

Reti complesse

Fabio Rinnone

Analisi e Gestione di Dati

Corso di Laurea Magistrale in Informatica

Università degli Studi di Catania

1 Introduzione

Reti biologiche e chimiche, reti neuronali, Internet, il World Wide Web (WWW) costituiscono dei sistemi complessi caratterizzati da un elevato numero di unità dinamiche connesse fra loro. Altre reti si trovano attorno a noi, come ad esempio le reti di individui, o le reti sociali, in cui le unità che le compongono sono individui e i collegamenti tra essi possono essere ad esempio i rapporti d'amicizia che intercorrono tra essi. Altre tipologie di reti, invece, sono effettivamente tangibili, basti pensare alle linee aeree o di autobus o metropolitane, oppure griglie di circuiti elettrici o la stessa Internet. Altre reti ancora, infine, sono astratte e possono, per esempio, essere reti di collaborazione tra individui.

Le reti complesse, o *complex networks*, sono reti la cui struttura è irregolare, complessa e che si evolve nel tempo. Il loro studio prende le basi dalla *teoria dei grafi* e dalla più recente *analisi delle Reti Sociali*. Proprio la teoria dei grafi costituisce la base di partenza per poter studiare nel dettaglio le reti complesse.

2 Struttura delle reti complesse

Seguono alcune definizioni preliminari per comprendere la struttura delle reti complesse.

Definiamo grafo non diretto $G = (\mathcal{N}, \mathcal{L}) = G_{N,K}$ la coppia di due insiemi \mathcal{N} ed \mathcal{L} dove gli elementi di $\mathcal{N} = \{n_1, n_2, \dots, n_N\}$ sono detti nodi o vertici e gli elementi di $\mathcal{L} = \{l_1, l_2, \dots, l_K\}$ sono coppie di elementi di \mathcal{N} e sono

detti archi o *links*. Il numero di elementi dell'insieme dei vertici e dell'insieme degli archi è rispettivamente N e K . Il grafo si definisce diretto se gli elementi di $\mathcal{L} = \{l_1, l_2, \dots, l_K\}$ sono coppie ordinate di elementi di \mathcal{N} . In un grafo non diretto un arco può essere denotato come $(i, j) = l_{ij}$. In un grafo diretto l_{ij} indica invece un arco che collega il nodo i con il nodo j e non viceversa, quindi avremo $l_{ij} \neq l_{ji}$. Definiamo *multigrafo* un grafo che contiene *loop* o *multiarchi*. In particolare, un *loop* è un arco che collega un nodo a se stesso, mentre i *multiarchi* sono archi che collegano medesime coppie di nodi.

Dato un grafo G con N nodi, il numero k di archi può variare da un minimo di 0 archi ad un massimo di $N(N-1)/2$. In quest'ultimo caso tutti i nodi sono adiacenti (due nodi sono detti *adiacenti* se sono collegati fra loro da un arco). Diremo che il grafo è *sparso* se $K \ll N^2$, mentre diremo che è denso se $K = \mathcal{O}(N^2)$. Un grafo si dice completo se $K = \binom{N}{2} = N(N-1)/2$ ed è denotato come K_N . Esempi particolari di grafi completi sono K_2 e K_3 , in particolare quest'ultimo è chiamato *triangolo*.

Dato un grafo $G = (\mathcal{N}, \mathcal{L})$, un suo *sottografo* $G' = (\mathcal{N}', \mathcal{L}')$ è tale che $\mathcal{N}' \subseteq \mathcal{N}$ e $\mathcal{L}' \subseteq \mathcal{L}$ e, se G' contiene tutti gli archi di G è detto *sottografo indotto* da \mathcal{N}' è indicato come $G' = G[\mathcal{N}']$.

Dato un nodo i di un grafo G , il sottografo dei vicini di i , indicato con G_i è il sottografo indotto da \mathcal{N}_i , dove \mathcal{N}_i è l'insieme di tutti i nodi adiacenti ad i .

Un cammino (*walk*) in un grafo G (da un nodo i a un nodo j) è una sequenza di nodi ed archi che comincia con i e termina con j . In generale in un cammino possono gli stessi nodi e gli stessi archi possono essere attraversati più volte. Il cammino di lunghezza minima viene definito cammino minimo, mentre un cammino chiuso, in cui il nodo i e il nodo j coincidono viene definito *ciclo*.

Un ciclo di lunghezza k viene tipicamente indicato con C_k . Un caso particolare di ciclo è C_3 che è un triangolo, dunque vale $C_3 = K_3$.

Un grafo si definisce *connesso* se per ogni coppia di nodi distinti i e j , vi è un cammino da i a j , altrimenti si definisce *non connesso* o *disconnesso*. Una *componente* è, invece, un sottografo indotto connesso. Una *componente gigante*, la cui definizione sarà importante più avanti, è una componente la cui dimensione è dello stesso ordine del numero di nodi del grafo, cioè N .

Un grafo può essere rappresentato mediante un'apposita matrice, detta di adiacenza, indicata tipicamente come \mathcal{A} , di dimensione $N \times N$ tale che:

$$a_{ij} = \begin{cases} 1 & \text{se } \exists l_{ij} \in \mathcal{L} \\ 0 & \text{altrimenti} \end{cases} .$$

In alternativa un grafo può essere rappresentato mediante una matrice di incidenza \mathcal{B} , di dimensione $N \times K$ tale che

$$b_{ik} = \begin{cases} 1 & \text{se il nodo } i \text{ è incidente all'arco } l_{ik} \\ 0 & \text{altrimenti} \end{cases} .$$

Un concetto importante relativo ai grafi e fondamentale nello studio delle proprietà delle reti complesse è il grado. Dato un nodo i di un grafo G , definiamo *grado* del nodo i la quantità:

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

dove a_{ij} sono gli elementi della matrice di adiacenza \mathcal{A} con $i, j = 1, 2, \dots, N$. Il grado di un nodo i è il numero di archi incidenti ad esso.

Tuttavia, se il grafo è diretto, occorre distinguere, per un certo nodo i , il grado entrante k_i^{in} (numero di archi entranti) e il grado uscente k_i^{out} (numero di archi uscenti), che valgono rispettivamente:

$$k_i^{\text{in}} = \sum_{j \in \mathcal{N}} a_{ji}$$

e:

$$k_i^{\text{out}} = \sum_{j \in \mathcal{N}} a_{ij} .$$

Banalmente vale $k_i = k_i^{\text{in}} + k_i^{\text{out}}$.

Indichiamo, invece, con $P(k)$ la probabilità che un certo nodo selezionato casualmente dal grafo abbia grado k . Nel caso di grafi diretti possiamo considerare due distribuzioni di probabilità, per il grado entrante e per il grado uscente, rispettivamente $P(k^{\text{in}})$ e $P(k^{\text{out}})$.

Le informazioni relative a come il grado sia distribuito tra i vertici del grafo possono anche essere ottenute calcolando i momenti della distribuzione.

In generale, in teoria della probabilità, il *momento* (di origine m e ordine k) di una variabile aleatoria è definito come il valore atteso della k -esima potenza dei valori:

$$\mu_{m,k} = \sum_{i=1}^n (x_i - m)^k p_i.$$

dove p_i denota la funzione di probabilità.

Nel caso di un grafo G definiamo il momento n -esimo di G come:

$$\langle k^n \rangle = \sum_k k^n P(k).$$

I momenti più significativi sono il primo ($n = 1$) e il secondo ($n = 2$), in particolar modo per $n = 1$, $\langle k \rangle$ costituisce il grado medio del grafo G .

Dato un nodo i definiamo grado medio dei vicini di i :

$$k_{nn,i} = \frac{1}{k_i} \sum_{j=1}^N a_{ij} k_j.$$

Il grado medio dei vicini di i è invece dato da::

$$k_{nn}(k) = \sum_{k'} k' P(k'|k).$$

dove $P(k'|k)$ è una probabilità condizionata ed in questo caso è la probabilità che un nodo con grado k sia collegato ad un nodo di grado k' .

Un'importanza rilevante, nella teoria dei grafi e nello studio delle reti complesse è la ricerca dei cammini minimi. In generale i cammini minimi in un grafo possono essere rappresentati mediante una matrice \mathcal{D} detta, appunto, matrice dei cammini minimi in cui ogni elemento d_{ij} è la lunghezza del cammino dal nodo i al nodo j . Il valore massimo di d_{ij} della matrice è detto *diametro* del grafo ($\text{Diam}(G)$). Tipicamente una misura utilizzata è la lunghezza del cammino medio data da:

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

Tuttavia è utilizzato spesso anche un approccio alternativo che consiste nel considerare piuttosto la media armonica delle lunghezze dei cammini minimi. In questo caso la misura è chiamata *efficienza* del grafo G ed è data da:

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

Un'altra misura che ha una rilevanza importante per comprendere le caratteristiche dei cammini in un grafo G è la cosiddetta *betweenness*. Questa misura si annovera tra quelle definite di *centralità* e servono a misurare l'importanza di un nodo all'interno della rete secondo determinati parametri. In particolare, dato un nodo i di un grafo G la *betweenness* sarà:

$$b_i = \sum_{j,k \in \mathcal{N}, j \neq k} \frac{n_{jk}(i)}{n_{jk}}.$$

dove n_{jk} è il numero di cammini minimi che collegano i nodi j e k , mentre $n_{jk}(i)$ è il numero di cammini minimi che collegano i nodi j e k che attraversano il nodo i .

Una proprietà importante relativa ai grafi è il cosiddetto *clustering*. Chiamato anche *transitività*, è una proprietà che dipende dal numero di triangoli che sono presenti in un grafo. In particolare, dato un grafo G , la *transitività* è così definita:

$$T = \frac{3 \times \# \text{ di triangoli presenti in } G}{\# \text{ di triple di vertici connessi in } G}.$$

Il fattore 3 nel numeratore è un fattore di normalizzazione che permette di assicurare che la *transitività* sia un indice compreso tra 0 e 1.

Una misura analoga ma alternativa è il *coefficiente di clustering* (locale) relativo a un certo nodo i del grafo. Il coefficiente di clustering è ottenuto contando il numero attuale di archi (indicato con e_i) in G_i (il grafo dei vicini del nodo i). In particolare il coefficiente di clustering è dato dal rapporto fra e_i , e il numero massimo possibile di archi in G_i , cioè $k_i(k_i - 1)/2$:

$$c_i = \frac{2e_i}{k_i(k_i - 1)}.$$

Il coefficiente di clustering del grafo è invece dato da:

$$C = \langle c \rangle = \frac{1}{N} \sum_{i \in \mathcal{N}} c_i.$$

Infine, una misura alternativa che può essere usata per misurare le proprietà di clustering di un grafo G è la cosiddetta *efficienza locale* definita

come:

$$E_{\text{loc}} = \frac{1}{N} \sum_{i \in \mathcal{N}} E(G_i), \quad (2)$$

dove $E(G_i)$ è l'efficienza di G_i così come definita nell'Equazione (1).

Le diverse occorrenze di pattern di interconnessione che occorrono un numero significativo di volte in un grafo G sono chiamati motivi (*motifs*). Il significato statistico di un motivo M , tipicamente viene descritto da un punteggio chiamato *Z-score* che è calcolato nel seguente modo:

$$Z_M = \frac{n_M - \langle n_M^{\text{rand}} \rangle}{\sigma_{n_M}^{\text{rand}}},$$

dove n_M è il numero di volte che il sottografo M appare in G , mentre $\langle n_M^{\text{rand}} \rangle$ e $\sigma_{n_M}^{\text{rand}}$ sono rispettivamente la media e la deviazione standard del numero di volte che il sottografo M compare in una rete casuale (random).

Dato un grafo $G = (\mathcal{N}, \mathcal{L})$ una *comunità* o *cluster* è un sottografo $G' = (\mathcal{N}', \mathcal{L}')$ i cui nodi sono altamente connessi. Una *cricca* è un sottografo completo massimale di tre o più nodi, mentre una *n-cricca* è un sottografo completo massimale dove la più grande distanza fra due nodi non è più grande di n .

