CAPITOLO 1

INTRODUZIONE ALLE RETI COMPLESSE

Negli ultimi anni si è compreso che sistemi anche molto diversi tra loro possono essere efficacemente descritti in termini di cosiddetti "networks" o reti complesse. Gli esempi vanno da reti tipo tecnologico, come Internet o il WWW, a reti di tipo biologico come reti metaboliche o proteiche, e perfino di tipo sociale, come ad esempio quelle che rappresentano le collaborazioni in ambito scientifico oppure la struttura delle grandi organizzazioni aziendali. In generale, il concetto di rete e' una schematizzazione molto generale di un sistema costituito da molte entità, dette nodi (che possono quindi rappresentare, nei vari casi, persone, computer, proteine,sostanze chimiche, eccetera) legate tra loro e interagenti mediante connessioni (che possono essere, di conseguenza, un cavo tra computer, un iperlink tra pagine web, una collaborazione tra persone, una reazione tra sostanze chimiche, ecc.). In tutti i casi citati, il comune denominatore è l'esistenza di proprietà topologiche complesse.

Lo studio delle caratteristiche strutturali delle reti complesse reali è di notevole importanza per la comprensione dei meccanismi e delle leggi generali che regolano il passaggio di informazioni, energia o altro attraverso i nodi della rete, ma anche comprendere i fattori di debolezza di una rete, che la rendono vulnerabile a disturbi esterni.

Nel corso del capitolo vedremo com'è possibile modellare la struttura di una rete, attraverso la rappresentazione mediante l'utilizzo dei grafi [1,2,3,4]. Poi saranno introdotte alcune particolari topologie di rete: le reti random e le reti a struttura regolare. Si dimostra che tutte le reti reali presentano una struttura che ha caratteristiche intermedie rispetto tra le due topologie sopra citate. In particolare , introdurremo le reti small-word che, poiché ha caratteristiche strutturali a metà strada tra quelli delle reti random e quella dei reticoli regolari, possono modellare con una buona approssimazione la struttura di alcune reti reali.

Nella parte finale del capitolo faremo riferimento alle reti dinamiche. Quest'ultime sono reti viste come entità dinamiche. Ossia la topologia delle connessioni della rete è soggetta ad evolversi e a adattarsi nel tempo, a causa d'azioni indotte dall'esterno, o a causa delle interazioni reciproche degli elementi interni della rete, o seguendo predefinite direttive di sviluppo. Approfondiremo l'argomento nella parte finale di questo capitolo.

1.1 Definizioni e concetti chiave

Una rete generica può essere rappresentata come un grafo $G = (N, E)$, nel quale N rappresenta l'insieme dei nodi della rete ed E è un insieme di coppie di nodi, dette archi, che rappresentano le interazioni tra i componenti della rete. Ad esempio, In Internet i nodi sono router mentre gli archi sono connessioni fisiche tra loro; nelle reti neurali i nodi sono neuroni e gli archi sono connessioni sinaptiche che li uniscono.

. Generalmente ci si riferisce ad un nodo tramite la posizione i all'interno dell'insieme N. Ciascuna connessione è definita da una coppia di nodi i e j e si denota con (i, j) o l_{ij}. In un *grafo orientato* l'ordine dei due nodi è importante: l_{ij} indica una connessione dal nodo i verso il nodo j, ragion per cui $l_{ij} \neq l_{ji}$. In un *grafo non orientato*, viceversa, l'ordine dei nodi non è importante, dunque l_{ij} indica l'esistenza di un collegamento tra i due nodi i e j. In particolare se esiste un arco che collega due nodi, questi ultimi si definiscono *adiacenti*. Una maniera usuale di disegnare i grafi è di rappresentare i nodi con dei punti, collegando con un tratto due punti se i corrispondenti nodi sono uniti da un arco. La figura 1.1 mostra un esempio di rappresentazione di un *grafo orientato* (a sinistra) e *non orientato* con sette nodi e otto connessioni. Poiché le reti che saranno oggetto della nostra analisi sono modellabili tramite *grafi non orientati,* nel prosieguo della trattazione ci riferiremo solo a tale tipologia di grafo.

