

Social Network Analysis

Introduzione

L'analisi dei social network è un campo accademico che studia reti complesse come reti di telecomunicazioni, reti informatiche, reti biologiche, reti cognitive e semantiche e reti sociali. Questo campo si basa su teorie e metodi provenienti dalla teoria dei grafi, dalla meccanica statistica, dall'analisi dei dati e dalla visualizzazione delle informazioni informatiche e dalla struttura sociale della sociologia.

Cos'è una rete?

Una rete è un modello di dati molto generico con ampie applicazioni in vari campi. **Essa consiste in un insieme di oggetti e una collezione di collegamenti che identificano una relazione tra coppie di oggetti. L'assunzione è che esista una connessione tra questi oggetti che ci aiuterà a comprendere meglio ciò che stiamo studiando.** In matematica, le reti vengono chiamate grafi, gli oggetti sono conosciuti come vertici o nodi, e i collegamenti tra di essi vengono chiamati archi. Lo studio di questi oggetti costituisce un sotto-campo della matematica noto come teoria dei grafi.

Sistemi e reti complesse

I sistemi complessi sono costituiti da molte parti interconnesse e presentano un'organizzazione complicata delle connessioni. Questi sistemi mostrano proprietà emergenti che non possono essere facilmente dedotte dalle proprietà delle singole parti. Un famoso aforisma di Aristotele afferma che "il tutto è più della somma delle sue parti". I sistemi complessi sono spesso rappresentati come reti complesse.

Le reti complesse possono essere trovate in molti contesti, come le connessioni tra i neuroni nel cervello, le interazioni tra geni e proteine, l'infrastruttura delle telecomunicazioni, le reti commerciali, le economie umane e il clima. Questi esempi illustrano come le reti siano presenti in diversi ambiti e come l'analisi delle reti possa aiutare a comprendere i fenomeni complessi che si verificano in questi sistemi.

Focus di SNA

L'analisi delle reti sociali si concentra sullo studio dei legami sociali tra individui. Nel corso della storia umana, le collezioni di legami sociali tra individui sono cresciute costantemente grazie ai progressi tecnologici che hanno facilitato i viaggi lontani, la comunicazione globale e l'interazione digitale. **L'analisi delle reti sociali utilizza i nodi per rappresentare gli individui e i collegamenti per rappresentare le relazioni personali tra loro.**

- L'esempio di Moreno con i sociogrammi è spesso considerato come uno dei primi esempi di analisi e visualizzazione delle reti sociali. Attraverso la **mappatura delle affinità sociali di un gruppo di individui, i sociogrammi di Moreno** visualizzano le relazioni tra gli alunni di una classe. Questo esempio mostra come le reti sociali possano essere utilizzate per **studiare le interazioni tra le persone.**
- Un altro esempio di analisi delle reti sociali è il caso del **club di karate di Zachary**. Zachary ha osservato i membri di un club di karate per un periodo di 2 anni e ha costruito una rete di amicizie tra i membri del club. Dopo una disputa tra l'amministratore del club e l'istruttore, il club si è diviso in due gruppi. Questo esempio evidenzia la **struttura di comunità nelle reti sociali e mostra come un algoritmo possa predire l'evoluzione sociale di un gruppo basandosi sulla rete di amicizie.**

Attraverso l'analisi delle relazioni tra individui, questa disciplina aiuta a comprendere i modelli sociali, le dinamiche e le caratteristiche delle reti sociali. Le applicazioni pratiche dell'analisi delle reti sociali spaziano dalla prevenzione della criminalità all'analisi delle dinamiche organizzative, alla previsione delle epidemie e alla comprensione dei comportamenti sociali.

Grafi

The Bridges of Königsberg:

Nel 1735 a Königsberg, capitale della Prussia orientale, una città commerciale dell'epoca, c'erano sette ponti che attraversavano il fiume Pregel che circondava la città. La domanda era se fosse possibile attraversare tutti e sette i ponti senza passare due volte sullo stesso ponte. Questo problema è stato risolto da **Leonard Euler**, un matematico, che ha costruito un grafo con quattro nodi (A, B, C, D), corrispondenti a quattro aree di terra, e sette archi, corrispondenti ai ponti. Euler ha offerto una dimostrazione matematica che un percorso di questo tipo non esiste. Non è possibile trovare un percorso continuo che attraversi i sette ponti senza passare due volte sullo stesso ponte. Un percorso che attraversa tutti i ponti può avere solo un punto di partenza e un punto di arrivo. **Un percorso del genere non può esistere in un grafo con più di due nodi con un numero dispari di archi.** Il grafo di Königsberg aveva quattro nodi con un numero dispari di archi, quindi non esisteva un percorso che soddisfacesse il problema. Successivamente, nel 1875, gli abitanti di Königsberg costruirono un nuovo ponte tra B e C, lasciando solo un nodo con un numero dispari di archi. Questo avrebbe dovuto consentire di trovare il percorso desiderato. **La dimostrazione di Euler fu la prima volta che qualcuno risolse un problema matematico utilizzando un grafo.** Da questo esempio possiamo trarre **due messaggi fondamentali:** alcuni problemi diventano più semplici e più trattabili se vengono rappresentati come grafi, e i grafi hanno proprietà che limitano o potenziano il loro comportamento.