Definiamo *spettro* di un grafo l'insieme di autovalori della sua matrice di adiacenza \mathcal{A} . In particolare un grafo $G_{N,K}$ ha N autovalori μ_i ($i = 1, 2, \dots, N$) e N autovettori associati \mathbf{v}_i ($i = 1, 2, \dots, N$). Nel caso in cui i grafi sono molto grandi solitamente si definisce la cosiddetta *densità spettrale* come:

$$\rho(\mu) = \frac{1}{N} \sum_{i=1}^N \delta(\mu - \mu_i).$$

Informazioni importanti sulle proprietà di connettività di un grafo G possono essere anche estratte dallo spettro delle matrici *normale* e *Laplaciana*. In particolare la *matrice normale* è definita come $\mathcal{N} = \mathcal{D}^{-1}\mathcal{A}$, dove \mathcal{D} è la matrice diagonale i cui elementi sono:

$$D_{ii} = \sum_j a_{ij} = k_i$$

e \mathcal{A} è la matrice di adiacenza di G . La *matrice Laplaciana* è invece definita come $\Lambda = \mathcal{D} - \mathcal{A}$. La matrice Λ è una matrice simmetrica semi-

definita positiva, il che significa che i suoi autovalori sono reali e non-negativi. Inoltre la matrice Laplaciana può essere espressa in termini della matrice d'incidenza e della sua trasposta come $\mathcal{D} = \mathcal{B}\mathcal{B}^t$.

Una proprietà interessante della matrice Λ che il primo suo autovalore $\lambda_1 = 0$ è banale e ad esso corrisponde l'autovettore $\mathbf{v}_1 = (1, 1, \dots, 1)$ ed è anche il più piccolo autovalore (poiché la matrice è semi-definita positiva).

Finora abbiamo visto alcune definizioni preliminari riguardanti grafi non pesati. I grafi non pesati sono caratterizzati da una natura binaria, in cui gli archi (tra due nodi) possono essere presenti o meno. Le reti in cui ogni arco ha, invece, associato un valore numerico sono dette reti *pesate*. Nelle reti pesate tale valore numerico misura il *peso* della connessione tra due nodi.

Un grafo pesato $G^W = (\mathcal{N}, \mathcal{L}, \mathcal{W})$ è una tripla dove:

- $\mathcal{N} = \{n_1, n_2, \dots, n_N\}$ è l'insieme dei nodi.
- $\mathcal{L} = \{l_1, l_2, \dots, l_K\}$ è l'insieme degli archi.
- $\mathcal{W} = \{w_1, w_2, \dots, w_K\}$ è l'insieme dei pesi.

Banalmente l'insieme degli archi è l'insieme dei pesi hanno la stessa dimensione K .

Per rappresentare i grafi pesati occorre introdurre un'ulteriore matrice, detta matrice dei pesi, denotata con W e di dimensione $N \times N$ dove ciascun valore w_{ij} è il peso dell'arco che connette il vertice i al vertice j . Banalmente $w_{ij} = 0$ se l'arco che collega i a j non esiste. Senza perdere di generalità tratteremo solo grafi pesati tali che $w_{ij} \geq 0$.

Definiamo peso del nodo (*node strength* o *node weight*) la quantità:

$$s_i = \sum_{j \in \mathcal{N}} w_{ij}.$$

Il peso del nodo è la generalizzazione del concetto di grado.

Il peso medio di un grafo pesato G sarà: $\langle w \rangle$.

Definiamo *disparità* nei pesi del nodo i la quantità:

$$Y_i = \sum_{j \in \mathcal{N}_i} \left[\frac{w_{ij}}{s_i} \right].$$

Denotiamo con $R(s)$ la probabilità che un vertice abbia peso s e con $k_{nn,i}^W$ il grado medio pesato dei vicini del nodo i :

$$k_{nn,i}^W = \frac{1}{s_i} \sum_{j \in \mathcal{N}} a_{ij} w_{ij} k_j.$$

In una rete pesata il minimo numero di archi tra un nodo i e un nodo j non determina necessariamente il cammino minimo per via del fatto che in questo caso interviene, come parametro di misura, il peso associato agli archi. Quindi dato un grafo pesato G definiamo *cammino minimo pesato* d_{ij} il cammino tale che somma dei pesi degli archi fra tutti i possibili cammini tra i e j sia la più piccola.

Definiamo coefficiente di clustering pesato di un nodo i :

$$C_i^W = \frac{1}{s_i(k_i - 1)} \sum_{j,m} \frac{(w_{ij} + w_{im})}{2} a_{ij} a_{jm} a_{mi},$$

dove $s_i(k_i - 1)$ è un fattore di normalizzazione che assicura che: $0 \leq C_i^W \leq 1$ con C_i^W coefficiente di clustering pesato (di tutti i nodi di G).

Definiamo, analogamente a quanto visto nelle reti non pesate, come $C^W(k)$ il coefficiente di clustering pesato di tutti i nodi di grado k .

Dato un grafo G^W definiamo motivo M l'insieme di tutti i sottografi topologicamente equivalenti di G^W . La definizione vale sia che il grafo pesato sia diretto che non diretto.

Dato $g(\mathcal{N}_g, \mathcal{L}_g, \mathcal{W}_g)$ sottografo di G^W definiamo le seguenti quantità:

- Intensità di g :

$$I(g) = \left(\prod_{(i,j) \in \mathcal{L}_g} w_{ij} \right).$$

- Coerenza di g :

$$Q(g) = \frac{I(g)}{\sum_{(i,j) \in \mathcal{L}_g} w_{ij}}.$$

In particolare $I(g)$ rappresenta la media geometrica dei pesi di g .

Analogamente allo *Z-score* per le reti non pesate, possiamo definire un indice per le reti pesate, chiamato *motif intensity score ZI* come:

$$(ZI)_M = \frac{I_M - \langle I_M^{\text{rand}} \rangle}{\sigma_{I_M}^{\text{rand}}},$$

dove I_m è l'intensità di M in G^W , $\langle I_M^{\text{rand}} \rangle$ e σ_M^{rand} sono rispettivamente la media e la deviazione standard delle intensità di un insieme di grafi random.

Definiamo, infine, il *motif-coherence score* ZQ come:

$$(ZQ)_M = \frac{Q_M - \langle Q_M^{\text{rand}} \rangle}{\sigma_{Q_M}^{\text{rand}}}.$$

2.1 Proprietà delle reti reali

Molte reti reali, come ad esempio le reti biologiche o le reti sociali, ma anche Internet e il World Wide Web, sono caratterizzate da un enorme numero di unità interconnesse fra loro. Queste reti godono, inoltre, di numerose proprietà che ne caratterizzano la topologia. Seguono alcune proprietà che contraddistinguono le reti reali.

Una proprietà interessante è la cosiddetta proprietà del piccolo mondo (*small-world*), che prende spunto dalla cosiddetta teoria del piccolo mondo, una branca della teoria dei grafi nata grazie ad una famosa osservazione dello studioso Stanley Milgram conseguenza dell'esame della lunghezza media di un qualunque percorso all'interno di una rete sociale. L'esperimento noto con l'espressione "*sei gradi di separazione*" afferma che una qualunque persona è collegata ad una qualunque altra attraverso una rete di conoscenze e relazioni con non più di 5 intermediari.

Più in generale la proprietà del piccolo mondo è una proprietà delle reti reali per la quale, indipendentemente dalla grandezza della rete, i cammini minimi tra due nodi scelti casualmente sono relativamente brevi (rispetto alla dimensione della rete). In particolar modo la lunghezza media L dei cammini minimi dipende logicamente dalla dimensione N della rete. Questa proprietà è osservata tipicamente nelle reti biologiche e tecnologiche.

2.2 Esempi di reti reali

Vi sono svariati esempi di reti reali osservabili e studiabili. In linea di massima gli esempi più interessanti possono essere suddivisi in tre macro-categorie principali:

- Reti tecnologiche.
- Reti biologiche.

- Reti sociali.

Per quanto riguarda ciascuna di queste categorie, negli anni, sono stati effettuati studi approfonditi sulla topologia di alcuni esempi specifici e sono stati mostrati e comparati i valori dei vari indici definiti attraverso delle vere e proprie prove empiriche.

I parametri che sono stati considerati sono i seguenti:

- N : numero di nodi.
- L : lunghezza media dei cammini minimi.
- $\langle k \rangle$: grado medio.
- C : coefficiente di clustering.

In particolar modo, per quanto riguarda la categoria delle reti tecnologiche sono state comparate le seguenti reti:

- AS2001: la rete dei sistemi autonomi su internet all'anno 2001.
- Routers: la rappresentazione mediante grafo di Internet.
- Gnutella: una rete P2P (Peer to Peer).
- WWW: il World Wide Web, ovvero la rete formata dalle pagine web e dai collegamenti (*hyperlinks*) tra le varie pagine web.

La rete che rappresenta il World Wide Web, caratterizzata da circa 10^8 nodi, si distingue dalle precedenti per essere rappresentata mediante grafo diretto, mentre le altre sono grafi non diretti.

I risultati ottenuti sono che tutte queste reti sono grafi sparsi, caratterizzati da un grado medio basso, godono tutte della proprietà del piccolo mondo e la lunghezza dei cammini minimi è bassa comparata alla dimensione. Tutti i grafi sono sparsi e caratterizzati da elevati valori di C .

Per quanto riguarda le reti biologiche sono stati considerati i seguenti casi:

- Protein: la rete delle interazione proteina-proteina nel lievito.
- Metabolic: la rete delle sostanze chimiche connesse ad altre mediante reazioni metaboliche.

In questo caso solo la rete Metabolic è rappresentata mediante grafo diretto. I risultati osservati, relativamente agli indici considerati, sono molto simili a quelli riscontrati nelle reti tecnologiche, in particolar modo per quanto riguarda la proprietà del mondo piccolo.

Per quanto riguarda infine gli esempi di reti sociali i casi analizzati sono:

- Math1999: il grafo delle collaborazioni tra matematici sulla base degli articoli scientifici in cui figurano come co-autori (anno di riferimento: 1999).
- Actors: la rete di collaborazione (nei film) degli attori basata sui dati dell'IMDB (Internet Movie Database <http://www.imdb.com>).
- e-mail: la rete degli scambi di messaggi di posta elettronica.

In questi casi, dalle osservazioni, è stato riscontrato che le reti sono caratterizzate da piccoli valori di L e da alti coefficienti di clustering.

Gli esempi mostrati finora riguardano reti reali rappresentabili mediante grafi non pesati. Vediamo, adesso, alcuni esempi di reti reali che possono essere rappresentate, invece, mediante grafi pesati.

Gli esempi principali possono essere suddivisi anche in questo caso in tre macro-categorie:

- Reti biologiche.
- Reti sociali.
- Reti tecnologiche.

Per quanto riguarda le reti biologiche due casi particolari sono:

- La rete delle reazioni metaboliche dell'*Escherichia coli*, rappresentate mediante reti pesate in cui w_{ij} rappresenta il flusso di metaboliti da i a j .
- La rete delle comunicazioni tra le cellule del sistema immunitario, rappresentata mediante grafo diretto pesato dove i nodi sono i vari tipi di cellule e i pesi degli archi sono il numero di differenti mediatori solubili secreti dalla cellula i che innescano la cellula j .

Per quanto riguarda le reti sociali, due esempi sono:

- Le reti di collaborazione (*collaboration networks*), di cui due casi interessanti sono i grafi degli attori nei film e i grafi delle collaborazioni scientifiche. Entrambi queste reti sono ad invarianza di scala e godono della proprietà del piccolo mondo. In particolare, nel secondo caso, il peso dell'interazione tra due vertici i e j della rete (dove i nodi sono i collaboratori) è stato definito come:

$$w_{ij} = \sum_p \frac{\delta_i^p \delta_j^p}{(n_p - 1)},$$

dove p è il contatore di tutti gli articoli scientifici,

$$\delta_i^p = \begin{cases} 1 & \text{se l'autore } i \text{ ha contribuito all'articolo} \\ 0 & \text{altrimenti} \end{cases},$$

n_p è il numero di autori dell'articolo p e $n_p - 1$ è un fattore di normalizzazione.

Per quanto riguarda le reti tecnologiche possiamo distinguere due macro-categorie:

- Internet.
- Reti dei trasporti pubblici (di terra, sottoterra o aerei). Un esempio è la rete dei voli diretti tra gli aeroporti i e j . Questa rete è ad invarianza di scala e gode della proprietà del piccolo mondo.

Una proprietà osservata nelle reti tecnologiche è che il coefficiente di clustering pesato $C^W(k)$ ha una variazione più limitata rispetto allo spettro di k il che significa che i nodi con grado elevato tendono a formare delle cricche con nodi del medesimo grado: questa proprietà è chiamata *rich-club phenomenon*.

2.3 Modelli di reti

Un aspetto molto importante nello studio delle reti complesse è lo studio di modelli matematici usati per la generazione di reti che quanto più possibile possano rispecchiare proprietà e caratteristiche delle reti reali. Sono stati, in particolar modo, elaborati svariati modelli che si applicano sia a reti non pesate che a reti pesate.

