Reti sociali: analisi, rappresentazione e proprietà strutturali

Annalisa Socievole, Sabrina Celia

RT- ICAR-CS-18-03 **Luglio 2018**

Consiglio Nazionale delle Ricerche, Istituto di Calcolo e Reti ad Alte Prestazioni (ICAR)
– Sede di Cosenza, Via P. Bucci 8-9C, 87036 Rende, Italy, URL: www.icar.cnr.it
– Sezione di Napoli, Via P. Castellino 111, 80131 Napoli, URL: www.icar.cnr.it
– Sezione di Palermo, Via Ugo La Malfa, 153, 90146 Palermo, URL: www.icar.cnr.it

Indice

1 ABSTRACT

La nozione di *rete sociale* e dei metodi per la loro analisi hanno avuto un notevole interesse da parte della comunitá scientifica a causa delle implicazioni che lo studio delle relazioni tra gli oggetti che costituiscono una rete possono produrre. Ricerche in questo ambito hanno messo in evidenza che le reti complesse permettono di rispondere a domande relative alla ricerca comportamentale e sociale di individui dando definizioni formali su aspetti sociali, economici e politici. In questo documento vengono fornite le nozioni di base sulle reti sociali, la loro rappresentazione matematica, le varie misure definite per valutare gli oggetti della rete e le tipologie di reti individuate.

2 INTRODUZIONE

2.1 SCOPO DEL DOCUMENTO

La capacitá delle reti di rappresentare molti sistemi del mondo reale ha suscitato notevole interesse nell' area di ricerca delle reti complesse. Reti di collaborazione, internet, il world wide web, reti biologiche, reti di comunicazione e trasporto, reti sociali, sono solo alcuni esempi di reti. Le reti, in generale, sono costituite da un insieme di oggetti e da un insieme di interconnessioni tra questi oggetti. Nelle reti sociali, gli oggetti sono persone e le connessioni rappresentano le relazioni sociali quali interessi comuni, amicizia, religione. L'analisi delle reti sociali é utilizzata in vari ambiti, quali scienze comportamentali, economia, marketing, ingegneria. Il concetto piú importante é relativo alle relazioni esistenti tra gli oggetti coinvolti, quali ad esempio comunicazioni tra i membri di un gruppo, transazioni economiche tra corporazioni, commercio e trattati tra nazioni. Lo studio di tali relazioni puó servire a vari scopi, quali risoluzione collaborativa di problemi, ricerca di avvisi, sviluppo di competenze imparando dagli altri nodi della rete.

L'obiettivo di questo documento é quello di fornire i concetti formali di base per la rappresentazione delle reti e la descrizione di una serie di misure che permettono di definire il ruolo di un oggetto rispetto agli altri presenti nella rete.

2.2 STRUTTURA DEL DOCUMENTO

Il documento é organizzato nel seguente modo. La prossima sezione introduce il concetto di rete e definisce i termini usati nel documento. Nella sezione 4 si descrivono le varie tipologie di reti. Nella sezione 5 sono elencate le proprietá che caratterizzano una rete. Nella sezione 6 sono descritti i vari modelli di grafo che rappresentano reti di diversa tipologia.

2.3 ACRONIMI E TERMINI CHIAVE

Reti sociali, grafi, effetto piccolo mondo, distribuzione di grado, indici di centralitá, modelli di grafo.

3 Concetti base

Una *rete* é un insieme di oggetti chiamati *vertici* oppure *nodi*, con connessioni tra di loro chiamate *rami* (se non sono orientate) o *archi* (se sono orientate). Nel mondo esistono innumerevoli sistemi che prendono la forma di reti (chiamate anche ""grafi” nella letteratura matematica). Alcuni esempi includono Internet, il World Wide Web, reti sociali o altre connessioni tra individui, reti organizzative e reti di relazioni di business tra compagnie, reti neurali, reti metaboliche, reti distribuite e molte altre. Lo studio delle reti, nella teoria matematica dei grafi, é uno dei pilastri fondamentali nella matematica discreta. La soluzione al problema dei ponti di Konigsberg proposta da Eulero nel 1735, é spesso citata come la prima dimostrazione corretta nella teoria delle reti. Le reti sono state anche studiate nelle scienze sociali.