Grafi come modelli di reti:

I grafi sono utili perché fungono da **modelli matematici** delle strutture di rete. Ad esempio, l'ARPANET, la rete di comunicazione che ha dato origine a Internet, può essere rappresentata come un grafo in cui i nodi sono dispositivi e gli archi rappresentano collegamenti diretti lungo i quali possono essere trasmessi messaggi. Nelle reti di informazione, i nodi sono risorse informative come pagine web o documenti, mentre gli archi sono connessioni logiche come hyperlink o citazioni. Nelle reti sociali, i nodi rappresentano persone o gruppi di persone, mentre gli archi rappresentano interazioni sociali di diverso tipo.

Proprietà e concetti fondamentali:

Una proprietà chiave di ogni nodo è il suo **grado**, che rappresenta il numero di archi che lo collegano ad altri nodi. In una rete non diretta, il numero totale di archi L è dato da $L = (1/2) \sum_i k_i$, dove k_i è il grado del nodo i . Un'altra proprietà importante è il grado medio $\langle k \rangle$, che per una rete non diretta è dato da $\langle k \rangle = (2L)/N$, dove N è il numero di nodi. Per le reti dirette, si fa distinzione tra il grado in ingresso (numero di archi che puntano al nodo i) e il grado in uscita (numero di archi che partono dal nodo i verso altri nodi). Il grado totale di un nodo è dato dalla somma del grado in ingresso e del grado in uscita. La distribuzione dei gradi fornisce la probabilità che un nodo selezionato casualmente abbia un certo grado. È possibile ottenere il numero di nodi con un certo grado dalla distribuzione dei gradi.

Matrice di adiacenza:

La matrice di adiacenza è utile per descrivere una rete attraverso un elenco dei suoi archi. In una rete diretta, la matrice di adiacenza ha N righe e N colonne, dove A_{ij} è uguale a 1 se c'è un arco dal nodo j al nodo i , altrimenti è 0. Il grado di un nodo può essere ottenuto direttamente dagli elementi della matrice di adiacenza. Per reti non dirette, il grado di un nodo è dato da $k_i = \sum_j A_{ij}$. Per reti dirette, il grado in ingresso è dato da $k_{in} = \sum_j A_{ij}$ e il grado in uscita da $k_{out} = \sum_j A_{ji}$.

Reti reali sono sparsi:

Teoricamente, il numero massimo di archi in una rete con N nodi è $N(N-1)/2$. Tuttavia, nella realtà, la maggior parte delle reti reali è sparsa, il che significa che il numero effettivo di archi è molto inferiore a questo valore massimo. Ad esempio, la rete del Karate Club di Zachary ha 34 nodi e 78 archi, che è molto inferiore al numero **massimo teorico** di 561 archi. **La sparsità ha importanti conseguenze per l'esplorazione**

e la memorizzazione delle reti reali. Ad esempio, è più efficiente memorizzare solo l'elenco degli archi anziché la matrice di adiacenza completa quando si lavora con una rete di grandi dimensioni.

Reti pesate:

A volte è necessario studiare reti pesate, in cui ogni arco ha un peso associato. Ad esempio, in una rete di chiamate telefoniche, il peso potrebbe rappresentare il numero totale di minuti trascorsi da due individui al telefono, mentre in una rete di amicizie potrebbe rappresentare il numero totale di anni trascorsi da due amici. In reti pesate, **gli elementi della matrice di adiacenza rappresentano il peso dell'arco.**

Percorsi:

Un percorso in una rete è una sequenza di archi in cui la destinazione di ogni arco è l'origine del successivo. La lunghezza del percorso è il numero di archi che lo compongono. Un tipo importante di percorso è un **ciclo**, che è **un percorso con almeno 3 archi in cui il primo e l'ultimo nodo sono gli stessi, ma gli altri nodi sono distinti. I cicli sono spesso presenti nelle reti di trasporto e comunicazione per consentire la ridondanza e fornire percorsi alternativi.**

Connettività:

La connettività di una rete si riferisce alla capacità di ogni nodo di raggiungere ogni altro nodo tramite un percorso. Se ogni coppia di nodi ha un percorso tra di loro, la rete è detta connessa. Tuttavia, in alcuni contesti, le reti possono non essere necessariamente connesse. Ad esempio, in una rete sociale potrebbero esistere due persone che non sono connesse tra loro. È possibile identificare le componenti connesse di una rete, che sono sottoinsiemi di nodi in cui ogni nodo ha un percorso verso ogni altro nodo all'interno del sottoinsieme. Una rete con una sola componente connessa è chiamata grafo connesso. **In una rete disconnessa, la matrice di adiacenza può essere organizzata in una forma a blocchi, con ciascun blocco che rappresenta una componente connessa separata.**

Componenti giganti:

Considera una rete sociale globale in cui esiste un arco tra due persone se sono amiche. Questa rete globale delle amicizie è probabilmente non connessa. Tuttavia, potrebbe esistere una componente connessa molto grande che comprende un numero significativo di nodi. Anche se la rete globale non è connessa, questa componente gigante può estendersi in molte parti del mondo. Questo fenomeno si osserva spesso nelle reti complesse di grandi dimensioni. La presenza di una componente gigante è importante per la possibile diffusione di malattie sessualmente trasmissibili, in quanto anche una persona che ha avuto un solo partner può essere parte di questa componente gigante e può esistere una serie di percorsi di trasmissione potenziale.

Percorso più breve:

Il percorso più breve tra due nodi i e j è il percorso con il **minor numero di archi**. Il percorso più breve è spesso chiamato distanza e viene indicato come d_{ij} . In una rete non diretta, $d_{ij} = d_{ji}$, mentre in una rete diretta di solito $d_{ij} \neq d_{ji}$.

Diametro:

Il diametro di una rete è la massima **distanza più breve all'interno della rete**, cioè la distanza più grande registrata tra tutte le coppie di nodi. La lunghezza media dei percorsi $\langle d \rangle$ è la distanza media tra tutte le coppie di nodi nella rete. $\langle d \rangle$ viene calcolato solo per i nodi che fanno parte della stessa componente connessa.

Strong and Weak Ties:

Il concetto di forza e debolezza dei legami sociali è fondamentale nello **studio del flusso delle informazioni** all'interno di una rete sociale. Negli anni '60, il sociologo **Granovetter** formulò la famosa ipotesi della "**forza dei legami deboli**". Egli intervistò persone che avevano recentemente **cambiato lavoro** per capire come avevano scoperto questa opportunità. Scoprì che molte persone apprendevano informazioni tramite contatti personali. Ma quali contatti personali? Granovetter propose **due diverse prospettive sull'amicizia** a distanza: **strutturale** e **interpersonale**. Gli amici a distanza spesso venivano descritti come conoscenti piuttosto che amici stretti. Ciò ci aiuta a riflettere sull'architettura delle reti sociali in generale, considerando come tali amicizie si estendano su diverse porzioni della rete e le conseguenze locali di amicizie forti o deboli.

Chiusura triadica:

È importante considerare come una rete si evolve nel tempo, osservando i meccanismi con cui i nodi entrano e escono dalla rete e come si formano e scompaiono i collegamenti. **Se due persone in una rete sociale hanno un amico in comune, è più probabile che diventino amici a loro volta in futuro.** La formazione di un collegamento tra B e C produce un triangolo e si parla di "**triadic closure**" perché il nuovo collegamento ha l'effetto di chiudere il terzo lato di questo triangolo. In generale, si osserva un numero significativo di nuovi collegamenti formati attraverso la triadic closure nel tempo. La triadic closure ha un ruolo fondamentale nelle reti sociali e ha motivato la formulazione di **misure per catturarne la presenza**, come il **coefficiente di clustering**. Il coefficiente di clustering **misura il grado in cui i vicini di un dato nodo sono collegati tra loro, ovvero la frazione di amici del nodo che sono collegati tra loro tramite collegamenti**. Questo coefficiente viene **calcolato per ogni nodo e fornisce un'indicazione della densità locale dei collegamenti nei suoi dintorni**. Più intensamente la triadic closure opera nella rete, maggiore sarà il coefficiente di clustering.

Average Clustering Coefficient:

Mentre il coefficiente di clustering locale è una proprietà di un singolo nodo, il coefficiente di clustering medio rappresenta la **media di tutti i coefficienti di clustering dei nodi nell'intero grafo**. Questa misura ci fornisce una visione globale del livello di clustering nella rete. Le ragioni per cui B e C diventano più probabilmente amici possono essere riassunte in tre punti: **opportunità** (aumentata), **fiducia** (che potrebbe mancare tra due persone non collegate tra loro) e **incentivi**.