2.3.1 Reti casuali

Il termine grafo (rete) casuale si riferisce alla natura disordinata degli archi che collegano i vari nodi di un grafo. Una rete casuale, in particolare, è una rete che viene tipicamente generata partendo da N nodi che vengono connessi fra loro mediante archi inseriti casualmente tra coppie di nodi.

Sono stati studiati diversi modelli per la generazione di reti casuali:

Modello ER¹. In questo modello il grafo generato casualmente viene denotato come $G_{N,K}^{ER}$, dove N e K sono rispettivamente il numero di nodi e il numero di archi del grado. Il modello è così strutturato: inizialmente abbiamo N iniziali disconnessi fra loro; successivamente vengono aggiunte casualmente connessioni fra coppie di nodi scelti casualmente fin quando il numero di archi generati non sia uguale a K .

Un approccio alternativo ma simile consiste nel connettere coppie di nodi con una probabilità $0 < p < 1$. In questo caso il numero di archi sarà casuale e il grafo sarà denotato come $G_{N,p}^{ER}$.

Il limite del modello ER e della sua variante probabilistica è tuttavia il fatto che esso non riproduce molte delle proprietà che caratterizzano invece le reti reali.

Nel caso della variante probabilistica del modello particolare attenzione si deve prestare al valore critico che può assumere la probabilità con cui viene generato un arco. È possibile dimostrare che tale valore è:

$$p_c = \frac{1}{N}.$$

In particolare è possibile dimostrare che:

- Se $p < p_c$ il grafo non ha componenti di dimensioni più grandi di $\mathcal{O}(\ln N)$ e nessuna componente ha cicli.
- Se $p = p_c$ la più grande componente ha dimensione $\mathcal{O}(N^{2/3})$.
- Se $p > p_c$ il grafo ha una componente di $\mathcal{O}(N)$ nodi con $\mathcal{O}(N)$ cicli e nessun'altra componente ha più di $\mathcal{O}(\ln N)$ nodi e più di un ciclo.

Per valori grandi di N e valore di grado medio $\langle k \rangle$ fissato, la distribuzione del grado è approssimata dalla distribuzione di Poisson:

¹Erdős and Rényi.

$$P(n) = e^{-\lambda} \frac{\langle k \rangle^n}{n!},$$

per questo motivo i grafi ER vengono anche chiamati *Poisson random graphs*.

Inoltre i grafi ER sono grafi non correlati, ovvero gli archi sono connessi ai nodi indipendentemente dal loro grado, il che significa che $P(k'|k)$ e $k_{nn}(k)$ sono indipendenti da k .

Configuration model. Questo modello, che è una generalizzazione del modello ER, è caratterizzato da una sequenza di gradi fissata. In particolare, dato un grafo G , definiamo sequenza di gradi una sequenza di N numeri interi $D = \{k_1, k_2, \dots, k_N\}$ tali che:

$$\sum_i k_i = 2K,$$

dove K è il numero di archi di G . L'insieme di tutti i grafi con N nodi e data sequenza di gradi D , in cui ogni grafo è considerato con eguale probabilità, è denotato come $G_{N,D}^{\text{conf}}$.

Nel modello la scelta di D viene effettuata in modo che la frazione di vertici con grado k tenderà, per valori grandi di N , alla distribuzione di grado desiderata $P(k)$.

Data, dunque, una sequenza D di gradi fissata opportunamente, le configurazioni casuali del grafo sono generate assegnando ad ogni vertice i un valore, detto *half-edge* (lett. metà arco), pari al valore del grado k_i atteso. Gli archi vengono poi generati accoppiando casualmente, con probabilità uniforme, due *half-edges* assieme.

È possibile dimostrare che se vale:

$$Q = \sum_k k(k-2)P(k) > 0$$

ed inoltre il grado massimo k_{max} del grafo non è troppo grande, il grafo casuale generato mediante il configuration model sarà caratterizzato dalla presenza di una componente gigante. Invece, se $Q < 0$ e k_{max} non è troppo grande, quasi certamente, la dimensione della più grande componente del grafo sarà $\mathcal{O}(k_{max}^2 \ln N)$.

Sia z_m il numero medio dei nodi vicini a distanza m . È possibile dimostrare che se $N \ll z_1$ e $N \ll z_2$ la lunghezza media dei cammini minimi

è:

$$L = \frac{\ln(N/z_1)}{\ln(z_2, z_1)}.$$

Infine il coefficiente di clustering del configuration model è:

$$C = \frac{\langle k \rangle}{N} \left[\frac{\langle k^2 \rangle - \langle k \rangle^2}{\langle k \rangle^2} \right]. \quad (3)$$

Modello WS². Il modello WS è un modello per costruire grafi casuali che godono della proprietà del mondo piccolo. I grafi generati dal modello sono denotati come $G_{N,K}^{WS}$.

Il modello è basato su una procedura di ricollegamento degli archi con probabilità p . Inizialmente abbiamo un anello di N nodi. Ogni nodo dell'anello è connesso simmetricamente ai suoi $2m$ più vicini, in modo che il totale degli archi aggiunti sia $K = mN$. Successivamente per ogni nodo, l'arco collegato al vicino più immediato in senso orario viene agganciato ad un altro nodo scelto con una probabilità p , oppure mantenuto alla stessa posizione con una probabilità $1 - p$.

Si può mostrare che per valori intermedi di p (in particolare per $p \neq 0$ e $p \neq 1$, i grafi generati godono della proprietà del piccolo mondo.

Sono stati elaborati svariati modelli che consentono di costruire reti casuali che rispecchiano la cosiddetta *legge di potenza*. Poiché si parla di network la distribuzione a legge di potenza riguarda il grado.

La legge di potenza, o *power-law* è una funzione del tipo:

$$f(x) = ax^k + o(x^k),$$

dove a, k sono costanti e $o(x^k)$ l'esponente di scala, ovvero una funzione asintoticamente piccola di x^k .

Una distribuzione di probabilità è detta *a legge di potenza* se è della forma:

$$p(x) \propto L(x)x^{-\alpha},$$

dove $L(x)$ è una funzione che “varia lentamente”, α è l'esponente della legge di potenza e inoltre vale:

²Watts and Strogatz.

$$\lim_{x \rightarrow \infty} \frac{L(tx)}{L(x)} = 1$$

con t costante.

Le reti con distribuzione (di grado) a legge di potenza sono dette anche reti ad invarianza di scala (o *scale-free*).

Un caso particolare di reti *scale-free* sono le reti statiche ad invarianza di scala, dove la distribuzione di grado è fissata prima.

Esistono svariati modelli per la generazione di reti random ad invarianza di scala.

Aiello et al. Modello che coinvolge due parametri α e γ . Il modello assegna probabilità uniforme a tutti i grafi random con grado k secondo la legge:

$$N(k) = e^{\alpha} k^{-\gamma}.$$

I grafi generati mediante il modello sono denotati come $G_{\alpha, \gamma}$. I valori N (numero di vertici) e K (numero di archi) sono determinati dai valori α e γ precedentemente fissati. In particolare α è il logaritmo del numero dei nodi con grado $k = 1$, mentre $e^{\alpha/\gamma}$ è il massimo grado del grafo G .

Attraverso l'equazione (3) è possibile calcolare il coefficiente di clustering per le reti con distribuzione di grado a legge di potenza. Il risultato che si ottiene è:

$$C \sim N^{(3\gamma-7)/(\gamma-1)}.$$

Poiché $3\gamma - 7 > 0 \Rightarrow 3\gamma > 7 \Rightarrow \gamma > \frac{7}{3}$ avremo che:

- Se $\gamma > \frac{7}{3} \Rightarrow C \rightarrow 0$.
- Se $\gamma < \frac{7}{3} \Rightarrow C$ aumenta all'aumentare della dimensione del sistema.

Possiamo fare anche delle considerazioni su L , ovvero sulla lunghezza media dei cammini minimi. In particolare:

- Se $\gamma > 3 \Rightarrow L$ cresce come $\log N$.
- Se $2 < \gamma < 3 \Rightarrow L = \mathcal{O}(\log \log N)$ quando il diametro di G è $\mathcal{O}(\log N)$ (ricordiamo che il diametro di G è la distanza massima tra due suoi nodi).

Vediamo altri modelli per la generazione di reti casuali ad invarianza di scala.

Fitness model. In questo modello ad ogni nodo $i \in N$ si assegna un peso (*fitness*) $w_i = i^{-\alpha}$, $i = 1, 2, \dots, N$, $\alpha \in [0, 1)$.

Successivamente si prendono due nodi i, j con una probabilità uguale ai rispettivi pesi normalizzati, ovvero:

- per il nodo i : $\frac{w_i}{\sum_l w_l}$ (per tutti i nodi l di G).
- per il nodo j : $\frac{w_j}{\sum_l w_l}$ (per tutti i nodi l di G).

Il modello prevede che i nodi i e j saranno connessi con un arco se tra loro, ad ogni passo, l'arco non esiste. La procedura di aggiunta degli archi sarà ripetuta più volte finché non saranno stati generati esattamente mN archi, con m fissato.

In definitiva avremo che il grado medio di G sarà: $\langle k \rangle = 2m$.

Osservazioni sul modello:

- Se $\alpha = 0 \Rightarrow$ otterremo un grafo casuale ER.
- Se $\alpha \neq 0 \Rightarrow$ otterremo un grafo con distribuzione di grado a legge di potenza. In particolare: $P(k) \sim k^{-\gamma}$ con $\gamma = 1 + 1/\alpha$ ($\alpha \in [0, 1)$).

Una variante del modello *fitness* è così strutturata: inizialmente abbiamo N nodi isolati. Successivamente associamo ad ogni nodo i un peso (*fitness*) η_i (con $\eta_i \in \mathbb{R}$) preso da una distribuzione di fitness $\rho(\eta)$.

Per ogni coppia di nodi i, j viene costruito un arco con probabilità:

$$f(\eta_i, \eta_j)$$

dove f è una funzione simmetrica dei suoi argomenti (ovvero una funzione che non varia invertendo i suoi argomenti).

Gradient networks. Una *gradient network* è un grafo diretto indotto da gradienti locali di un campo scalare distribuito fra i nodi. (Il gradiente di una funzione è il vettore che ha come componenti le derivate parziali della funzione).

Il modello più semplice per la generazione di *gradient networks* consiste nel considerare anzitutto:

- Una network S , detta *sottostrato*.

- Uno scalare h_i associato ad ogni nodo i che descrive il potenziale del nodo.

La rete è costruita come una collezione di archi diretti che puntano da ciascun nodo al nodo vicino della network (il sottostrato) che ha il più alto potenziale.

Vediamo, adesso, alcuni modelli per la generazione di reti ad invarianza di scala (*scale-free*) evolute, ovvero a reti che rispecchiano maggiormente le proprietà che intercorrono nella maggior parte delle reti reali.

Modello BA³. Modello ispirato al concetto di World Wide Web (WWW). I concetti di base del modello sono:

- Evoluzione.
- Attaccamento preferenziale (*preferential attachment*).

L'idea di base del modello è che nel WWW i siti web con grado elevato acquisiscono nuovi collegamenti (*links*) o punteggi più alti rispetto ai nodi con grado più basso.

I grafi casuali generati mediante il modello BA sono denotati come $G_{N,K}^{BA}$.

Il modello è così strutturato: inizialmente abbiamo m_0 nodi isolati. Successivamente ad ogni passo temporale $t = 1, 2, \dots, N - m_0$ (N è il numero di nodi), un nuovo nodo j con m collegamenti (con $m < m_0$) viene aggiunto alla rete. In questo caso avremo che la probabilità che un arco colleghi il nodo j con un nodo i già esistente sarà linearmente proporzionale al grado attuale del nodo i , ovvero:

$$\prod_{j \rightarrow i} \frac{k_i}{\sum_l k_l}. \quad (4)$$

D'ora in avanti con la notazione dell'equazione (4) denoteremo la probabilità che un arco colleghi il nodo j ad un nodo i della rete: il cosiddetto *attaccamento preferenziale*.

Nel modello BA avremo dunque che all'istante di tempo t il grafo G avrà $N = m_0 + t$ nodi e $K = mt$ archi. Il grado medio del grafo generato sarà $\langle k \rangle = 2m$ ad istanti di tempo abbastanza grandi. Osservazione: il modello BA produce grafi non diretti.

³Barbási-Albert.