Negli ultimi anni si é registrato un nuovo movimento sostanziale nella ricerca sulle reti: si é passati dall'analisi di piccoli grafi e delle proprietá di singoli vertici o spigoli, allo studio delle proprietá statistiche di grafi di grandi dimensioni. Questo nuovo approccio ha tratto vantaggio dalla disponibilitá di computer e reti di comunicazioni piú potenti, che permettono di analizzare dati in una scala piú grande rispetto a quella precedente. Questo cambiamento di scala corrisponde ad un cambiamento nell'approccio analitico. Molti problemi che precedentemente erano stati risolti su piccole reti non hanno piú significato nelle grandi reti. C'é un'altra ragione per cui l'approccio allo studio delle reti é cambiato negli ultimi anni. Per reti formate da decine o centinaia di vertici é relativamente facile produrre una rappresentazione grafica della rete, con nodi e linee, e rispondere a specifiche domande sulla sua struttura esaminando la figura. L'occhio umano é uno strumento analitico molto potente per analizzare graficamente le reti ed é un modo eccellente per capirne la struttura. Con una rete costituita da milioni o bilioni di vertici, questo approccio non puó essere usato. Non si puó disegnare una figura formata da un milione di vertici, senza un computer che abbia uno strumento adatto per farlo ed anche in questo caso non é possibile un'analisi diretta fatta con l'occhio umano. Lo sviluppo recente di metodi statistici per quantificare reti di grandi dimensioni é un tentativo di imitare l'occhio umano nell'analisi delle reti nel XX secolo.

Un insieme di vertici uniti da rami é solo un tipo semplice di rete; esistono molti modi in cui le reti possono essere rese piú complesse. Per esempio, potrebbero esserci piú di un tipo differente di vertici in una rete o piú di un tipo differente di rami. Vertici o rami possono avere una varietá di proprietá numeriche e non associate con essi. Prendendo l'esempio di una rete sociale di persone, i vertici possono rappresentare uomini o donne, persone di nazionalitá differente, localitá, etá o altre cose. I rami possono rappresentare legami di amicizia, ma potrebbero rappresentare anche ostilitá, conoscenza professionale oppure vicinanza geografica. Possono avere pesi, che rappresentano, per esempio, quanto si conoscono due persone. I grafi orientati possono essere ciclici, questo significa che contengono anelli chiusi di rami, oppure aciclici, quando non ne hanno. Possono anche avere *hyperedges*, rami che collegano piú di due vertici insieme. I grafi che contengono hyperedges sono chiamati *hypergraphs*. Gli hyperedges possono essere usati per indicare un legame familiare in una rete sociale, per esempio n individui connessi tra di loro in virtú del fatto di appartenere alla stessa famiglia, possono essere rappresentati da un n-ramo che li collega. I grafi possono essere anche partizionati naturalmente in modi diversi. Ad esempio esistono *grafi bipartiti*: grafi che contengono vertici di due tipi, con rami che collegano solo vertici di tipo diverso. Inoltre i grafi possono evolversi nel tempo, con vertici o rami che appaiono o scompaiono, oppure proprietá variabili associate ad essi. In generale ci sono molti altri modi di rendere complesso un grafo.

Nella tabella 1 é riportato un glossario dei termini piú usati nel seguito. Nei paragrafi che seguono verrá data una classificazione generale delle reti presenti nel mondo reale e verranno descritte le proprietá piú importanti per lo studio delle reti.

Vertice	Unitá fondamentale di una rete, anche chiamato nodo.
Ramo	Linea non orientata che collega due vertici.
Arco	Linea orientata che collega due vertici.
Orientato/non-orientato	Un collegamento é orientato se é percorso solo in una direzione e non-orientato se é percorso in entrambe le direzioni. Un grafo é orientato se contiene solo archi. Un grafo non-orientato puó essere rappresentato da un grafo orientato avente due archi tra ogni coppia di vertici connessi, uno in ogni direzione.
Grado	Il numero di collegamenti connessi ad un vertice. Il grado di un vertice non é necessariamente uguale al numero di vertici adiacenti ad esso, dato che potrebbe esserci piú di un collegamento tra due vertici. Un grafo orientato ha sia un grado in ingresso che uno in uscita per ogni vertice, cioé il numero di collegamenti in ingresso e in uscita rispettivamente.
Componente	La componente alla quale un vertice appartiene é l'insieme di vertici che possono essere raggiunti da esso tramite cammini percorsi lungo i rami del grafo. In un grafo orientato un vertice ha sia una componente in ingresso che una in uscita, rispettivamente gli insiemi di vertici dai quali il vertice puó essere raggiunto o che possono essere raggiunti da esso.
Cammino Geodetico	Un cammino geodetico é il piú piccolo cammino che attraversa la rete da un vertice ad un altro. Potrebbe esserci piú di un cammino geodetico tra due vertici.
Diametro	Il diametro di una rete é la lunghezza (in numero di rami) del piú lungo cammino geodetico tra ogni coppia di vertici.

Tabella 1: Glossario dei termini

4 Tipologie di reti

Nel mondo reale possiamo distinguere diverse categorie di reti: reti sociali, reti di informazione, reti tecnologiche, reti biologiche e reti random.