Un'importante proprietà di una rete è la raggiungibilità di due nodi differenti. Due nodi non adiacenti, infatti, potrebbero non essere raggiungibili l'uno dall'altro.Un *cammino* (walk) dal nodo i al nodo j è una sequenza di nodi adiacenti che comincia in i e finisce in j. Due nodi, dunque, sono raggiungibili l'uno dall'altro se esiste un cammino che li unisca. La lunghezza del cammino è definita come il numero d'archi attraversati. Un *percorso* (path) è un cammino in cui nessun nodo è visitato più di una volta. Il percorso di lunghezza minima è detto *percorso più breve* (*shortest path)*.

Rappresenteremo una rete attraverso la sua matrice delle *adiacenze A*, una matrice quadrata di dimensione pari al numero N di nodi della rete, il cui elemento generico a_{ij} $(i, j = 1, ..., N)$ è pari a uno se il collegamento l_{ij} esiste, e zero viceversa.

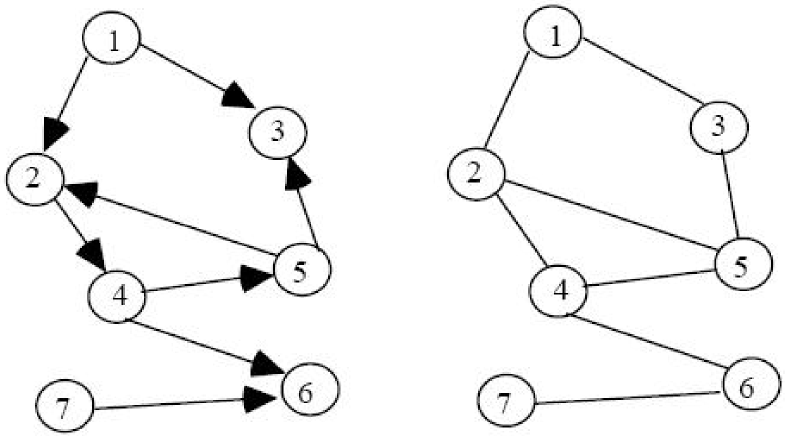

Figura 1.1: rappresentazione grafica di un grafo orientato (a sinistra) e non orientato (a destra) composto da sette nodi e otto archi

Nel nostro caso di studio essendo il grafo non orientato la matrice è simmetrica.

1.1.1 Grado di un nodo

Il *grado* k_i di un nodo *i* è il numero di collegamenti che caratterizzano il nodo, ed si definisce considerando la matrice delle adiacenze *A* come:

$$k_i = \sum_{j \in N} a_{ij} .$$

1.1.2 Shortest path

I percorsi più brevi giocano un ruolo fondamentale per il trasporto d'informazioni all'interno di una rete. Supponiamo che un utente debba inviare un pacchetto da un computer ad un altro della rete Internet: se il pacchetto giunge a destinazione tramite il percorso più breve tra i due nodi, si garantisce un trasferimento più veloce e si minimizza lo sfruttamento di risorse della rete. Si usa rappresentare la lunghezza di tutti i percorsi più brevi di un grafo *G* in un'unica matrice D, in cui ciascun elemento $d_{i\,j}$ è la lunghezza del percorso più breve tra il nodo *i* e *j*. Un'importante proprietà della rete è la *lunghezza media dei percorsi più brevi* (nota anche come distanza media), definita come la media dei percorsi più brevi, calcolata su tutte le coppie di nodi della rete [5,6]:

$$L = \frac{1}{N(N-1)} \sum_{i,j \in N, i \neq j} d_{ij}.$$

E' auspicabile che una rete abbia un valore di L che sia il più basso possibile, poiché ciò implica che ciascuna coppia di nodi è in grado di comunicare con un minor numero di passi.

1.1.3 Coefficiente di clustering

La proprietà di clustering, o transitività, di un grafo misura il grado di conoscenza di una rete, ossia la tendenza di due nodi che sono adiacenti ad un nodo comune di essere connessi l'uno con l'altro. In altre parole la proprietà misura l'importanza di un nodo, infatti, essa si basa sulla valutazione del numero d'archi del grafo qualora il nodo, preso in esame, fosse estratto dalla rete.