Esiste una variante del modello che produce grafi diretti⁴. In questo modello si introduce il concetto di età a di un nodo i inteso come il numero di nodi aggiunti dopo il nodo i . Sia, inoltre, k_i il grado del nodo i e $N_a(k)$ il numero di nodi di età a di grado k . È possibile dimostrare che:

$$N_a(k) = \sqrt{1 - \frac{a}{N}} \left(1 - \sqrt{1 - \frac{a}{N}}\right)^k.$$

Osservazione sui modelli visti fino ad ora: la distanza media dei nodi nel modello BA è minore rispetto ad un grafo casuale ER con medesimi valori di N e K .

Sono state proposte varie generalizzazioni del modello BA, al fine di rendere il modello quanto più in grado di fornire una più realistica rappresentazione di reti reali. Seguono alcune fra queste generalizzazioni.

Modello DMS⁵. In questo modello la probabilità che un arco colleghi un nodo j con un nodo i già esistente è:

$$\prod_{j \rightarrow i} \frac{k_i + k_0}{\sum_l (k_l + k_0)},$$

dove k_0 è una costante tale che $-m < k_0 < \infty$. In questo modello k_0 , in funzione del valore che assume, gioca un ruolo nell'attrattività iniziale del nodo.

Modello di Krapivski et al. In questo modello si assume:

$$\prod_{j \rightarrow i} \frac{k_i^\alpha}{\sum_l k_l^\alpha}.$$

In questo caso, quindi, la probabilità che un arco colleghi un nodo j ad un nodo i esistente è non-lineare. Un modello con un attaccamento preferenziale di questo tipo è importante perché permette di provare che linearità dell'attaccamento preferenziale è una condizione necessaria per ottenere una distribuzione di grado a legge di potenza. Infatti in questo caso otterremo una distribuzione di probabilità esponenziale.

Modello AB⁶. Il modello è una generalizzazione del modello BA. Il perché di questa generalizzazione è dovuto al fatto che il modello originale BA non consente cambi nella topologia della rete dopo che essa è stata genera-

⁴Krapivsky-Redner.

⁵Dorogostev-Mendes-Samukhin.

⁶Albert-Barabási.

ta. Questo limite comporta un problema: reti reali, come il WWW o le reti sociali, sono caratterizzate da cambi continui nel tempo di connessioni tra i nodi.

In questo modello le regole sono quelle che seguono:

1. Ad ogni passo, con probabilità p viene selezionato un nodo esistente e vengono aggiunti m nuovi archi da esso verso m nodi scelti.
2. Successivamente, con probabilità q , vengono riagganciati m archi selezionati arbitrariamente.
3. Con probabilità $1-p-q$ viene aggiunto un nuovo nodo e viene collegato con m nodi già esistenti.

Si può dimostrare che il modello produce una rete con distribuzione di grado a legge di potenza dove $P(k)$ dipende da p e q .

Modello BB⁷. Modello caratterizzato dal seguente attaccamento preferenziale:

$$\prod_{j \rightarrow i} \frac{\eta_i k_i}{\sum_l \eta_l k_l},$$

dove η_i è la *fitness* del nodo i ed è ottenuta da una distribuzione data $\rho(\eta)$.

Modello RB⁸. Il modello RB riproduce molto bene l'organizzazione gerarchica di molte reti reali. Esso combina le proprietà delle reti con distribuzione di probabilità a legge di potenza con un elevato grado di clustering.

Modello HK⁹. Modello pensato per rafforzare le proprietà di clustering delle reti generate. Esso è basato su un metodo di incorporazione di un protocollo di generazione di triangoli (si veda la definizione di clustering) all'interno del modello BA.

Il modello è caratterizzato dai seguenti passi:

1. Si aggiungono nuovi nodi con m archi (analogamente al modello BA).
2. Si connettono i nodi creati o ai vicini di un nodo connesso precedentemente o usando una opportuna regola di attaccamento preferenziale.

⁷Bianconi-Barabási.

⁸Ravasz-Barabási.

⁹Halme-Kim.

Modello di Jin-Girvan-Newman. Modello ispirato alle dinamiche delle reti sociali.

Tali dinamiche sono essenzialmente le seguenti:

- Due individui con amici in comune hanno un'alta probabilità di conoscerne altri.
- Le relazioni tra individui che hanno pochi incontri diminuiscono con il tempo.
- Un individuo ha una bassa capacità nelle relazioni.

Le reti prodotte da questo modello sono caratterizzate da alti coefficienti di clustering.

Modello KE¹⁰. Modello chiamato anche *structured scale-free*.

In questo modello ciascun nodo può trovarsi in due differenti stati: *attivo* o *inattivo*.

Inizialmente avremo un grafo completo di nodi tutti attivi. Successivamente, ad ogni istante di tempo, viene aggiunto un nodo j con m archi in uscita. Ciascuno degli m nodi attivi ricevono quindi in ingresso un arco dal nodo j . A questo punto il nodo j viene attivato mentre uno fra gli m nodi attivi viene disattivato.

Si può calcolare la probabilità \prod_i^{deact} che un nodo i venga disattivato:

$$\prod_i^{\text{deact}} = \frac{1}{k_i + a} \left(\sum_{m \in \mathcal{N}_{\text{act}}} \frac{1}{k_m + a} \right),$$

dove k_i è il grado del nodo i , a è una costante tale che $a > 0$ e N_{act} è l'insieme dei nodi attivi.

La procedura sarà ripetuta tutte le volte che si riterrà opportuno.

Il modello produce una rete con più alto coefficiente di clustering rispetto al modello BA. Tuttavia essa non godrà della proprietà del mondo piccolo.

Inoltre la lunghezza media dei cammini minimi sarà proporzionale a N : $L \sim N$.

Modello SPSK¹¹. Modello basato su tre meccanismi fondamentali:

1. *Duplicazione*: un nodo selezionato arbitrariamente viene duplicato e vengono duplicate anche le sue connessioni.

¹⁰Klemm-Eguiluz.

¹¹Salé-Pastor-Satones-Kepler.

2. *Divergenza*: le connessioni dei nodi duplicati vengono modificate in base ad una certa probabilità δ .
3. *Mutazione*: le connessioni sono aggiunte dai nodi duplicati agli αN nodi che non sono vicini al nodo originale.

Il vantaggio di questo modello è che produce reti a distribuzione di probabilità a legge di potenza, tuttavia il coefficiente di clustering di tali reti non è realistico.

Modello VFMV¹². Il modello ha le stesse caratteristiche del modello SPSK, tuttavia *duplicazione* e *divergenza* sono implementate in modo differente. In particolare la *duplicazione* è siffatta: da un nodo A scelto casualmente viene generato un duplicato A' ; successivamente A viene connesso ad A' con una certa probabilità p . La *divergenza* è così implementata: viene eliminato, con probabilità q , un arco che collega i vicini in comune di A e A' .

Il modello produce delle reti con proprietà più corrispondenti a quelle delle reti reali.

Vediamo adesso alcuni modelli per generare reti pesate. Il modo più semplice di costruire reti pesate consiste nel considerare grafi casuali con una distribuzione di probabilità $P(k)$ data e assumere che i pesi degli archi siano delle variabili casuali indipendenti, determinate da una distribuzione di probabilità dei pesi $Q(w)$. Segue la descrizione dei principali modelli.

Modello YJBT¹³. Semplice modello per la generazione di reti pesate ad invarianza di scala. La topologia del modello è la stessa del modello BA: abbiamo m_0 nodi iniziali; ad ogni istante di tempo viene aggiunto un nodo j con $m \leq m_0$ archi connessi alla rete. La probabilità che un nodo j venga collegato ad un nodo i (attaccamento preferenziale) è la stessa del modello BA. La differenza con il modello BA sta nel fatto che ad ogni arco viene assegnato un peso:

$$w_{ji} = \frac{k_i}{\sum_m k_m}.$$

Per quanto riguarda distribuzione di grado e distribuzione di peso avremo:

- $P(k) \sim k^{-\gamma}$ con $\gamma = 3$.
- $R(s) \sim s^{-\gamma_s}$ con γ_s dipendente da m .

¹²Vásquez-Flammini-Maritan-Vespignani.

¹³Yook-Jeong-Barabási-Tu.

Modello ZTZH¹⁴. Generalizzazione del modello YJBT. Secondo questo modello ad ogni nuovo collegamento l_{ij} si assegna:

- con probabilità p peso:

$$w_{ji} = \frac{k_i}{\sum_l k_l},$$

- con probabilità $1 - p$ peso:

$$w_{ji} = \frac{\eta_i}{\sum_l \eta_l},$$

dove η_i è un parametro di fitness assegnato a ciascun nodo i .
Considerazioni sul modello:

- Se $p = 1$ il modello ZTZH equivale al modello YJBT.
- Se $p = 0$ i pesi sono determinati interamente dalla fitness dei nodi.

Per quanto riguarda la distribuzione di peso avremo che: $R(s) \sim s^{-\gamma_s}$ con γ_s dipendente dalla probabilità p . La distribuzione di grado $P(k)$ è, invece, logaritmica.

Modello AK¹⁵. Modello in cui la crescita strutturale della rete è accoppiata con i pesi degli archi. La procedura di costruzione della rete è la seguente: ad ogni passo un nuovo nodo j è aggiunto e collegato mediante un arco ($m=1$) ad un nodo, detto *target* i con una probabilità che è proporzionale al peso del nodo (*strenght*) di i :

$$\prod_{j \rightarrow i} \frac{s_i}{\sum_l s_l}. \quad (5)$$

Osservazione: in questo modello la probabilità (5) è la corrispondente rispetto ai pesi dell'attaccamento preferenziale adottato nel modello BA per le reti non pesate.

Il modello AK ha un'analogia con le reti reali, in particolare con Internet. Infatti, in Internet, i nuovi router vengono connessi ai router già esistenti, in proporzione alla larghezza di banda disponibile. Un'altra analogia si ha con

¹⁴Zheng-Trimper-Zheng-Hui.

¹⁵Antal-Krapivski.

le reti di collaborazioni, in cui gli autori con più articoli tendono ad avere ulteriori collaborazioni scientifiche rispetto a chi ha pubblicato meno articoli.

I tre ultimi modelli visti, ovvero i modelli YJBT, ZTZH e AK, hanno una caratteristica in comune: i pesi vengono assegnati alla creazione degli archi e rimangono fissi per tutto il corso della creazione della rete.

Vediamo, adesso, modelli in cui, invece, i pesi si evolvono dinamicamente. **Modello BBV**¹⁶. Il modello funziona come segue:

- Primo passo: abbiamo m_0 vertici connessi fra loro con archi con peso assegnato w_0 .
- Ad ogni passo: viene aggiunto un vertice j con m archi avente peso iniziale w_0 . Gli m archi vengono “attaccati” ai vertici i già esistenti in accordo con la distribuzione di probabilità (5) (la medesima del modello AK).

Dopo l’aggiunta del nuovo arco l_{ij} vengono riassegnati i pesi degli archi che collegano i ai suoi vicini L secondo la seguente regola:

$$w_{il} \rightarrow w_{il} + \Delta w_{il}, \quad \forall l \in \mathcal{N}_i,$$

dove \mathcal{N}_i è l’insieme dei vicini di i e:

$$\Delta w_{il} = \delta \frac{w_{ij}}{s_i},$$

con δ un certo parametro.

Senza perdere di generalità è possibile fissare $w_0 = 1$. In questo caso il modello dipenderà esclusivamente dalla scelta del parametro δ .

Avremo le seguenti proprietà:

- $Q(w) \sim w^{-\gamma_w}$ quando $\gamma_w = 2 + 1/\delta$.
- $P(k) \sim k^{-\gamma}$.
- $R(s) \sim s^{-\gamma_s}$ con $\gamma = \gamma_s = (4\delta + 3)/(2\delta + 1)$.

In definitiva il modello produce reti casuali pesate ad invarianza di scala con un esponente $\gamma \in [2, 3]$.

¹⁶Barrat-Barthélemy-Vespignani.

Osservazione: se $\delta = 0$ il modello risulta topologicamente equivalente al modello BA.

Modello DM¹⁷. È un modello simile per risultati ottenuti al modello BBV. In questo modello i nodi con peso (*strenght*) elevato attraggono nuovi archi e, successivamente, i pesi degli archi di questi nodi vengono opportunamente modificati. In particolare archi di peso elevato aumentano progressivamente i loro pesi ed attraggono nuove connessioni.

Il modello è caratterizzato dalle seguenti regole:

- inizialmente abbiamo una configurazione arbitraria di nodi e di archi costruita a partire da un solo arco iniziale di peso 1.
- Ad ogni istante di tempo:
 - viene scelto un arco con probabilità proporzionale al suo peso ed il suo peso viene incrementato di una certa costante δ .
 - Un nuovo vertice l viene aggiunto e viene collegato alle due estremità dell'arco selezionato con due archi di peso 1 come in Figura 1.