La forza dei legami deboli:

Vediamo come la triadic closure si collega alle interviste di Granovetter. Supponiamo che le informazioni sulle nuove opportunità di lavoro siano relativamente scarse. Se qualcuno è a conoscenza di un'opportunità di lavoro interessante, ha informazioni a cui tu non hai accesso. Consideriamo questa semplice rete: A ha quattro amici, ma il collegamento con B raggiunge una parte diversa della rete. Possiamo supporre che A, C, D ed E siano esposti a opinioni/informazioni simili, a differenza di B.

Un collegamento che unisce due nodi A e B è chiamato "**bridge**" se l'eliminazione di quel collegamento farebbe sì che A e B si trovino in due componenti diverse. Quel collegamento è l'unico percorso tra A e B.

Ma dalla nostra discussione sulle componenti giganti e sul fenomeno dei "mondi piccoli", sappiamo anche che i ponti sono estremamente rari nelle reti sociali reali. In una rete sociale più grande, avremmo probabilmente dei "**local bridges**". Un collegamento tra due nodi A e B in un grafo è un "local bridge" se i suoi estremi A e B non hanno amici in comune. In altre parole, l'eliminazione del collegamento aumenterebbe **la distanza tra A e B in modo strettamente superiore a 2**. La "**span**" di un "local bridge" è la distanza che i suoi estremi avrebbero l'uno dall'altro se il collegamento venisse eliminato. Ad esempio, il collegamento A-B è un "local bridge" con una span di 4, mentre nessun altro collegamento nel grafo è un "local bridge" perché, per ogni altro collegamento, gli estremi rimarrebbero a una distanza di 2.

Tornando a Granovetter, se un nodo come A vuole ottenere informazioni davvero nuove, spesso può ottenerle attraverso amici connessi da un "local bridge". Il gruppo coeso di cui fa parte A è formato da persone che hanno approssimativamente le stesse informazioni di A.

Granovetter ha suggerito una versione più formale di questa ipotesi: se un nodo A ha legami forti con altri due nodi B e C, e non esiste alcun collegamento (sia forte che debole) tra B e C, allora si dice che il nodo A viola la "Strong Triadic Closure" (STC) property. Se A non viola questa proprietà, si dice che A soddisfa la STC property.

La proprietà STC è troppo estrema per essere applicata a tutti i nodi di una grande rete sociale, ma è **utile per analizzare le conseguenze strutturali dei legami forti e deboli.**

Possiamo affermare che se un nodo A in una rete soddisfa la proprietà STC e partecipa ad almeno due legami forti, allora ogni "local bridge" a cui A partecipa deve essere un legame debole. Possiamo giustificare questa affermazione come una dichiarazione matematica.

La presenza di "local bridges" che sono legami deboli è importante perché i legami deboli ci connettono a nuove fonti di informazione e opportunità. La loro proprietà di "local bridge" è direttamente correlata alla loro debolezza. **I legami deboli svolgono un ruolo duplice: sono connessioni deboli ma allo stesso tempo preziose per accedere a parti della rete difficili da raggiungere.** Questo doppio ruolo rappresenta la forza dei legami deboli.

Tie Strength and Network Structure in Large-Scale Data:

Da molti anni le previsioni di Granovetter sono rimaste senza essere testate su grandi reti sociali. È stato difficile trovare dati che catturassero la forza dei legami in contesti realistici. Tuttavia, con l'avvento delle comunicazioni digitali, la situazione ha cominciato a cambiare, anche se le **definizioni di legami deboli e ponti locali impongono una netta dicotomia:** un collegamento è considerato o un legame forte o un legame debole, un collegamento è considerato un ponte locale o non lo è. **Abbiamo bisogno di gradazioni più fluide quando lavoriamo con dati reali su larga scala.**

Struttura e forza dei legami nelle reti di comunicazione mobile:

Onnela et al. hanno studiato la rete di "chi parla con chi" gestita da un provider di telefonia mobile. I nodi sono gli utenti dei telefoni cellulari e un collegamento unisce due nodi se si sono chiamati a vicenda. **Come possiamo distinguere tra legami forti e deboli?**

Forza dei legami: La forza del collegamento può essere definita come una quantità numerica, ad esempio il numero totale di minuti trascorsi al telefono dai due nodi. Possiamo ordinare tutti i collegamenti in base alla forza del legame e chiederci in quale percentile si colloca ciascun collegamento.