Il coefficiente di clustering è una misura di tale proprietà introdotta da Watts e Strogatz [6]. Prima di tutto occorre definire la quantità c_i, il coefficiente di clustering relativo ad un nodo i. Quest'ultimo quantifica la probabilità che $a_{jm} = 1$ per i nodi j e m che sono entrambi adiacenti al nodo i, ed è calcolato contando il numero di archi nel grafo G_i che è il sottografo ottenuto da G:

- Estraendo il nodo i e i nodi connessi ad esso in maniera diretta;
- Eliminando il nodo i e tutti gli archi che insistono su di esso;

l'espressione del coefficiente di clustering è ottenuta dal rapporto del numero di archi in G_i fratto $k_i(k_i - 1)/2$, il numero massimo possibile di archi in G_i [5,6] :

$$c_i = \frac{\sum_{j,m} a_{ij} a_{jm} a_{mi}}{k_i(k_i - 1)}.$$

Il coefficiente di clustering del grafo è quindi dato dalla media dei c_i, su tutti i nodi in G:

$$C = \frac{1}{N} \sum_{i \in N} c_i.$$

Per costruzione, $0 \leq c_i \leq 1$, e $0 \leq C \leq 1$. Un alto valore del coefficiente C indica che sono presenti molte connessioni tra nodi vicini. Al limite, per una rete totalmente connessa C è pari a uno.

1.2 Topologie di rete

In questa sezione introdurremo alcune topologie di rete, e i relativi parametri caratteristici.

1.2.1 Reticolo regolare

Nella topologia a *reticolo regolare* i nodi della rete sono posizionati in modo regolare, formando una struttura cristallina, e ognuno di essi è caratterizzato dallo stesso numero di archi, e dunque dallo stesso grado k. La figura 1.2 (a) 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 può avvenire solo tramite un numero piuttosto alto di nodi intermedi. Ragion per cui la distanza media L risulta essere elevata, tanto più elevata quanto più grande è il numero dei nodi della rete. Al contrario l'alto numero di connessioni locali fa sì che tale topologia abbia ottime qualità locali, essendo caraterizzata da un valore elevato del coefficiente di clustering C.

1.2.2 Grafo random

Lo studio dei grafi di tipo random fu introdotto da Erdös e Rényi nel 1959, e attualmente tale topologia di grafo è tra le più 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 è, 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 d'archi raggiunge la quantità prestabilita K [7].

Per com'è costruito il grafo, i nodi, a differenza del reticolo regolare, non presenteranno il medesimo grado k. Ciò nonostante l'insieme dei valori assunti dal parametro k è 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

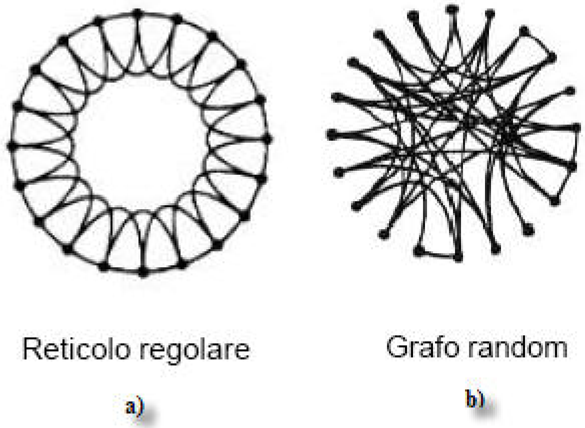

Figura 1.2: Rappresentazione grafica di un grafo a struttura regolare(a) e di un grafo random (b)

essi prevale nettamente sugli altri. Per alti valori di N, la distribuzione dei gradi dei nodi è ben approssimata da una distribuzione di Poisson.

I grafi random sono caratterizzati da buone qualità globali, infatti, la distanza 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 qualità locali non buone, essendo caratterizzato, infatti, da un basso coefficiente di Clustering .

1.2.3 Reti small-word

Partendo dal presupposto che molte reti biologiche e sociali presentano caratteristiche intermedie fra il reticolo ordinato ed il grafo random, Watts & Strogatz hanno sviluppato un nuovo modello di grafo, che prende il nome di *rete small-word* (rete piccolo-mondo) [6]. 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 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 proprietà, il reticolo regolare ha buone qualità locali (basso coefficiente di clustering), ma una distanza media L elevata. Viceversa, il grafo random è caratterizzato da buone qualità globali (basso valore di L),ma presenta anche un basso valore del coefficiente di clustering.