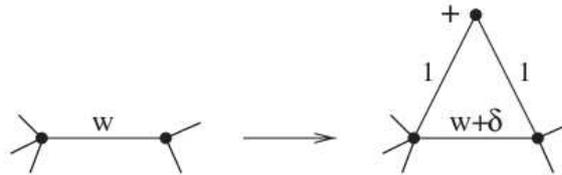


Figura 1: Crescita della rete nel modello DM.

Se $\delta = 0$ il modello è identico al modello DMS.

In questo modello la legge di potenza è rispettata e l'esponente della legge sarà:

$$\gamma_w = 2 + \frac{2}{\delta} \text{ con } \gamma = \gamma_s = 2 + \frac{1}{1 + \delta}.$$

¹⁷Dorogovstev-Mendes.

3 Reti spaziali

Le reti spaziali sono una particolare classe di reti dello spazio reale, ovvero reti i cui nodi occupano una posizione precisa nello spazio Euclideo a due o a tre dimensioni i cui archi rappresentano collegamenti fisici reali.

Alcuni esempi tipici di reti spaziali sono:

- Reti neurali.
- Information/Communication networks.
- Reti degli aeroporti/strade/treni/metropolitane.

Una proprietà fondamentale delle reti spaziali consiste nel fatto che il grado dei nodi è vincolato, poiché il numero di archi che possono essere collegati ad un singolo nodo dipende dallo spazio fisico effettivamente disponibile. Inoltre due nodi distinti hanno meno probabilità di essere collegati a causa del “costo” degli archi dipendente dalla loro distanza. Inoltre nelle reti spaziali la correlazione tra grado e clustering è banale: ad esempio in Internet o nella rete dei collegamenti elettrici negli Stati Uniti si osserva che $C(k)$ è indipendente da k mentre in altre reti si ha che $C(k) \sim k^{-1}$.

Un problema tipico nello studio delle reti spaziali è verificare se in esse è verificata la proprietà del mondo piccolo. In particolare è stato provato che la proprietà non vale nelle reti vincolate fortemente da costanti geografiche. Avremo:

$$N_i(r) \sim \exp(\alpha r),$$

dove $N_i(r)$ è il numero di nodi che possono essere raggiunti da un nodo i in al massimo r passi.

Occorre, dunque, ridefinire il concetto di *small-world* nelle reti spaziali. L’idea di base del concetto di mondo piccolo, come già visto, è la presenza di un numero elevato di connessioni all’interno di singoli cluster e la presenza di poche connessioni globali che connettono fra di loro i vari cluster. Per questo motivo, per analizzare opportunamente le reti spaziali e per definire la proprietà del piccolo mondo, occorre definire una nuova misura specifica per esse. Come esempio esplicativo è stata considerata la rete delle connessioni Internet negli Stati Uniti. Data quindi la topologia della rete (spaziale) l’idea di base è quella di suddividere la rete in regioni, detti cluster locali con un

numero variabile di nodi. A questo punto introduciamo una nuova misura $C_{i(q)}$ definita come la connettività di un nodo i appartenente ad un'area q ovvero il rapporto tra il numero di collegamenti del nodo i verso regioni fuori da q e il numero di link da i verso nodi all'interno della regione q .

Sono stati proposti vari modelli per la generazione di reti ad invarianza di scala nello spazio euclideo. Seguono alcuni di questi modelli.

Xulvi-Brunet-Sokolov. In questo modello ad ogni passo un nuovo nodo j è aggiunto su un reticolo uni-dimensionale con condizioni a contorno periodiche. Il nodo aggiunto viene poi collegato ad un nodo target con probabilità:

$$\prod_{j \rightarrow i} \sim k_i l_{ij}^{-\alpha},$$

dove l_{ij} è la distanza euclidea tra i e j e $\alpha > 0$ è un parametro specifico del modello.

Un reticolo, in algebra, è un sottogruppo discreto di \mathbb{R}^n . Un reticolo uni-dimensionale è caratterizzato da una sola dimensione x , quindi se volessimo calcolare la distanza euclidea tra due punti del reticolo essa sarà una distanza mono-dimensionale calcolata nel seguente modo per due punti $P(p_x)$ e $Q(q_x)$ (dove p_x e q_x sono rispettivamente le uniche coordinate dei punti P e Q):

$$d = \sqrt{(p_x - q_x)^2} = |p_x - q_x|.$$

Nel caso del modello di Xulvi-Brunet-Sokolov, da simulazioni empiriche è stato possibile constatare che se $\alpha < 1$ la distribuzione di probabilità è a potenza reale: $P(k) \sim k^{-3}$ (analogamente al modello BA). Quando $\alpha > 1$ la distribuzione di probabilità è invece esponenziale: $P(k) = a \exp(-bk^\gamma)$. Tuttavia, per quanto riguarda la lunghezza media dei cammini minimi, la proprietà $L \sim \ln N$ vale per ogni valore di α .

Manna-Sen. Modello studiato in un reticolo di D dimensioni. In questo modello la distribuzione di probabilità segue la legge di potenza:

$$q(\ell) \sim \ell^{-\delta},$$

dove δ è calcolato esattamente per tutto il range di valori di α , in particolare:

- $\delta(\alpha) = \alpha - D + 1$ se $\alpha \leq 2D$.

- $\delta(\alpha) = D + 1$ se $\alpha > 2D$.

Vediamo adesso alcune tecniche per la modellazione di reti geografiche reali. Internet, nella fattispecie, è un esempio di rete caratterizzata da una precisa struttura geografica. Un modello che permette la generazione di reti che hanno le proprietà di Internet è basato sui seguenti concetti:

- I router sono distribuiti casualmente nello spazio euclideo.
- La probabilità che due router i, j siano connessi decade rispetto alla loro distanza secondo la seguente legge:

$$P(\ell_{ij}) = \beta \exp(-\ell\alpha D),$$

dove D è la distanza massima fra due nodi, $a > 0$ è un parametro usato per controllare il rapporto fra archi a breve e lunga distanza e $\beta \leq 1$ è un parametro usato per controllare il grado medio della rete.

Un modello basato su questi concetti tuttavia produce una rete statica. Sono possibili delle varianti che consentono la generazione di reti che si evolvono, quindi di reti non statiche. È stato elaborato, ad esempio, un modello in cui il layout fisico dei nodi è un insieme di tipo frattale. In particolare l'aggiunta di nodi è governata da due meccanismi in competizione:

- Attaccamento preferenziale.
- Distanza geografica.

Secondo questo modello un quadrato di dimensione lineare L viene suddiviso in due quadrati piccoli $l \times l$. Ad ognuno di questi quadrati viene assegnata una densità di popolazione $\rho(x, y)$ e una dimensione del frattale D_f . Ad ogni passo un nuovo nodo j viene collocato nella mappa e la sua posizione viene determinata probabilisticamente. In particolare il nodo j viene collocato in posizione (x, y) con probabilità lineare rispetto a $\rho(x, y)$. Il nodo j aggiunto viene poi collegato con m archi a nodi già presenti nel sistema.

La probabilità che un nuovo nodo j sia collegato ad un nodo i con grado k e distanza ℓ_{ij} è:

$$\prod_{j \rightarrow i} \sim k_i^\delta \ell_{ij}^{-\alpha},$$

dove σ è l'esponente che controlla l'attaccamento preferenziale e α è l'esponente che controlla il costo della distanza nodo-nodo. Questo significa che se aumentiamo il valore di σ favoriamo i collegamenti ai nodi di grado elevato, se aumentiamo α favoriamo i collegamenti brevi nodo-nodo.

Vediamo adesso ulteriori modelli.

Modello HOT. Il modello HOT (*Heuristically optimized trade-off*) permette di generare reti che descrivono anche in questo caso Internet. In particolare, in questo modello, Internet è descritta come un albero che si evolve. Ad ogni passo un nuovo nodo j è collocato in posizione casuale in un quadrato di dimensione lineare L e viene poi collegato ad uno dei nodi già presenti. Otterremo:

- Una minimizzazione della distanza euclidea fra i due nodi.
- Un nodo nuovo arrivato presumibilmente verrà collegato ad un nodo collocato al centro della rete.

Più nel dettaglio, un nuovo nodo j sarà connesso ad un nodo i tale che minimizzi la somma dei due obiettivi:

$$i : \min_{i < j} [\alpha l_{ij} + C_i],$$

dove α è un parametro costante, l_{ij} è la distanza euclidea fra i e j e C_i è una qualunque misura di centralità per il nodo i .

Chiaramente la qualità del modello dipende molto dalla scelta del parametro α , ma anche dalla scelta opportuna della misura di centralità. In particolare se $\alpha < \alpha_c$, con α_c un certo valore critico, le distanze euclidee non sono più rilevanti e la rete avrà una topologia a stella. Una stella, in dettaglio, è un grafo bipartito completo senza cicli, quindi un grafo tale che l'insieme dei suoi vertici si può partizionare in due sottoinsiemi tali che ogni vertice di uno di questi sottoinsiemi è collegato solo a vertici dell'altro. In questo caso il grafo è pure completo, quindi esistono tutti gli archi che connettono gli elementi del primo sottoinsieme con gli elementi del secondo.

Se, invece, α cresce veloce almeno quanto \sqrt{N} otterremo una rete con N nodi: il grafo risultante sarà, in questo caso, una versione dinamica del minimo albero di copertura ¹⁸ (*spanning tree*) Euclideo.

¹⁸Albero che contiene tutti i vertici di un grafo. Minimo se la somma dei pesi dei suoi archi è minima.

Infine, se $\alpha > \alpha_c$ e α cresce più lentamente di \sqrt{N} i gradi obbediranno alla legge di potenza.

Infine, un parametro interessante che può essere introdotto e che descrive caratteristiche tipiche delle reti geografiche è la cosiddetta *effettiva lunghezza dell'arco* (i, j) calcolabile come:

$$\lambda\sqrt{N}\ell_{ij} + (1 - \lambda),$$

dove $0 \leq \lambda \leq 1$ è un parametro che serve per definire la preferenza che ha l'utente per misurare la distanza ed è, di fatto, un'unità di misura. Un esempio tipico potrebbe essere il miglio (o il metro) per il calcolo delle distanze automobilistiche o il numero di cambi per quanto riguarda i trasporti aerei.

4 Robustezza statica e dinamica

Il concetto di robustezza si riferisce alla capacità di una rete di evitare malfunzionamenti se una parte di essa si danneggia. Nello specifico possiamo avere due tipi di robustezza:

1. Robustezza statica: atto della cancellazione dei nodi senza bisogno di ridistribuire quantità di informazioni che sono state trasportate nella rete (ad es. reti sociali).
2. Robustezza dinamica: caso in cui è necessaria la ridistribuzione dei flussi di dati nella rete (ad es. Internet in cui al danneggiamento di un router è necessario ridistribuire opportunamente il traffico nella rete).

4.1 Tolleranza statica agli errori

Per tolleranza statica agli errori si intende l'abilità del sistema a mantenere le sue proprietà di connettività anche quando vengono cancellati suoi nodi o archi. La cancellazione casuale di nodi o archi nella rete viene detta *errore*. Si parlerà invece di *attacco* quando ad essere cancellati sono una particolare categoria di nodi o archi, ad esempio i nodi altamente connessi della rete. Il problema della tolleranza statica agli errori può essere trattato mediante la teoria della percolazione.

La percolazione è un modello matematico usato per descrivere in modo statistico le connessioni a lunga distanza di sistemi contenenti numerosi elementi collegati tra loro mediante relazioni aleatorie. Un esempio tipico di modellazione tramite percolazione è il problema della sezione finita del grafo di \mathbb{Z}^2 (percolato bidimensionale) in cui ogni tratto (arco del grafo) è aperto con una certa probabilità p . In particolare il valore della probabilità p influisce sulla generazione del sottografo casuale di \mathbb{Z}^2 ottenuto mediante la cancellazione dei tratti chiusi. Il problema, in questo caso, consiste nel trovare un percorso da un vertice esterno ad un vertice interno x che sia composto unicamente da tratti aperti. Questo problema è noto come “problema dell’esistenza dei cammini aperti”. Se abbiamo un cammino di lunghezza infinita avremo un cosiddetto cluster infinito.

Osservazioni:

- quando p aumenta, aumentano le dimensioni dei cammini aperti.
- Quando p diminuisce i cammini che si formano tendono ad essere finiti.
- Per un certo valore di p , denotato con p_c e detto probabilità critica, esiste un cammino infinito (il valore p_c è variabile e dipende dalla natura del reticolo).

Abbiamo due tipi di percolazione:

- Percolazione dei punti (*site-percolation*).
- Percolazione dei tratti (*bond-percolation*).