Ponti locali: I dati contengono molto pochi collegamenti che costituiscono un ponte locale. Possiamo **rilassare la definizione e considerare "quasi" ponti locali.** Onnela et al. definiscono l'"**overlap del vicinato**" di un collegamento tra A e B come il **numero di nodi che sono vicini sia ad A che a B diviso il numero di nodi che sono vicini ad almeno uno dei due.**

Con queste definizioni, Onnela et al. hanno formulato una serie di domande basate sulle previsioni di Granovetter:

Dipende dall'intensità del legame l'overlap del vicinato? Le previsioni sulla forza dei legami indicano che **l'overlap del vicinato dovrebbe aumentare all'aumentare della forza del legame.** I legami deboli servono a connettere diverse comunità strettamente coese, ognuna contenente un maggior numero di legami forti? Onnela et al. hanno affrontato questa domanda nel seguente modo: hanno eliminato un collegamento alla volta dalla rete, in ordine decrescente di forza. La componente gigante si è ridotta costantemente, diminuendo gradualmente di dimensione. Poi hanno fatto lo stesso partendo dai legami più deboli. La componente gigante si è ridotta più rapidamente e i suoi resti si sono spezzati improvvisamente una volta che è stato rimosso un numero critico di legami deboli. Cosa possiamo apprendere da questo? **I legami deboli forniscono la struttura connettiva più cruciale per mantenere insieme diverse comunità e mantenere integra la componente gigante.**

Questi risultati dimostrano come la forza dei legami e la struttura della rete siano strettamente correlate e come i legami deboli svolgano un ruolo fondamentale nella coesione sociale e nell'integrità della rete. È

importante considerare non solo la quantità di interazioni, ma anche la loro intensità per comprendere appieno le dinamiche delle reti sociali.

Tie Strength on Facebook:

Nel 2009, i ricercatori di Facebook hanno analizzato i legami di amicizia segnalati nei profili degli utenti. Hanno cercato di capire fino a che punto ciascun legame fosse utilizzato per l'interazione sociale e quali fossero i legami forti tra gli amici di un utente. Hanno **definito 3 categorie di legami**:

- Comunicazione **reciproca**: entrambi gli utenti si sono scambiati messaggi l'uno con l'altro.
- Comunicazione **unidirezionale**: l'utente ha inviato uno o più messaggi all'altro (indipendentemente dal fatto che fossero ricambiati).
- Relazione **mantenuta**: l'utente ha seguito le informazioni sull'amico (indipendentemente dal fatto che si sia svolta una comunicazione).

Le categorie non sono mutuamente esclusive. Ci sono due regioni distinte con una grande quantità di chiusura triadica, in cui sopravvivono molti più legami nelle categorie di comunicazione e relazioni mantenute nella regione superiore (amici delle scuole superiori, colleghi).

Tie Strength on Twitter:

Huberman, Romero e Wu hanno fatto qualcosa di simile su Twitter. Hanno considerato per ciascun utente il numero di persone che seguiva e hanno definito i legami forti come gli utenti ai quali l'utente in questione aveva inviato almeno due messaggi.

Closure, Structural Holes, and Social Capital:

Fino ad ora ci siamo concentrati sul ruolo giocato dai diversi tipi di legami in una rete sociale, ma cosa succede con i nodi? **Non tutti i nodi sono uguali.** Ad esempio, il nodo A è soggetto a una considerevole chiusura triadica ed ha un alto coefficiente di clustering. **L'embeddedness di un legame è definita dal numero di vicini comuni condivisi dai due estremi. Se il valore di embeddedness è 0, allora si tratta di un ponte locale. Gli edge con un'embeddedness significativa indicano una maggiore fiducia e sicurezza nelle transazioni (sociali, economiche, ecc.) tra le due persone coinvolte.**

Le buche strutturali, d'altra parte, rappresentano un vantaggio per il nodo B. B occupa una posizione che riempie lo spazio vuoto tra due gruppi di nodi. Questa posizione offre a B l'accesso a informazioni provenienti da diverse fonti e l'opportunità di generare idee innovative combinando tali informazioni. Tuttavia, le interazioni di B sono meno integrate all'interno di un gruppo e meno protette dalla presenza di amici reciproci.

Tutti questi concetti sono correlati alla nozione di **capitale sociale**, che è difficile da definire. Possiamo pensare alle **strutture sociali come facilitatori per individui e/o gruppi.**

Betweenness Measures and Graph Partitioning:

Ci concentreremo su metodi per suddividere una rete in una serie di regioni dense con connessioni sparse. Ci riferiamo a questo problema come "graph partitioning". Esistono due classi principali di metodi:

- **Divisive**: identificano e rimuovono i collegamenti tra regioni densamente connesse (dal più generale al più specifico).
- **Agglomerative**: individuano i nodi che probabilmente appartengono alla stessa regione e li uniscono insieme (dal più specifico al più generale).

Esistono molteplici approcci e metodi differenti che vengono utilizzati per la suddivisione dei grafi. Comprendere quali metodi funzionino meglio in specifici contesti è oggetto di ricerca.