4.1 RETI SOCIALI

Una rete sociale é un insieme di persone o gruppi di persone con alcuni tipi di contatti o interazioni tra di loro [17] [37]. Modelli di amicizia tra gli individui [20] [5], relazioni di business tra le compagnie [30] [26] e matrimoni misti tra le famiglie [19] sono tutti esempi di reti che sono state studiate nel passato. Tra le discipline accademiche, le scienze sociali sono quelle che da piú tempo studiano le reti del mondo reale [9] [17]. I primi lavori svolti su questo tema sono quelli di Jacob Moreno negli anni ’20 e ’30 [20], il cosidetto ""southern women study” di Davis et al [12] e i modelli matematici di Anatol Rapoport [3]. Negli ultimi anni, lo studio di comunitá di business [30] e gli schemi di relazioni sessuali [8] [15] [25] hanno attratto particolare attenzione.

Le tradizionali reti sociali studiate in passato spesso soffrivano di problemi di inaccuratezza, soggettivitá e piccole dimensioni dei campioni. La raccolta dei dati veniva effettuata solitamente interrogando direttamente i partecipanti, utilizzando questionari o interviste. Questi metodi limitano la dimensione della rete che puó essere osservata. Inoltre il rilevamento dei dati era influenzato dal giudizio soggettivo da parte degli intervistati; per esempio il modo in cui un partecipante definisce un amico, potrebbe essere molto diverso da come lo fa un altro. A causa di questi problemi, molti ricercatori si sono rivolti ad altri metodi per studiare le reti sociali. Una fonte di molti dati, relativamente affidabili, sono le reti di collaborazione. In questo tipo di reti i partecipanti collaborano in gruppi di diverso tipo e i link tra gli individui sono stabiliti dall’appartenenza agli stessi gruppi. Un classico esempio é la rete di collaborazione degli attori del cinema, che é documentata nel Database online Internet Movie (www.IMDb.com). In questa rete gli attori collaborano nei film e due attori vengono considerati connessi se sono apparsi insieme in almeno un film. Le proprietá statistiche di questa rete sono state analizzate da molti autori [1] [18].

4.2 RETI DI INFORMAZIONE

La seconda categoria di reti é quella chiamata in genere *rete di informazione* (e spesso anche "rete di conoscenza”). Un classico esempio di rete di informazione é la rete di citazioni tra pubblicazioni accademiche [22]. Molti articoli citano i lavori fatti in precedenza da altri su argomenti correlati. Queste citazioni formano una rete in cui i vertici sono gli articoli ed un arco che va dall’articolo A all’articolo B indica che A cita B. La struttura della rete di citazioni riflette la struttura dell’informazione memorizzata nei vertici, da qui il termine "rete di informazione”. Le reti di citazioni sono acicliche perché le pubblicazioni possono solo citare altre pubblicazioni che sono giá state scritte. In questo modo tutti gli spigoli della rete puntano a ritroso nel tempo, rendendo impossibili i cicli chiusi, o estremamente rari. Lo studio delle reti di citazioni ha il grosso vantaggio dell’enorme disponibilitá di dati e della loro accuratezza.

Un altro importante esempio di rete di informazione é il World Wide Web, che é una rete di pagine Web contenenti informazioni e collegate tra di loro tramite collegamenti ipertestuali [2]. Il web non deve essere confuso con Internet, che é una rete fisica di computer collegati insieme da fibre ottiche ed altri tipi di connessioni dati. Diversamente da una rete di citazioni, il World Wide Web é ciclico; non c’é un ordine naturale di siti e nessun vincolo che previene la comparsa di cicli chiusi.

Infine le reti di preferenze sono esempi di reti di informazione con due tipi di vertici (quindi bipartite), che rappresentano gli individui e l’oggetto delle loro preferenze, ad esempio libri o film, con rami che collegano ogni individuo ai libri o i film preferiti. Un esempio ampiamente studiato di una rete di preferenze é il database EachMovie delle preferenze dei film.

4.3 RETI TECNOLOGICHE

La terza classe di reti sono le reti tecnologiche, reti artificiali progettate tipicamente per la distribuzione di alcuni prodotti o risorse, come elettricitá o informazione. La rete elettrica é un buon esempio di questa classe. Essa é una rete che trasmette linee ad alta tensione trifase che ricopre un paese o una porzione di un paese. Altre reti distributive che sono state oggetto di studio includono la rete delle rotte aeree [7], reti stradali [21], ferrovie [36] e il traffico pedonale [10]. La rete telefonica e le reti di distribuzioni come quelle usate dagli uffici postali ricadono anche in questa categoria. Un altro importante esempio di rete tecnologica é Internet, la rete di connessioni fisiche tra computer.