Nel primo caso (percolazione dei punti) i vertici del grafo sono dichiarati aperti con probabilità p , nel secondo caso gli archi del grafo sono dichiarati aperti con probabilità p .

Lo studio dei fallimenti nelle complex networks è possibile mediante la teoria standard della percolazione. Possiamo analizzare, in particolare, vari esempi pratici. Consideriamo, ad esempio, Internet e il World Wide Web e supponiamo che dalla rete sia rimossa una frazione di vertici f . Osserveremo che sia in Internet che nel World Wide Web anche elevati tassi di rimozione casuale di nodi, la componente gigante della rete resiste, ammesso che la rimozione avvenga in modalità di attacco. Questo comportamento è tuttavia in contrasto con quanto osservato in grafi casuali. Ad esempio in una rete

casuale con $N = 10.000$ nodi e grado medio $\langle k \rangle = 4$ sono stati osservati comportamenti differenti, sia nel caso di errori che di attacchi. Altre osservazioni empiriche sono state effettuate in reti ad invarianza di scale determinate con i modelli BA e KE, facendo ricorso ai concetti di efficienza (locale e globale) così come definiti nelle Equazioni (1) e (2): i risultati numerici indicano che nelle reti ad invarianza di scala se vengono rimossi al massimo il 2% di nodi casualmente sia l'efficienza locale che quella globale non sono intaccate. Tuttavia se vengono rimossi nodi altamente connessi (attacco) sia l'efficienza locale che quella globale precipitano considerevolmente.

Il problema consiste dunque nello studiare le condizioni che debbano essere verificate affinché in una rete, alla rimozione di una frazione f di nodi o alla rimozione di una frazione f di archi, la componente gigante sia mantenuta anche dopo che si è verificato il danno. Nel primo caso (rimozione dei nodi) si farà ricorso alla percolazione dei punti, nel secondo caso (rimozione degli archi) si farà ricorso alla percolazione degli archi. Il problema deve, inoltre, essere studiato sia nel caso in cui la rete sia correlata sia nel caso che non lo sia. In particolare una rete si dice non correlata se non esiste una correlazione tra i gradi dei vertici, ovvero se consideriamo un vertice x del grafo avente grado k e distribuzione $P(k)$, se scegliamo a caso un nodo vicino, esso avrà un valore di $P(k)$ differente. Una rete si definisce, altresì, correlata se esiste invece una correlazione tra i vertici del grafo.

4.2 Resilienza in reti non correlate

Per *resilienza* si intende la capacità di ripristino di un servizio a seguito di un *disaster recovery*. In particolare la resilienza è la capacità che ha un sistema ad adattarsi alle condizioni d'uso e di resistere all'usura in modo continuare a garantire la disponibilità dei servizi erogati. In altre parole la resilienza rappresenta una sorta di indice di fragilità.

In generale i passi da seguire sono i seguenti:

1. Supponiamo di avere un grafo G con N vertici, grado medio $\langle k \rangle$ e distribuzione di grado $P(k)$. In un primo passo analizziamo le condizioni per cui nel grafo G esista una componente gigante.
2. Eliminiamo dal grafo ed eliminiamo una frazione di nodi (o di archi) e ripetiamo l'analisi: dobbiamo verificare se dopo il danno subito la componente gigante esiste ancora.

Prendiamo il caso delle reti non correlate. In questo caso supponiamo che venga rimosso un vertice scelto casualmente avente grado k con probabilità $P(k)$: la distribuzione di probabilità di uno qualunque dei suoi vicini, essendo la rete non correlata, sarà diversa da $P(k)$, in particolare la distribuzione di uno di tali vertici vicini sarà:

$$q_k = \frac{kP(k)}{\langle k \rangle}.$$

Questo poiché avremo kN_k archi collegati ai vertici di grado k per un totale di $\sum_k kN_k$ archi (N_k è il numero di nodi con grado k). In questo caso se il danno alla rete non danneggia le proprietà *fat-tail* della distribuzione del grado la componente gigante resisterà anche dopo il danno. La distribuzione di probabilità deve quindi mantenersi di tipo *fat-tailed* sia prima che dopo il danno. Questo significa che la distribuzione di probabilità abbia elevata simmetria. In particolare una distribuzione di probabilità si dice simmetrica se gode della proprietà di simmetria rispetto ad un certo valore fissato x_0 .

Successivamente alla cancellazione di una frazione di vertici f , un vertice con grado k si ritroverà collegato ad un certo numero k' di vertici esistenti.

In termini di probabilità $P(k)$ è possibile dimostrarne che la condizione di esistenza della componente gigante in G è data da:

$$\langle k^2 \rangle - 2\langle k \rangle > 0 = \sum_k k(k-2) - P(k).$$

4.3 Resilienza in reti correlate

Prendiamo adesso il caso delle reti correlate. Supponiamo quindi che esista una correlazione nel grado dei vertici adiacenti del grafo G . In questo caso la distribuzione di grado non è un parametro sufficiente, quindi occorre pure tenere conto della probabilità condizionata che un vertice di grado k' sia raggiunto seguendo qualche arco proveniente da un vertice di grado k :

$$P(k'|k) = \frac{E_{k'k}}{\sum_{k'} E_{k'k}} = \frac{E_{k'k}}{kN_k},$$

dove $E_{k'k}$ è la matrice simmetrica che misura il numero di archi tra i vertici di grado k' e k e N_k è il numero di vertici di grado k . La probabilità che un arco preso casualmente connetta due vertici di grado rispettivamente k' e k è:

$$(2 - \delta_{k'k})e_{k'k}$$

con:

$$e_{k'k} = \frac{E_{k'k}}{\sum_k \sum_{k'} E_{k'k}} = \frac{E_{k'k}}{\langle k \rangle N}$$

da cui si ha che:

$$P(k'|k) = \frac{e_{k'k}}{q_k}..$$

Supponiamo quindi che dalla rete vengano rimossi alcuni archi o vertici. Vogliamo, anche in questo caso, verificare l'esistenza della componente gigante anche dopo la rimozione di tali archi e vertici. Nel caso in cui avverrà una rimozione di vertici tratteremo il problema mediante la percolazione dei punti, nell'altro caso mediante la percolazione dei tratti.

Mentre nel caso delle reti non correlate per effettuare l'analisi è sufficiente conoscere il grado k , questo non possibile per le reti correlate. In questo caso occorre lavorare anche con la probabilità condizionata che un vertice di grado k' sia raggiunto seguendo qualche arco che proviene da un vertice di grado k ($P(k'|k)$). Sia $E_{k'k}$ la matrice simmetrica che misura il numero di archi tra vertici di grado k , avremo:

$$P(k'|k) = \frac{E_{k'k}}{kN_k}$$

dove N_k è il numero di vertici di grado k . È possibile dimostrare che $P(k'|k) = e_{k'k}/q_k$.

Supponiamo che siano verificate le condizioni di esistenza della componente gigante nella rete. Supponiamo quindi che dalla rete venga rimossa una frazione di nodi f . Poiché in questo caso la rimozione di un nodo è indipendente dal grado del nodo stesso anziché considerare come parametro la probabilità che un nodo abbia grado k ($P(k)$) considereremo la probabilità che un nodo scelto casualmente abbia grado k e non sia stato rimosso, data da:

$$(1 - f)P(k).$$

Analogamente anziché considerare la probabilità $P(k'|k)$ considereremo la probabilità che se selezioniamo un nodo casualmente e seguiamo uno dei

suoi archi raggiungiamo un nodo di grado k' che non è stato rimosso. Tale probabilità è data da:

$$(1 - f)P(k'|k).$$

Attraverso la definizione di queste probabilità è possibile calcolare le condizioni di persistenza della componente gigante dopo la rimozione della frazione di nodi. Analogo procedimento si può effettuare mediante percolazione dei tratti nel caso di rimozione di una frazione di archi. La conclusione a cui si arriva è che, in generale, nelle reti dove non vi sono correlazioni di grado tra i vertici, dopo il danneggiamento, si potrà constatare che la componente gigante tenda a diventare più piccola, mentre ciò non accade nelle reti correlate. Addirittura non si può escludere che nel primo caso la componente gigante possa anche cancellarsi. In generale quindi le reti correlate risultano essere più resistenti agli attacchi.

4.4 Tolleranza agli attacchi

Il problema da analizzare riguarda in questo caso la tolleranza di una rete non correlata agli attacchi, ovvero a casi in cui dalla rete vengono rimossi nodi altamente connessi fra loro. In generale la rimozione mirata di nodi in reti non correlate ad invarianza di scala è più efficace se paragonata alla rimozione casuale di nodi. Ciò è dovuto all'alto grado di pochi vertici della rete la cui rimozione può danneggiare tutta la rete.

Nelle reti correlate il problema della tolleranza dagli attacchi non è ancora stato esplorato approfonditamente.

4.5 Effetti su reti dinamiche

Fino ad ora abbiamo focalizzato l'attenzione sulle proprietà statiche delle reti, mostrando quali sono gli effetti della rimozione di nodi o archi in esse. Tuttavia, soprattutto per quanto riguarda le reti reali, occorre anche tenere conto delle dinamiche dei flussi di dati che attraversano la rete. In particolare occorre tenere conto di vari parametri, come ad esempio le differenti capacità di trasmissione o il carico variabile. In generale si deve tenere conto di due problematiche principali:

1. La rottura di nodi a valanga (cascata).

2. Il problema della congestione.

Vediamo inizialmente il primo caso. La cancellazione di nodi nelle reti reali, che può essere dovuta a danneggiamenti casuali o attacchi mirati, può provocare problemi nella ri-distribuzione del flusso di dati attraverso la rete, che, in alcuni casi, può non essere tollerato o può risultare essere difficile da gestire. A tal proposito sono stati elaborati vari modelli per gestire queste situazioni.

Fiber bundle model. Questo modello si applica alle reti ad invarianza di scala e, in particolare, consta di un insieme di regole ben precise che si applicano ai nodi. Il modello descrive come l'errore si propaga a cascata sui nodi adiacenti a quelli rimossi dalla rete.

Secondo questo modello il sistema (ovvero la rete) è sottoposta ad una pressione esterna, detta carico (*load*). Il carico è un fattore che permette di prevenire i danni alla rete.

Il modello garantisce, nello specifico, le seguenti proprietà:

- Robustezza della rete in caso di fallimenti ripetuti.
- Possibilità di conoscere in anticipo che il sistema può essere danneggiato.

Più nel dettaglio, in questo modello i carichi sono distribuiti su una rete in accordo con una certa distribuzione. Quando occorre un sovraccarico in alcuni nodi (*overloading*) che causa il danno (*failure*), tali nodi saranno rimossi dalla rete. In questo caso il carico dei nodi cancellati verrà ridistribuito sui nodi adiacenti secondo apposite regole prefissate.

Una variante del modello consiste nel distribuire i carichi agli archi, anziché ai nodi, e di fissare per essi le regole per la loro ri-distribuzione.

Motter-Lai. Anche questo modello è basato sulla ri-distribuzione dei carichi dei nodi eliminati sui nodi adiacenti. In particolare il carico di un certo nodo i ad un istante t di tempo è $b_i(t)$ ovvero la *betweenness* del nodo i . Definiamo inoltre capacità del nodo i , denotata con C_i , il massimo carico che il nodo i può sopportare e assumiamo che:

$$b_i(t) \leq C_i.$$

Se all'istante t' si verifica il sovraccarico il nodo viene cancellato e il carico viene ri-distribuito ai vicini purché si garantisca la condizione $b_i(t') \leq C_i$ per i nodi rimanenti.

Crucitti et al. Variazione del modello precedente in cui anziché eliminare i nodi sovraccaricati, la comunicazione attraverso tali nodi viene solamente degradata. Il modello può essere considerato simile ad un modello che descrive la congestione delle reti.

Tale modello assume che le reti siano pesate e che per ogni arco $i - j$ sia associato un peso $w_{ij} \in (0, 1]$ che misuri l'efficienza della comunicazione attraverso esso. Il nodo sovraccaricato non viene rimosso all'istante t in cui si sovraccarica, ma all'istante $t + 1$ viene solamente aggiornato opportunamente il valore del peso $w_{ij}(t + 1)$.

Caneras et al. Modello utilizzato per descrivere reti elettriche: ogni nodo della rete è caratterizzato da una potenza, il cui valore è positivo per i generatori (di corrente) e negativa per gli utenti. In particolare, per un arco (i, j) il valore del peso w_{ij} rappresenta la potenza del flusso ed un valore z_{ij} rappresenta l'impedenza, ovvero la forza di opposizione al passaggio di corrente elettrica. La potenza del flusso della rete è calcolata mediante equazioni definite opportunamente.