Betweenness:

Come possiamo identificare i collegamenti da rimuovere? Una metrica utile è il "betweenness" (centralità di intermediazione).

Il problema 1 riguarda la scelta degli archi da rimuovere quando si considerano i "ponti" (bridges). Ci possono essere diversi ponti, quindi quale dovremmo rimuovere per primi?

Il problema 2 riguarda la mancanza di ponti locali (local bridges). Tuttavia, i ponti locali sono importanti perché fanno parte del percorso più breve tra coppie di nodi appartenenti a regioni diverse della rete. Senza un ponte locale specifico, i percorsi potrebbero diventare più lunghi.

Possiamo definire una nozione di flusso nel grafo. Per ogni coppia di nodi A e B collegati da un percorso, abbiamo un flusso di 1 unità lungo gli archi da A a B. Questo flusso si divide equamente lungo tutti i possibili percorsi più brevi da A a B. Se ci sono k percorsi più brevi, $1/k$ unità di flusso passano lungo ciascun percorso.

Definiamo **il betweenness di un arco come la quantità totale di flusso che porta in considerazione il flusso tra tutte le coppie di nodi che utilizzano l'arco. In modo simile, possiamo calcolare il betweenness a livello di nodi.**

Il metodo di Girvan-Newman:

Trova l'arco (o gli archi) con il betweenness più alto e lo rimuove. Ricalcola tutti i betweenness e rimuove l'arco (o gli archi) con il betweenness più alto (procedura nidificata). Continua fino a quando rimangono archi nel grafo. **Girvan e Newman hanno mostrato l'effetto del loro metodo su dataset reali, ad esempio il Karate Club di Zachary. Il metodo funziona bene per reti di dimensioni moderate (alcune migliaia di nodi). Tuttavia, per reti più grandi, calcolare il betweenness ad ogni passo risulta costoso e si possono utilizzare metodi approssimati o metodi agglomerativi.**

Reti nei loro contesti circostanti:

Nel campo dell'analisi delle reti sociali, ogni individuo all'interno di una rete possiede un insieme unico di caratteristiche personali. Le somiglianze tra le persone possono influenzare significativamente la formazione dei collegamenti all'interno di una rete. Tuttavia, non possiamo trascurare i contesti circostanti, ovvero i fattori esterni ai nodi e agli archi della rete che influenzano l'evoluzione stessa della rete.

Omofilia:

Un concetto chiave nell'analisi delle reti sociali è l'omofilia, che indica la tendenza delle persone ad essere simili ai loro amici. Di solito, i nostri amici non assomigliano a un campione casuale della popolazione generale, ma sono simili a noi in vari aspetti, come l'età, l'appartenenza razziale o etnica e caratteristiche come il luogo di residenza, il lavoro, l'occupazione, gli interessi e le credenze. Sebbene molti di noi abbiano amicizie che superano tali confini, nel complesso i collegamenti all'interno di una rete sociale tendono a connettere persone simili tra loro. L'omofilia può quindi guidare la formazione dei collegamenti all'interno di una rete sociale.

La chiusura triadica e i contesti sociali:

Abbiamo precedentemente ipotizzato che la chiusura triadica, che si riferisce alla tendenza delle persone a formare collegamenti con gli amici dei propri amici, sia motivata da meccanismi intrinseci come la fiducia, le opportunità e gli incentivi. Tuttavia, i contesti sociali forniscono anche una base naturale per la chiusura triadica. Ad esempio, se A è amico di B e C, l'omofilia suggerisce che B e C siano simili tra loro. Pertanto, c'è una probabilità elevata che si sviluppi un'amicizia tra B e C, anche se non sono consapevoli di avere A come amico comune. In questo modo, i contesti sociali influenzano la formazione dei collegamenti tra persone simili all'interno di una rete.

Misurare l'omofilia:

Per valutare se una rete manifesta omofilia per una particolare caratteristica (come l'età), è possibile applicare un semplice test. Ad esempio, supponiamo di avere la rete di amicizia di una classe scolastica e ipotizziamo

che la rete manifesti omofilia di genere. Se i ragazzi tendono ad avere amici maschi e le ragazze tendono ad avere amiche femmine, possiamo verificare se la rete è coerente con l'omofilia di genere. Possiamo utilizzare una misura numerica per confrontare la frazione effettiva di collegamenti tra generi con quella che ci si aspetterebbe in un caso di assenza di omofilia. Se la frazione di collegamenti tra generi è significativamente inferiore a un valore di riferimento (come $2pq$), allora ci sono evidenze di omofilia nella rete.