4.4 RETI BIOLOGICHE

Molti sistemi biologici possono essere rappresentati come reti. Il piú classico esempio di rete biologica é la rete metabolica, che é una rappresentazione di sottostrati e prodotti metabolici con archi che li connettono se esiste una reazione metabolica che agisce su un dato sottostrato producendo un determinato prodotto. Un'altra importante classe di rete biologica é la rete genetica. L'espressione di un gene puó essere controllata dalla presenza di altre proteine, sia attivatori che inibitori, in modo che lo stesso genoma forma una rete di commutazione, con i vertici che rappresentano le proteine e gli archi che rappresentano la dipendenza della produzione di proteine dalle proteine di altri vertici. Le reti neurali sono un'altra classe di reti biologiche di considerevole importanza. Misurare la topologia di reti neurali reali é estremamente difficile, ma é stato fatto con successo in alcuni casi presenti in letteratura.

4.5 RETI RANDOM

Probabilmente il piú semplice modello di rete é il grafo random, studiato per la prima volta da Rapoport [33]. In questo modello, i rami sono posizionati in modo random (casuale) tra un numero fissato di n vertici per creare una rete in cui ognuno dei $\frac{1}{2}n(n-1)$ rami possibili é presente con una certa probabilitá p ed il numero di rami connessi ad ogni vertice (grado) é distribuito secondo una distribuzione binomiale o una distribuzione di Poisson. Molte caratteristiche interessanti di reti reali che hanno attirato l'attenzione dei ricercatori negli ultimi anni, tuttavia, riguardano i modi in cui le reti reali non si comportano come dei grafi random.

5 Proprietá delle reti

In questa sezione descriviamo alcune caratteristiche comuni a reti di diverso tipo, con particolare attenzione a quelle caratteristiche che rendono una rete reale diversa da una random.

5.1 DISTRIBUZIONE DEI GRADI

Il grado di un vertice in una rete é il numero di rami incidenti su questo vertice. Definiamo p_k come la frazione di vertici nella rete che hanno grado k. Equivalentemente, p_k é la probabilitá che un vertice scelto a caso in modo uniforme abbia grado k. In un grafo random del tipo studiato da Erdos e Rényi [29], ogni ramo é presente o assente con la stessa probabilitá, quindi la distribuzione dei gradi é binomiale oppure Poissoniana. Le reti reali sono molto spesso diverse dai grafi random per quanto riguarda la distribuzione dei gradi. Lontani dall'avere una distribuzione di Poisson, i gradi dei vertici in molte reti reali sono altamente asimmetrici a destra, questo significa che la loro distribuzione ha una lunga coda di valori che sono sopra la media.

5.2 TRANSITIVITÁ O CLUSTERING

In molte reti si trova che se il vertice A é connesso al vertice B ed il vertice B al vertice C, c'é un'alta possibilitá che il vertice A sará connesso al vertice C. Nel linguaggio delle reti sociali, l'amico del tuo amico potrebbe anche essere un tuo amico. In termini di topologia, transitivitá significa la presenza di un'alto numero di triangoli nella rete (insiemi di tre vertici ognuno dei quali é connesso agli altri due). Matematicamente questa proprietá puó essere quantificata con il coefficiente di clustering $\mathscr{C}$:

$$\mathscr{C} = \frac{3\times \text{ numero di triangoli nella rete}}{\text{numero di triple connesse di vertici}}$$

dove una tripla connessa é un singolo vertice con rami che vanno su un'altra coppia di vertici. C é quindi la probabilitá media che due vertici aventi uno stesso vicino nella rete, siano essi stessi vicini.

5.3 CENTRALITÁ DI GRADO

L'insieme dei gradi di tutti i nodi di una rete possono essere descritti da un unico indice, detto *centralitá di grado* che calcola la variazione dei gradi dei nodi diviso la variazione del grado massimo possibile in una rete della stessa grandezza.

$$Cd(\mathcal{N}) = \frac{1}{(\mid V \mid -1)(\mid V \mid -2)} \sum_i (Cd_{max} - Cd_i) \tag{1}$$

dove Cd_{max} é il piú grande valore di centralitá di grado nella rete e Cd_i é il grado di centralitá del vertice i.

La centralitá di grado é una misura di dispersione del grado dei nodi, quando vale zero significa che tutti i nodi sono collegati allo stesso modo, se vale 1 significa che un singolo nodo é collegato con tutti gli altri, mentre gli altri nodi della rete sono connessi sono a questo, quindi si ha una grafo a forma di stella.