Wang-Xu. Modello relativo ai fallimenti a cascata nei reticoli *coupled-map* (*coupled-map lattices*), ovvero in sistemi dinamici che modellano il comportamento di sistemi non-lineari (tali sistemi sono infatti utilizzati per descrivere reti complesse).

In questo modello l'abbattimento di un singolo nodo può essere sufficiente ad innescare una reazione a catena sull'intera rete se l'ampiezza del danno è superiore ad una certa soglia fissata.

Watts. Il modello di Watts è tipicamente applicato a sistemi sociali ed è usato per risolvere il problema della decisione binaria con esternalità. Supponiamo di avere una rete con N nodi, detti agenti. Ogni agente osserva lo stato dei suoi k vicini che è un valore binario: 0 o 1. L'agente che osserva adotterà lo stato 1 se la frazione Θ dei suoi vicini osservati avrà stato 1. Il sistema funziona nel seguente modo: la configurazione iniziale è data dall'intera popolazione di agenti con inizialmente stato 0. Ad un certo istante una piccola frazione di vertici $N_1 \ll N$ viene perturbata e passa allo stato 1. A questo punto i vertici vicini, negli istanti successivi, cambieranno il loro stato secondo la regola descritta prima.

5 Congestione nelle reti di comunicazione

Sono stati elaborati vari modelli che descrivono le congestioni nelle reti di comunicazione. Ne elenchiamo alcuni.

Ohira-Sawatari. Il modello di Ohira-Sawatari è un modello di base che è punto di partenza per gran parte dei successivi modelli descritti. Il modello si applica al traffico di rete e alla trasmissione dei pacchetti attraverso essa. In questo modello i nodi si suddividono in due classi:

- Host: nodi che inviano e ricevono messaggi.
- Router: nodi che smistano opportunamente i messaggi attraverso la rete.

Ad ogni istante di tempo vengono generati p pacchetti. A ciascun pacchetto viene assegnato un nodo s sorgente (*source*) e un nodo t destinazione (*target*). Si suppone che i router siano in grado di mettere in coda un numero illimitato di pacchetti. I pacchetti vengono accodati su ogni router mediante approccio FIFO (*First In First Out*). Ogni router decide qual è il collegamento migliore (arco) su cui immettere il pacchetto affinché raggiunga la destinazione. In particolare la strategia usata consiste nell'utilizzare un algoritmo standard per la ricerca dei cammini minimi in un grafo. Nel caso in cui un router abbia a disposizione più di una scelta possono essere usate due strategie differenti:

- Strategia deterministica.
- Strategia probabilistica.

Nel primo caso il nodo meno congestionato, ovvero il nodo con il minor numero di pacchetti che sono stati inoltrati fino a quel momento, viene scelto come nodo a cui inoltrare il pacchetto.

Nel secondo caso viene scelto come nodo a cui inoltrare il pacchetto quello con la maggiore probabilità di essere più congestionato.

Esistono generalizzazioni del modello caratterizzate dalle seguenti proprietà:

- Locazione casuale di host e router.

- Incorporazione di un punteggio per la creazione del pacchetto, dipendente dal tasso di congestione del nodo.

Una variante del modello di Ohira-Sawatari consiste nell'utilizzare un approccio LILO (*Last In Last Out*) per accodare i pacchetti nei vari router.

Modelli RD/CS. Il modello RD (*Random Diffusion*) è caratterizzato dalla seguente proprietà: il pacchetto viene inviato dal nodo corrente i ad uno dei suoi vicini con eguale probabilità. Nel modello CS (*Cyclic Search*) viene utilizzato un meccanismo di ricerca locale. In particolare, ad ogni istante di tempo, viene cercato i nodi vicini ed i nodi vicini successivi di un certo nodo i . Se tra questi nodi è presente la destinazione allora il pacchetto viene inoltrato attraverso il cammino che porta ad esso, altrimenti viene adottato il modello RD.

Vi sono, inoltre, una collezione di modelli in cui i pacchetti seguono i cammini minimi dal nodo sorgente s al nodo destinazione t . In questa classe di modelli, ad ogni istante di tempo, ogni pacchetto si muove dalla sua attuale posizione i al nodo successivo nel cammino che porta a j con una probabilità q_{ij} (detta qualità del canale). La qualità del canale è media geometrica delle capacità dei due nodi del canale:

$$q_{ij} = \sqrt{q_i q_j}.$$

Se $q_{ij} = 0$ il canale è disabilitato, se $q_{ij} \simeq 1$ il pacchetto si muove rapidamente e se $q_{ij} \simeq 0$ il pacchetto si muove lentamente.

Echenique et al. In questo modello si suppone che un nodo i abbia un pacchetto la cui destinazione è t . Ad ogni istante di tempo il pacchetto viene spostato dal nodo i al nodo vicino j tale che minimizzi la distanza tra j e t .

6 Processi di diffusione

Gli automi cellulari sono sistemi in cui ogni nodo della rete rappresenta un agente che può trovarsi in un numero finito di stati. Il tempo è discreto e ad ogni istante di tempo lo stato successivo di ogni agente è computato come una funzione del suo stato e dello stato dei suoi vicini nella rete.

Distinguiamo due categorie di modelli di diffusione:

- Modelli di diffusione delle epidemie.

- Modelli di diffusione del rumore.

I modelli di diffusione delle epidemie sono modelli che descrivono come avviene la diffusione di una malattia in una popolazione e ne riproducono le dinamiche, cercando di disegnare anche delle strategie per controllarne o annientarne la diffusione.

I modelli di diffusione del rumore sono modelli che descrivono come avviene la diffusione del rumore nel modo più efficiente possibile, ma non come cancellarlo totalmente. I modelli di questo tipo si applicano tipicamente alle reti di calcolatori e consistono nel disegnare regole che permettono alla rete di trasmettere correttamente le informazioni nonostante la presenza del rumore.

6.1 Diffusione delle epidemie

Esistono due classi di modelli che descrivono la diffusione di epidemie tra individui di una popolazione:

- Modello SIR: *Susceptible-infected-removed*.
- Modello SIS: *Susceptible-infected-susceptible*.

Il modello SIR descrive le malattie di una popolazione assumendo che ogni individuo possa trovarsi in tre stati differenti:

- *Susceptible* (suscettibile): persona che può contrarre la malattia.
- *Infected* (infetto): persona che ha contratto la malattia e che può trasmetterla.
- *Removed* (rimosso): persona che è guarita dalla malattia e che o ne è diventata immune o è deceduta.

Introduciamo due parametri fondamentali: il tasso di trasmissione (della malattia) λ ed il tasso di recupero μ (dalla malattia).

Nel modello SIR diremo che un individuo j della popolazione trasmette la malattia all'individuo i con un tasso di trasmissione λ e lo indicheremo con:

$$S(i) + I(j) \xrightarrow{\lambda} I(i) + I(j).$$

Diremo, inoltre, che un individuo i guarisce con un tasso di recupero μ e lo indicheremo con:

$$I(i) \xrightarrow{\lambda} R(i).$$

Poiché nei casi reali non è detto che da una malattia, dopo la guarigione, si diventi immuni (o eventualmente si deceda), ma può accadere che alla guarigione si ridiventi suscettibili alla stessa malattia si introduce il modello SIS, in cui, stavolta, non è contemplato lo stato *removed*. In questo caso la guarigione di un individuo è descritta dalla seguente relazione:

$$I(i) \xrightarrow{\lambda} S(i).$$

Nel modello SIR dunque è contemplata l'ipotesi che la malattia possa essere debellata, mentre nel modello SIS no.

Il rapporto

$$\sigma = \frac{\lambda}{\mu}$$

descrive come l'epidemia possa diffondersi in una popolazione a partire, tipicamente, da un singolo individuo infetto. Indicheremo con σ_c il valore critico della transazione epidemica. In particolare avremo che:

- Se $\sigma < \sigma_c$ allora sarà impossibile avere un'epidemia infinita.
- Se $\sigma > \sigma_c$ allora un'epidemia infinita sarà possibile con una certa probabilità.

Nella costruzione di modelli che descrivono la diffusione di epidemie, tipicamente si suppone che gli individui con i quali un certo individuo suscettibile entra in contatto siano scelti casualmente dalla popolazione. In una prima approssimazione, tuttavia, supponiamo che tale ipotesi, detta di *omogeneità*, non valga. Eliminare questa ipotesi è conveniente perché consente di descrivere i modelli in termini di equazioni differenziali ordinarie. In questo caso il sistema può essere interpretato nel seguente modo: gli individui infetti decadono nella classe degli individui rimossi con tasso μ , mentre gli individui suscettibili diventano infetti con un tasso proporzionale alla densità di individui sia infetti che suscettibili. Naturalmente si deve supporre che i valori λ e μ siano costanti fisse. In un modello costruito con queste assunzioni

vi sarà una soglia critica anche per il valore di λ e la indicheremo con λ_c . In questo caso avremo che:

- Se $\lambda > \lambda_c$ allora la malattia infetterà una frazione finita della popolazione.
- Se $\lambda < \lambda_c$ allora il totale di individui infetti (incidenza epidemica) sarà infinitesimamente piccolo nel limite di grandi popolazioni.

6.2 Diffusione in reti non correlate

Assumere che l'ipotesi di omogeneità valga è tuttavia importante per capire la diffusione delle epidemie: ad esempio molte malattie sessualmente trasmissibili non potrebbero essere comprese senza la conoscenza delle strutture spaziali della loro trasmissione.

Il problema della diffusione infinita dell'epidemia in una data rete e a partire da un singolo individuo infetto può essere descritta, nel modello SIR, mediante la teoria della percolazione. In particolare la probabilità che occorra un'epidemia infinita sarà uguale alla possibilità che il nodo originariamente infetto appartenga ad un cluster infinito. Grafi eterogenei con distribuzione di grado asimmetrica sono particolarmente importanti per descrivere casi reali di trasmissione di malattie. Per esempio, nel caso delle malattie sessualmente trasmissibili, la distribuzione di grado segue una legge di potenza. Per questo motivo si considera tipicamente il modello SIR in reti eterogenee non correlate con una generica distribuzione di grado $P(k)$ e una connettività media $\langle k \rangle$ finita.

Le equazioni differenziali che descrivono il modello SIR sono espresse in funzione dei seguenti parametri:

- $s(t)$: densità (frazione) di nodi suscettibili all'istante t .
- $\rho(t)$: densità (frazione) di nodi infetti all'istante t .
- $r(t)$: densità (frazione) di nodi rimossi all'istante t .

Questi parametri, in una rete con grado k , li indicheremo come $s_k(t)$, $\rho_k(t)$ e $r_k(t)$.

Ad ogni intervallo di tempo dovrà valere la condizione di normalizzazione:

$$s_k(t) + \rho_k(t) + r_k(t) = 1.$$

In un successivo passo occorrerà studiare come i parametri $s_k(t)$, $\rho_k(t)$ e $r_k(t)$ si evolvono con il passare del tempo. Alla fine otterremo che $\Theta(t)$ sarà la probabilità che un dato link della rete punti ad un nodo infetto e il suo valore sarà dato da:

$$\Theta(t) = \frac{\sum_k k P(k) \rho_k(t)}{\langle k \rangle}.$$

Si può inoltre dimostrare che la soglia epidemica λ_c vale:

$$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle}.$$

In particolare la soglia epidemica è inversamente proporzionale alle fluttuazioni della connettività $\langle k^2 \rangle$. Se $\langle k^2 \rangle < \infty$ allora la soglia epidemica sarà a valore finito.

È stato dimostrata l'assenza di soglie epidemica in reti ad invarianza di scala nel modello SIS. Non è tuttavia così per le reti casali, in cui, invece, questo valore è finito.

È possibile anche dimostrare l'esistenza di strategie di immunizzazione in reti ad invarianza di scala: in particolare si può mostrare che tali reti non acquisiscono immunità globale se la densità dei nodi immunizzati casualmente è non realisticamente alta. Invece se l'immunizzazione dei nodi è mirata, quindi non casuale, oltre che derivabile dalla connettività dei nodi, avremo una soglia epidemica finita e, potenzialmente, sarà possibile immunizzare totalmente la rete.

Una strategia standard per il controllo della diffusione di epidemie in reti non correlate è la vaccinazione di individui scelti casualmente: questo poiché la probabilità di raggiungere un particolare nodo seguendo archi scelti casualmente è proporzionale al grado dei nodi.