Selezione e influenza sociale:

Nell'analisi delle reti sociali, i concetti di selezione e influenza sociale sono fondamentali per comprendere come si formano e si evolvono le connessioni tra le persone. La selezione si riferisce alla tendenza delle persone a formare amicizie con individui simili a loro, sia a livello personale che in base alle caratteristiche immutabili o mutevoli. Le caratteristiche immutabili sono presenti sin dalla nascita e influenzano la formazione delle connessioni, mentre le caratteristiche mutevoli, come i comportamenti, le attività, gli interessi, le credenze e le opinioni, possono influenzare la rete sociale in modo più complesso.

D'altra parte, l'influenza sociale indica che le connessioni sociali esistenti in una rete influenzano le caratteristiche individuali dei nodi. Le persone possono modificare i loro comportamenti per avvicinarli a quelli dei loro amici. Quindi, l'influenza sociale rappresenta il rovescio della selezione: anziché le caratteristiche individuali guidare la formazione dei collegamenti, i collegamenti esistenti nella rete contribuiscono a plasmare le caratteristiche delle persone.

L'intreccio di selezione e influenza sociale:

Quando si osserva un istante specifico di una rete, può essere difficile distinguere gli effetti distinti e il contributo specifico della selezione e dell'influenza sociale. Ad esempio, potremmo chiederci se le persone si sono adattate ai comportamenti dei loro amici per assomigliare a loro o se hanno cercato amici che erano già simili a loro stessi. Per affrontare tali domande, sono necessari studi longitudinali delle reti sociali, che analizzano i cambiamenti comportamentali che si verificano dopo modifiche alle connessioni di una persona o dopo che una persona cambia il proprio comportamento. Questi studi consentono di valutare l'impatto distintivo della selezione e dell'influenza sociale sulla formazione delle reti.

Equilibrio tra selezione e influenza sociale:

L'equilibrio tra selezione e influenza sociale può variare in diverse situazioni. Gli studi suggeriscono che l'influenza sociale può essere più moderata rispetto alla selezione, sebbene entrambi gli effetti siano presenti. Comprendere questi meccanismi è fondamentale per identificare possibili interventi nelle reti sociali. Ad esempio, se l'uso di droghe manifesta omofilia in una rete sociale, possiamo implementare programmi di intervento mirati a studenti specifici, con l'obiettivo di influenzarli a smettere di usare droghe. Tuttavia, è importante considerare se l'omofilia osservata si basa principalmente sull'influenza sociale o sulla selezione, poiché ciò può influire sull'efficacia degli interventi.

Comprendere l'equilibrio tra selezione e influenza sociale offre informazioni cruciali per la progettazione di interventi efficaci nelle reti sociali.

Temi critici per i Big Data

Il tema "Critical Questions for Big Data" affronta le problematiche e le sfide associate alla gestione dei Big Data, basandosi sul documento di Boyd e Crawford del 2012. I Big Data sono definiti come insiemi di dati così ampi da richiedere supercomputer per la loro elaborazione, ma questa definizione è considerata obsoleta. La quantità di dati disponibili può essere considerevole, ma non è questa la caratteristica principale dei Big Data. Invece, i Big Data riguardano la capacità di cercare, aggregare e mettere in relazione grandi insiemi di dati.

I Big Data possono essere interpretati in due modi:

- come uno strumento potente per affrontare questioni sociali
- come una manifestazione preoccupante del Grande Fratello, che permette invasioni della privacy, riduzione delle libertà civili e aumento del controllo da parte di entità statali e aziendali

Come per tutti i fenomeni socio-tecnologici, ci sono sfumature e osserviamo un aumento dell'automazione nella raccolta e nell'analisi dei dati, algoritmi che possono estrarre e illustrare schemi di comportamento umano su larga scala. Tuttavia, quali sistemi guidano e regolamentano queste pratiche?

Ricerche di Boyd e Crawford

Boyd e Crawford identificano quattro forze che regolano i sistemi sociali, applicabili anche ai Big Data:

- il mercato
- la legge
- norme sociali
- l'architettura (codice).

Queste forze entrano in conflitto nella gestione dei Big Data. Ad esempio, il mercato vede nei Big Data un'opportunità, mentre la legge solleva preoccupazioni sulla privacy. Le norme sociali e il codice affrontano l'idea di personalizzazione, che offre informazioni più rilevanti ma pone anche domande etiche e frammenta il pubblico.

Il mercato:

I ricercatori offrono sei provocazioni riguardanti i Big Data nel contesto dei social media. Queste provocazioni pongono domande difficili che non hanno risposte semplici. I Big Data non si riferiscono solo a grandi insiemi di dati e agli strumenti e alle procedure utilizzati per manipolarli e analizzarli, ma anche a un cambiamento computazionale nel pensiero e nella ricerca. I Big Data sono diventati un sistema di conoscenza in grado di influenzare la nostra comprensione delle reti umane e delle comunità. Tuttavia, gli strumenti specializzati dei Big Data hanno limiti e restrizioni intrinseche. Ad esempio, piattaforme come Twitter e Facebook offrono funzioni di archiviazione e ricerca molto limitate.