5.4 BETWEENNESS CENTRALITY

La *betweenness centrality* di un vertice i é il numero dei cammini geodetici tra altri vertici che attraversano i [23].

$$B_i = \sum_{j,k,j \neq k} \frac{n_{jk}(i)}{n_{jk}} \tag{2}$$

dove $n_{jk}(i)$ é il numero cammini minimi tra i vertici j e k che passano attraverso i, e n_{jk} é il numero totale di cammini minimi tra j e k.

É stato mostrato che la betweenness sembra seguire una legge di potenza per molte reti e propongono una classificazione delle reti in due categorie in base all'esponente di questa legge di potenza. La betweenness centrality puó essere vista anche come una misura della resistenza in quanto ci dice quanti cammini geodetici continuano ad esistere dopo la rimozione di un vertice. Si puó definire anche la betweenness centrality per tutto il grafo come

$$B_{\mathscr{N}} = \frac{1}{|V|-1}\sum_i (B_{max} - B_i) \tag{3}$$

dove B_{max} é il massimo valore di betweenness centrality nel grafo.

5.5 RESISTENZA DI RETE

La proprietá di resistenza delle reti dalla rimozione di vertici é strettamente connessa alla distribuzione dei gradi. Se alcuni vertici vengono rimossi dalla rete, la lunghezza media dei cammini tra coppie di vertici aumenterá, fino a giungere al punto in cui coppie di vertici prima connessi non lo saranno piú e quindi la comunicazione tra di essi lungo la rete non sará piú possibile. Il livello di resistenza a questo tipo di rimozione varia da rete a rete. I vertici possono essere rimossi in diversi modi e le reti possono mostrare un livello di resistenza diverso in base alla politica di rimozione. Per esempio, é possibile rimuovere i vertici in modo casuale oppure concentrarsi su classi specifiche di vertici, come quelli con grado piú alto. La resistenza di rete ha una particolare importanza nel campo dell'epidemiologia dove la "rimozione" di vertici dalla rete puó corrispondere, ad esempio, alla vaccinazione di individui contro una malattia. Poiché la vaccinazione non protegge soltano gli individui vaccinati contro la malattia, ma potrebbe impedire la diffusione di quest'ultima anche tra individui collegati con essi, un'attenta analisi dell'efficacia delle diverse strategie di vaccinazione puó apportare vantaggi sostanziali nella sanitá pubblica.

5.6 MIXING PATTERNS

In molte categorie di reti possono essere presenti tipi diversi di vertici e la probabilitá di connessione tra vertici spesso dipende dalla tipologia cui essi appartengono. Per esempio, in una rete alimentare che rappresenta le specie di prede e di predatori in un ecosistema, i vertici rappresentano le piante, gli erbivori e i carnivori. Molti rami collegano le piante agli erbivori e molti altri gli erbivori ai carnivori. Ma ci sono pochi rami che collegano gli erbivori ad altri erbivori oppure i carnivori alle piante. Nelle reti sociali questi tipi di collegamenti selettivi sono chiamati *assortative mixing* e sono stati studiati ampiamente nella letteratura. L'*assortative mixing* puó essere quantificato tramite un "coefficiente di *assortativity*", il quale puó essere definito in diversi modi. Ad esempio Gupta et al. [34] definisce il seguente coefficiente:

$$Q = \frac{\sum_i P(i|j) - 1}{N - 1}$$

Dove $P(i|j)$ é la probabilitá condizionale che un mio vicino sia di tipo j dato che io sono di tipo i.

5.7 CORRELAZIONI DI GRADO

Un caso speciale di *assortative mixing* é il mixing basato sul grado dei vertici, anche comunemente chiamato correlazione di grado: i vertici con grado alto sono collegati preferibilmente con altri vertici di grado alto? Oppure i vertici con grado alto preferiscono collegarsi con quelli di grado basso? Entrambi i comportamenti sono stati osservati nelle reti reali. La correlazione di grado é di particolare interesse in quanto, essendo il grado una proprietá della topologia del grafo, essa puó dare luogo ad alcuni interessanti effetti sulla struttura della rete. In letteratura sono stati proposti diversi modi di quantificare la correlazione di grado [35] [16] [31].

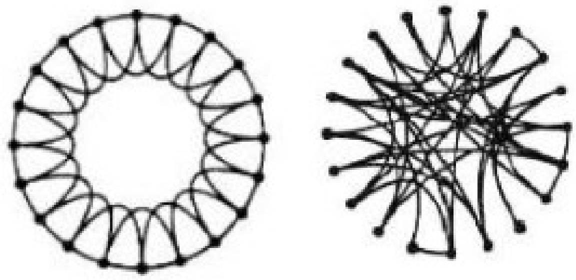

Figura 1: Un esempio di reticolo regolare (a sinistra) e di grafo random (a destra)

6 Modelli di grafo

Nelle sezioni seguenti verrá data una breve descrizione dei modelli di grafo piú diffusi, in particolare nella sezione 6.1 verrano trattati i reticoli regolari, nella sezione 6.2 i grafi random e infine nella sezione 6.3 il modello *piccolo-mondo*.