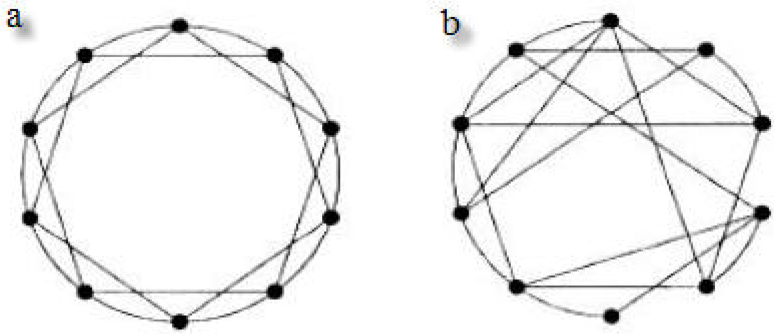

Figura 1.3: Fasi della costruzione di una rete Small-world a partire da un reticolo regolare (a) di 10 nodi e con una probabilità di rewiring pari a =0,3 (b)

La costruzione del grafo 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 si dipartono da esso. Con probabilità $0 \leq p \leq 1$ ogni arco viene tagliato dalla parte del nodo di destinazione (cioè non quello in esame) e riconnesso in maniera random ad un altro nodo (procedura di *rewiring*), purché ogni coppia di nodi abbia al più un arco che li colleghi, e nessun nodo sia direttamente collegato a se stesso. In figura 1.3 è mostrata la procedura di costruzione del grafo a partire da un reticolo regolare(a) fino a giungere alla small-word(b) con una probabilità *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.

Un notevole proprietà del modello di Watts e Srogatz è che, man mano che *p* aumenta, la lunghezza media dei percorsi più 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 probabilità *p*, il modello è caratterizzato da un basso valore della distanza media e allo stesso tempo da un alto valore del coefficiente di clustering. Si dimostra inoltre è sufficiente anche il rewiring di un numero relativamente basso di nodi, ossia è sufficiente anche un piccolo valore di *p*, per garantire che la rete abbia buone qualità globali e locali (basso *L* e alto *C*).

1.3 Reti dinamiche

Di recente è stato sviluppato un nuovo approccio di studio per le reti, che considera queste ultime come entità di tipo dinamico. Ossia la topologia delle connessioni della rete è soggetta ad evolversi e a adattarsi nel tempo, a causa d'azioni indotte dall'esterno, o a causa delle interazioni reciproche degli elementi interni della rete, o seguendo predefinite direttive di sviluppo. Questa tipologia d'analisi nasce dalla necessità di modellare in maniera adeguata alcuni casi reali come ad esempio gli ecosistemi, le reti sociali e biologiche, i mercati finanziari o anche per descrivere problematiche che emergono dallo studio delle connessioni wireless e dei telefonini cellulari.

Una recente ricerca del 2004 di J. Wang e P. De Wilde [8], ad esempio, utilizza un approccio di tipo dinamico per descrivere il processo di diffusione dei virus in Internet tramite e-mail. Molti virus da computer utilizzano il seguente meccanismo per diffondersi nella rete: per prima cosa raggiungono la casella di posta di un utente come allegati ad un messaggio di posta elettronica, quindi, se attivati, sono in grado di replicare se stessi e inviarsi a tutti gli indirizzi di posta contenuti nella rubrica dell'utente attaccato. Per modellare lo scenario descritto viene introdotto un modello di rete dinamica, in cui i nodi sono indirizzi e-mail di utenti, e gli archi sono record della rubrica di ciascun utente contenenti indirizzi di posta elettronica. Man mano che il sistema si evolve nuove connessioni saranno aggiunte e altre vecchie saranno eliminate, simulando il processo periodico di riorganizzazione della rubrica che ogni utente compie, cancellando vecchi indirizzi e-mail e aggiungendone dei nuovi. Attraverso l'inserimento d'opportuni parametri è possibile modellare la fase dinamica di cancellazione e aggiornamento.

Nel corso del prossimo capitolo vedremo come il paradigma delle reti dinamiche potrà essere applicato per la caratterizzazione di modelli di moto collettivo. Questi modelli si applicano a sistemi complessi, ossia aggregati costituiti da un numero elevato di membri, interdipendenti l'uno dall'altro. Lo scopo dei modelli è modellare la dinamica del sistema, il passaggio delle informazioni tra i membri del sistema e le modalità con cui queste informazione sono sfruttate per modificare il comportamento globale della comunità. Vedremo come i modelli di moto collettivo e le reti, viste con l'approccio dinamico, presentino numerosi punti di contatto.