Le rivendicazioni di oggettività e accuratezza nei Big Data sono fuorvianti perché tutti i ricercatori interpretano i dati. Anche se un modello matematico può essere corretto o un esperimento sembra valido, non appena un ricercatore cerca di comprenderne il significato, inizia il processo di interpretazione. Non tutte le interpretazioni sono uguali, e nemmeno tutti i numeri sono neutrali. Le decisioni di progettazione che determinano cosa verrà misurato derivano anche dall'interpretazione. Ad esempio, nel caso dei dati dei social media, il processo di pulizia dei dati e le scelte sulle metriche da utilizzare sono influenzate dalle interpretazioni e dagli obiettivi dei ricercatori. Ciò può portare a una visione distorta o parziale della realtà sociale.

La rappresentatività:

Un'altra provocazione riguarda la questione della rappresentatività dei dati nei Big Data. Molti studi si basano su dati provenienti da piattaforme online, ma ciò può portare a un'immagine distorta della società, in quanto alcune popolazioni o gruppi possono essere sottorappresentati o completamente esclusi. Ciò solleva interrogativi sulla validità e l'accuratezza delle conclusioni tratte dai Big Data.

Responsabilità e dell'accountability:

Una terza provocazione riguarda la questione della responsabilità e dell'accountability nei confronti dei Big Data. Chi è responsabile delle azioni e delle decisioni basate sui dati? Quali sono le implicazioni etiche e legali delle pratiche di raccolta e utilizzo dei Big Data? Queste sono domande complesse che richiedono una riflessione approfondita e una regolamentazione adeguata.

La trasparenza e l'accessibilità

Un'altra sfida riguarda la trasparenza e l'accessibilità dei Big Data. Spesso i dati sono raccolti e analizzati da aziende o organizzazioni che possono avere interessi commerciali o politici. Ciò solleva preoccupazioni sulla trasparenza delle pratiche di raccolta e analisi dei dati e sulla possibilità per gli individui di comprendere e controllare l'uso dei loro dati personali.

Norme sociali e legge

Infine, la quinta provocazione riguarda il potenziale impatto dei Big Data sulla disuguaglianza sociale.

L'accesso e la capacità di utilizzare i Big Data sono concentrati in mano a poche aziende e istituzioni che dispongono delle risorse necessarie per sfruttarli appieno. Ciò può portare a un divario digitale e a un aumento delle disuguaglianze nel potere e nell'informazione.

Big Data

Il termine Big Data rappresenta un dataset di dimensioni molto grandi. I problemi dei big data:

- Data quality: la qualità delle informazioni raccolte non è sempre ottimale. Da tenere in considerazione:
- Data sparsity: quando esistono valori senza informazioni
- Data noise: informazioni inutili (ridondanze, non di interesse)
- Representativeness: un piccolo campione che rappresenta la popolazione - Data bias: elencati precedentemente.
- Data acquisition: limitazioni degli API che vengono utilizzati, keyword specifiche che possono limitare la raccolta di dati.
- Data filtering: rimozione di informazioni inutili/non necessarie, il filtro dei testi potrebbe vincolare un processo di analisi (stepwords per analizzare la personalità di un utente), rimozione di utenti non più attivi all'interno della piattaforma di interesse.

Bias

I bias sono fattori che influenzano il comportamento/decisioni di un individuo o un gruppo di individui.

- Population biases: quando si prende come campione un gruppo di individui che fa parte della popolazione ma che non la rappresenta del tutto. Come ad esempio gli studenti delle superiori che vengono presi come esempio per i teenagers ma non li rappresentano tutti.
- Behavioural biases: differenza nel comportamento di un utente in piattaforme diverse o contesti diversi.
- Content production biases: contenuti diversi generati dal bias comportamentale dell'utente
- Linking biases: differenza degli attributi di un utente all'interno della rete causato dal loro collegamento con altri nodi o dalla loro attività all'interno della rete (sempre causato dal bias comportamentale).
- Temporal variations: differenze nel tempo. (primacy e recency per esempio)
- Functional biases: causati dai meccanismi di una specifica piattaforma e che sono possibili solo all'interno di tale ambiente.
- Normative biases: causati da normative scritte o non scritte ma accettati da tutti all'interno di una piattaforma.
- External biases: culturali, disinformazione, eventi con grande impatto che vengono riflessi all'interno di una piattaforma social.