6.1 RETICOLO REGOLARE

Nella topologia a reticolo regolare i nodi della rete sono posizionati in modo regolare, formando una struttura cristallina, e ognuno di essi é caratterizzato dallo stesso numero di archi incidenti e dunque dallo stesso grado k. La figura 1 mostra un esempio di grafo a struttura regolare con venti nodi e quattro connessioni per nodo.

Essendo le connessioni solo locali, la comunicazione tra nodi distanti puó avvenire solo tramite un numero piuttosto alto di nodi intermedi. Ragion per cui la distanza geodetica media l risulta essere elevata, tanto piú elevata quanto piú grande é il numero dei nodi della rete. Al contrario l'alto numero di connessioni locali fa sì che tale topologia abbia ottime qualitá locali, essendo caratterizzata da un valore elevato del coefficiente di clustering C.

6.2 GRAFO RANDOM

Lo studio dei grafi di tipo random fu introdotto da Erdös e Rényi nel 1959, e attualmente tale topologia di grafo é tra le piú studiate. La costruzione di un grafo random, con N nodi e K archi, parte dalla condizione iniziale di N nodi privi di connessioni. Il grafo é, dunque, generato inserendo un arco tra coppie di nodi scelti in maniera casuale, impedendo connessioni multiple tra la stessa coppia di nodi, ripetendo l'operazione fino a quando il numero di archi raggiunge la quantitá prestabilita K.

Per come é costruito il grafo, a differenza del reticolo regolare i nodi non presenteranno il medesimo grado k. Ció nonostante l'insieme dei valori assunti dal parametro k é limitato e circoscritto ad un piccolo intervallo di valori. Pur differendo tra loro, i nodi hanno comunque un valore del grado pressoché comparabile e nessuno di essi prevale nettamente sugli altri. Per alti valori di N, la distribuzione dei gradi dei nodi é ben approssimata da una distribuzione di Poisson.

I grafi random sono caratterizzati da buone qualitá globali, infatti, la distanza geodetica media l aumenta lentamente all'aumentare di N, e risulta abbastanza piccola anche in reti con un elevato numero di nodi. Di contro, questa topologia di grafo presenta qualitá locali non buone, essendo caratterizzato, infatti, da un basso coefficiente di clustering.

Per informazioni piú dettagliate sulle caratteristiche matematiche dei grafi random si rimanda ai lavori presenti in letteratura. In particolare, per quanto riguarda i classici grafi random Poissoniani [3], [4], [29], [28] e [24]; per quanto riguarda i grafi Markov e quelli esponenziali si rimanda a [32], [27] e [11];

6.3 IL MODELLO *PICCOLO MONDO*

Partendo dal presupposto che molte reti biologiche e sociali presentano caratteristiche intermedie fra il reticolo regolare ed il grafo random, Watts & Strogatz ([13], [14], [18]) hanno sviluppato un nuovo modello di grafo che prende il nome di rete small-world (rete piccolo-mondo). Il loro scopo era introdurre un modello che fosse caratterizzato da un alto coefficiente di clustering e allo stesso tempo presentasse un valore della distanza geodetica media l basso, come essi avevano rilevato negli studi compiuti sulle reti di tipo sociale. Considerando le due topologie introdotte in precedenza, nessuna delle due presenta entrambe le proprietá, il reticolo regolare ha buone qualitá locali (alto coefficiente di clustering), ma una distanza media l elevata. Viceversa, il grafo random é caratterizzato da buone qualitá globali (basso valore di l), ma presenta anche un basso valore del coefficiente di clustering.

Questo tipo di rete é diventata famosa dopo l'esperimento effettuato da Stanley Milgram negli anni '60, che selezionó casualmente un gruppo di americani del Midwest e chiese loro di mandare una lettera ad un estraneo che abitava nel Massachusetts. Ognuno di essi conosceva il nome del destinatario, il suo impiego, e la zona in cui risiedeva, ma non l'indirizzo preciso. Quindi ognuno doveva mandare la lettera ad una persona da loro conosciuta, che a sua volta avrebbe fatto lo stesso, e cosí via, fino a che la lettera non fosse arrivata al destinatario finale.

Il risultato fu che alcune lettere raggiunsero il destinatario stabilito in circa sei passaggi. Questo risultato fu una delle prime dimostrazioni dirette dell' *effetto piccolo mondo*, cioé che molte coppie di vertici in numerose reti sembrano essere connesse da un cammino piccolo che attraversa la rete. Consideriamo un grafo non-orientato e definiamo l la lunghezza media del cammino minimo (in altre parole la distanza geodetica media) tra coppie di nodi nella rete:

$$l = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij} \tag{4}$$

dove d_{ij} é la distanza geodetica (cioé, il percorso piú breve) dal vertice i al vertice j. La quantitá l puó essere misurata per una rete di n vertici ed m rami in tempo $O(mn)$ usando una semplice ricerca in ampiezza [6], anche chiamata ""algoritmo di burning" nella letteratura fisica. La definizione 4 di l é problematica nelle reti che hanno piú di una componente. In questi casi, esistono coppie di vertici che non sono connessi da un cammino. Convenzionalmente si assegna infinito come distanza geodetica ad ognuna di queste coppie, cosicché il valore di l diventa infinito. Per evitare questo problema, su questo tipo di reti l é definito come distanza geodetica media tra tutte le coppie di vertici che sono connesse da un cammino. Un approccio alternativo, e forse piú soddisfacente, é quello di definire l come "media armonica" della distanza geodetica tra tutte le coppie, il reciproco della media dei reciproci:

$$l^{-1} = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij}^{-1}$$

In questo modo i valori infiniti di d_{ij} non contribuiscono alla somma. Negli ultimi anni l'effetto piccolo mondo ha assunto un significato piú preciso in lettertura: le reti presentano questo effetto se, fissato il grado medio dei vertici, il valore di l diminuisce con velocitá logaritmica, o minore, con la dimensione della rete.

La costruzione di un grafo piccolo mondo parte dalla configurazione iniziale di reticolo regolare con N nodi e K archi, quindi si scorrono in maniera ordinata tutti i nodi del grafo e per ogni nodo si considerano tutti gli archi che partono da esso. Con probabilitá $0 \leq p \leq 1$ ogni arco viene tagliato dalla parte del nodo di destinazione (cioé non quello in esame) e riconnesso in maniera random ad un altro nodo (procedura di rewiring), purché ogni coppia di nodi abbia al piú un arco che li colleghi, e nessun nodo sia direttamente collegato a se stesso. In figura 2 é mostrata la

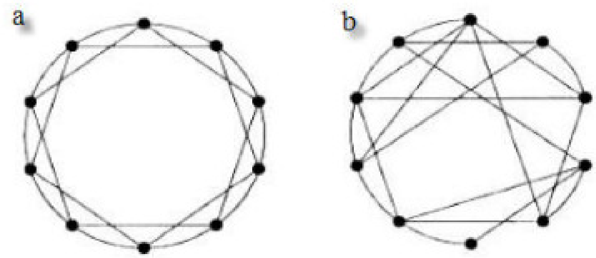

Figura 2: Fasi della costruzione di una rete Small-world a partire da un reticolo regolare (a) di 10 nodi e con una probabilitá di rewiring pari a 0.3 (b)

procedura di costruzione del grafo a partire da un reticolo regolare(a) fino a giungere alla rete small-world(b) con una probabilitá p di rewiring pari a 0.3. I casi limite nella scelta di p, sono $p = 1$ e $p = 0$. Nel primo caso otteniamo un grafo di tipo random, viceversa rimaniamo nell'ambito del reticolo regolare di partenza.

Una notevole proprietá del modello di Watts & Strogatz é che, man mano che p aumenta, la lunghezza media dei percorsi piú brevi (l) passa da valori alti tipici di un reticolo regolare a valori medio-bassi caratteristici di un grafo random e allo stesso tempo il coefficiente di clustering C passa da valori alti a valori bassi. Conseguentemente, a valori intermedi della probabilitá p, il modello é caratterizzato da un basso valore della distanza media e allo stesso tempo da un alto valore del coefficiente di clustering. Si dimostra inoltre che é sufficiente anche il rewiring di un numero relativamente basso di nodi, ossia é sufficiente anche un piccolo valore di p, per garantire che la rete abbia buone qualitá globali e locali (basso l e alto C).

7 CONCLUSIONI

In questo documento é stato introdotto il concetto di rete, si sono descritte le varie tipologie di reti esistenti e le proprietá che le caratterizzano. Inoltre é stata data una panoramica dei vari modelli di grafo che rappresentano diverse tipologie di reti. Questi concetti sono basilari per comprendere le tecniche che verranno utilizzate per la definizione di servizi innovativi.

8 Indice delle figure

Elenco delle figure

9 Indice delle tabelle

Elenco delle tabelle

Riferimenti bibliografici

[1] Adamic L. A. and Huberman B. A. Power-law distribution of the world wide web. *Science*, 287:2115, 2000.

[2] Huberman B. A. *The Laws of the Web*. MIT Press, Cambridge, MA, 2001.

[3] Rapoport A. Contribution to the theory of random and biased nets. *Bulletin of Mathematical Biophysics*, 19:257–277, 1957.

[4] Rapoport A. and Solomonoff R. Connectivity of random nets. *Bulletin of Mathematical Biophysics*, 13:107–117, 1951.

[5] Rapoport A. and Horvath W.J. A study of a large sociogram. *Behavioral Science*, 6:279–291, 1961.

[6] Magnanti T. L. Ahuja R. K. and Orlin J. B. *Network Flows: Theory, Algorithms and Applications*. Prentice Hall, Upper Saddle River, New Jersey, 1993.

[7] Barthelemy M. Amaral L. A. N., Scala A. and Stanley H. E. Classes of small-world networks. *Proc. Natl. Acad. Sci. USA*, 97:11149–11152, 2000.

[8] Moody J. Bearman P. S. and Stovel K. *Chains of affection: The structure of adolescent romantic and sexual networks*. Preprint, Department of Sociology, Columbia University, 2002.

[9] Freeman L. C. Some antecedents of social network analysis. *Connections*, 19:39–42, 1996.

[10] Hyman J.M. Chowell G. and Eubank S. *Analysis of a real world network: The City of Portland, Technical Report BU-1604-M*. Department of Biological Statistics and Computational Biology, Cornell University, 2002.

[11] Strauss D. On a general class of models for interaction. *SIAM Review*, 28:513–527, 1986.

[12] Gardner B. B Davis A. and Gardner M. R. *Deep South*. University of Chicago Press, Chicago, 1941.

[13] Watts D.J. Networks, dynamics and the small world phenomenon. *Am. J. Sociol*, 105:493–592, 1999.

[14] Watts D.J. Small worlds. Technical report, Princeton University Press, Princeton, 1999.

[15] Jones J. H. and Handcock M. S. *An assessment of preferential attachment as a mechanism for human sexual network formation*. Preprint, University of Washington, 2003.

[16] Newman M. E. J. Assortative mixing in networks. *Phys. Rev. Lett.*, 89:208701, 2002.

[17] Scott J. *Social Network Analysis*. A Handbook, Sage Publications, London, 2nd Edition, 2000.

[18] Watts D. J. and Strogatz S. H. Collective dynamics of small-world networks. *Nature*, 393:440–442, 1998.

[19] Padgett J.F. and Ansell C.K. Robust action and the rise of the medici,1400-1434. *Am. J. Sociol.*, 98:1259–1319, 1993.

[20] Moreno J.L. *Who Shall Survive?* Beacon House, Beacon NY, 1934.

[21] Sanwalani V. Kalapala V. K. and Moore C. *The structure of the United States road network*. Preprint, University of New Mexico, 2003.

[22] Egghe L. and Rousseau R. *Introduction to Informetrics*. Elsevier, Amsterdam, 1990.

[23] Freeman L. A set of measures of centrality based upon betweenness. *Sociometry*, 40:35–41, 1977.

[24] Molloy M. and Reed B. A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms*, 6:161–179, 1995.

[25] Morris M. Sexual networks and hiv, aids 97:. *Year in Review*, 11:209–216, 1997.

[26] Mizruchi M.S. *The American Corporate Network,1904-1974*. Sage, Beverley Hills, 1982.

[27] Frank O. and Strauss D. Markov graphs. *J. Amer. Stat. Assoc.*, 81:832–842, 1983.

[28] Erdős P. and Rényi A. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5:17–61, 1960.

[29] Erdos P. and Renyi A. On random graphs. *Publicationes Mathematicae*, 6:290–297, 1959.

[30] Mariolis P. Interlocking directorates and control of corporations: The theory of bank control. *Social Science Quarterly*, 56:425–439, 1975.

[31] Vazquez A. Pastor-Satorras R. and Vespignani A. Dynamical and correlation properties of the internet. *Phys. Rev. Lett.*, 87:258701, 2001.

[32] Holland P.W. and Leinhardt S. An exponential family of probability distributions for directed graphs. *J. Amer. Stat. Assoc.*, 76:33–65, 1981.

[33] Solomonoff R. and Rapoport A. Connectivity of random nets. *Bulletin of Mathematical Biophysics*, 13:107–117, 1951.

[34] Gupta S., Anderson R.M., and May R.M. Networks of sexual contacts: Implications for the pattern of spread of hiv. *AIDS*, 2:807–817, 1989.

[35] Maslov S. and Sneppen K. Specificity and stability in topology of protein networks. *Science*, 296:910–913, 2002.

[36] Latora V. and Marchiori M. Is the boston subway a small-world network? *Physica A*, 314:109–113, 2002.

[37] Faust K. Wasserman S. *Social Network Analysis*. Cambridge University Press, Cambridge, 1994.