Sono stati riscontrati dei casi in cui reti costruite con modelli *scale-free*, in cui il secondo momento della distribuzione del grado $\langle k^2 \rangle$ diverge, che presentano una soglia epidemica finita: in questi casi le reti sono anche caratterizzate da un alto coefficiente di clustering. È dimostrabile che un elevato coefficiente di clustering previene la diffusione delle epidemie.

6.3 Diffusione su reti correlate

Per quanto riguarda le reti correlate è stato mostrato, nel modello SIR, che le proprietà sono le medesime di quelle riscontrate nel caso delle reti non correlate, con l'unica differenza che l'incidenza epidemica risulta più bassa.

Nel modello SIS invece si ha:

$$\lambda_c = \frac{1}{\Lambda'_{\max}},$$

dove Λ'_{\max} è il più grande autovalore della matrice:

$$C'_{kk'} = kP(k'|k).$$

6.4 Malattie cicliche

Le malattie cicliche sono malattie i cui focolai si manifestano ciclicamente nel tempo. Esempi tipici sono il vaiolo, la varicella, il morbillo. Questi particolari tipi di malattie vanno trattati diversamente in quanto sono più complesse sia per quanto riguarda la loro dinamica che per quanto riguarda la topologia.

Un esempio di modello che descrive le malattie cicliche è caratterizzato da una rete rappresentata mediante un reticolo bidimensionale a cui sono associati due parametri:

- u_{ij} detto attivatore.
- v_{ij} detto inibitore.

I due parametri sono applicati a ciascun arco da i a j . Suddetti parametri si associano poi all'equazione standard di Fitz-Hugh-Nagumo (FHN). Il modello FHN, in particolare, è un modello che descrive la depolarizzazione (diminuizione del valore assoluto del potenziale) della membrana cellulare ed è descritto da un particolare sistema differenziale di due equazioni.

7 Diffusione del rumore

Finora abbiamo visto come l'inclusione di topologie complesse nei modelli che descrivono la diffusione delle epidemie cambiano notevolmente i risultati che si otterrebbero altrimenti su grafi regolari o casuali. Ora ci porremo

come obiettivo quello di produrre dei modelli basati su quelli visti relativi alla diffusione delle epidemie, ma che permettono, ad esempio, di controllare la diffusione dei virus tra computer.

La differenza tra questi modelli e quelli relativi alle malattie sta nel fatto che chi progetta le regole è libero di disegnarle in modo tale da poter ottenere i risultati che desidera, anziché dover studiare un modello basato su un processo reale già esistente. Modelli progettati in questo modo si dicono basati sul *rumore*.

Il modello più significativo è il modello DK (Daley/Kendal). Esso è caratterizzato da un grafo di N elementi (individui). Ognuno di questi elementi può trovarsi in uno dei seguenti stati:

- I: ignorante (*ignorant*), ovvero individuo che non ha sentito il rumore.
- S: propagato (*spreader*), ovvero individuo che ha propagato il rumore.
- R: soffocato (*stifler*), ovvero individuo che ha sentito il rumore ma non l'ha più propagato.

Occorre fare un'ulteriore osservazione circa le differenze che intercorrono tra questo modello e i modelli standard per la diffusione delle malattie: in quest'ultimi il cambiamento di stato (ad esempio il passaggio da infetto a suscettibile) non dipende dallo stato corrente, in un dato istante di tempo, dei suoi vicini.

Le regole di transizione nei modelli basati sul rumore sono le seguenti:

$$\begin{aligned} I(i) + S(j) &\xrightarrow{\lambda} S(i) + S(j) \\ S(i) + S(j) &\xrightarrow{\alpha} R(i) + S(j) \\ S(i) + R(j) &\xrightarrow{\alpha} R(i) + R(j) \end{aligned}$$

dove i e j sono due nodi vicini. Dalle regole di transizione si evince che la diffusione del rumore avviene tramite il contatto tra nodi propagati e nodi ignoranti. I parametri λ e α sono rispettivamente il tasso di trasmissione e la probabilità che i nodi *spreader* diventino *stifler*.

Se supponiamo che valga l'ipotesi di omogeneità, il modello DK sarà descritto tramite equazioni differenziali in funzione del tempo: $i(t), s(t), r(t)$. Tuttavia, nei casi di reti con topologie molto complesse, per ottenere risultati

credibili sono necessarie simulazioni numeriche (ad esempio mediante metodi Monte Carlo).

8 Sincronizzazione e dinamiche collettive

8.1 Introduzione

La sincronizzazione è il processo per cui molti sistemi adeguano una certa loro proprietà a causa di condizioni esterne. La forma più semplice di sincronizzazione è quella completa che consiste in un totale agganciamento delle traiettorie di sistemi caotici nel corso del tempo.

L'origine del termine *sincronizzazione* risiede in una parola del greco antico e significa “condividere lo stesso tempo”. Lo studio dei fenomeni di sincronizzazione risale già al XVII secolo quando Huygens osservò per primo che due orologi a pendolo, debolmente accoppiati per mezzo di una trave in comune, si evolvevano nel tempo con la stessa fase. Ultimamente il fenomeno della sincronizzazione è stato osservato in molti fenomeni fisici, quali il sistema di illuminazione delle lucciole, la dinamica dei neuroni, etc.

Il concetto di sincronizzazione si applica ai sistemi dinamici, di cui le reti complesse sono una classe specifica. I sistemi dinamici sono dei modelli di evoluzione deterministici, di cui le reti complesse sono una classe specifica. Essi sono caratterizzati da:

- Uno spazio delle fasi, dove i punti rappresentano gli stati possibili del sistema.
- Una legge di evoluzione, cioè una relazione matematica che determina come si evolve il sistema.

Ad un istante di tempo fissato t il sistema si troverà in un certo stato e l'evoluzione del sistema sarà rappresentata da un punto che si muove nello spazio delle fasi. Il punto, in questo caso, seguirà delle traiettorie ottenibili mediante la risoluzione di un apposito sistema di equazioni. Chiaramente, se le equazioni del sistema dovessero essere non-lineari, la soluzione non sarà ottenibile analiticamente.

Nel caso della sincronizzazione fra due sistemi si parlerà di accoppiamento. L'accoppiamento fra due sistemi può essere di due tipi:

- Unidirezionale: in questo caso il sistema globale è formato da due sistemi che realizzano una configurazione di tipo *master-slave*. In questo caso il sistema master si evolve liberamente e condizionerà il comportamento del sistema slave.
- Bidirezionale: in questo caso i sottoinsiemi sono mutualmente accoppiati tra loro ed entrambi aggiusteranno reciprocamente i loro ritmi.

Supponiamo di considerare un sistema caotico. Per sistema caotico s'intende un sistema dinamico avente le seguenti caratteristiche:

- Dipendenza sensibile alle condizioni iniziali, ovvero a variazioni infinitesime delle condizioni iniziali che corrispondono a variazioni finite in uscita (il cosiddetto *effetto farfalla*).¹⁹
- Imprevedibilità: non si può prevedere in anticipo l'andamento del sistema su lunghi tempi rapportati al tempo caratteristico del sistema.
- L'evoluzione nello spazio delle fasi traccia una figura molto complessa, tipicamente un frattale.

In generale un sistema caotico non è asintoticamente stabile, ha traiettorie strettamente dipendenti dalle condizioni iniziali che diventano molto rapidamente non correlate. Due sistemi caotici si sincronizzano quando le traiettorie di uno dei due sistemi convergono ai valori dell'altro sistema.

8.2 Master-Stability-Function

L'approccio *Master-Stability-Function* (MSF) consente di ottenere la sincronizzazione fra sistemi dinamici complessi. Originariamente pensato per l'accoppiamento di traiettorie di oscillatori, è stato successivamente esteso al caso delle reti complesse di sistemi dinamici accoppiate con topologie arbitrarie.

La MSF è il miglior strumento per analizzare la stabilità della sincronizzazione fra sistemi. Secondo questo approccio un singolo sistema è descritto

¹⁹Principio anticipato da Turing: «Lo spostamento di un singolo elettrone per un miliardesimo di centimetro, a un momento dato, potrebbe significare la differenza tra due avvenimenti molto diversi, come l'uccisione di un uomo un anno dopo, a causa di una valanga, o la sua salvezza.» *Macchine calcolatrici ed intelligenza*, 1950)

da un'equazione del tipo $\dot{\mathbf{x}} = F(\mathbf{x})$ dove \mathbf{x} è un vettore di componenti e F una funzione generalmente non lineare.

L'approccio consiste nell'accoppiare N sistemi di questo tipo. L'equazione che descriverà ciascuno di questi sistemi sarà:

$$\dot{\mathbf{x}}_i = F(x_i) - c \sum_{j=1}^{N_c} a_{ij} \Gamma x_j, \quad i = 1, 2, \dots, N_c$$

dove $\mathbf{x}_i \in \mathbb{R}^n$ è il vettore di stato del sistema i , $c > 0$ è uno scalare che descrive l'intensità dell'accoppiamento e $\Gamma \in \mathbb{R}^{N_c \times N_c}$ è una matrice costante che definisce quale variabile di stato è usata per l'accoppiamento. La matrice $A = (a_{ij} \in \mathbb{R}^{N_c \times N_c})$ definisce, infine, la topologia della rete di accoppiamento tra i nodi. Se vi è un accoppiamento tra i nodi i e j allora $a_{ij} = 1$, altrimenti $a_{ij} = 0$. Un'altra condizione necessaria è che le righe di A abbiano somma nulla.

Quando tutti i nodi saranno in sincronizzazione avremo che la soluzione unica del sistema sarà $x_1 = x_2 = \dots = x_{N_c}$.

Occorre poi introdurre il concetto di esponente di Lyapunov. In un sistema dinamico gli esponenti di Lyapunov di un certo punto p danno una misura di quanto sensibilmente le orbite del sistema siano dipendenti dai dati iniziali. Essi sono quindi degli indicatori della presenza di dinamiche caotiche. Gli esponenti di Lyapunov misurano in particolare la velocità media di allontanamento delle orbite dei punti vicini a p per tempi sufficientemente lunghi. Ad un certo punto p sono associati un numero di esponenti di Lyapunov pari alla dimensione dello spazio e, se l'esponente di Lyapunov massimo è λ e la distanza euclidea tra p ed un punto vicino è abbastanza piccola, tale distanza avrà un'evoluzione nel tempo che per tempi grandi t sarà circa $\delta(t) = c \cdot e^{\lambda t}$. Questo in realtà non accade per ogni punto ma per quasi ogni punto. Se il massimo esponente di Lyapunov del sistema è positivo allora il sistema presenterà una dipendenza sensibile alle condizioni iniziali (effetto farfalla) e sarà quindi caotico.

In altre parole: qualunque sistema dinamico, il cui attrattore abbia almeno un esponente di Lyapunov positivo è caotico. Il valore numerico dell'esponente fornisce un'indicazione precisa sulla scala dei tempi dopo i quali le dinamiche del sistema diventano imprevedibili. Per attrattore s'intende un insieme verso il quale evolve un sistema dinamico dopo un tempo sufficientemente lungo. Perché un insieme possa essere definito attrattore le traiettorie

che arrivano ad essere sufficientemente vicine ad esso devono rimanere vicine anche se leggermente perturbate.

Una volta costituito il sistema di N equazioni che descrive gli N sistemi dinamici, ne consideriamo uno (relativo ad un sistema dinamico). A questo punto costruiamo l'equazione del sistema accoppiato ma sottoposto ad una piccola perturbazione rispetto ad un punto di equilibrio. Ne calcoliamo dunque gli esponenti di Lyapunov (secondo la loro definizione). Consideriamo poi l'esponente più grande: esso sarà la MSF. Se la MSF è negativa allora un'eventuale disturbo dello stato di sincronizzazione decrescerà asintoticamente a zero e quindi la sincronizzazione sarà mantenuta. Altrimenti (se la MSF è positiva) eventuali disturbi si amplificheranno facendo perdere la sincronizzazione.

La MSF è correntemente utilizzata come framework per lo studio dei comportamenti sincronizzati nelle reti complesse: sono state infatti sviluppate teorie che combinano l'approccio basato su MSF e la teoria dei grafi. Mediante l'applicazione della MSF è possibile dimostrare come le reti pesate, in particolare, offrano in generale un aumento considerevole di sincronizzazione.

9 Applicazioni

Le nozioni descritte nelle sezioni precedenti possono essere applicate ad una vasta gamma di reti reali. Vedremo come queste nozioni si possono applicare sia per quanto concerne la struttura che per quanto concerne la dinamica delle reti complesse.

9.1 Reti sociali

Una rete sociale è un insieme di attori, che possono essere individui o gruppi sociali, e relazioni di diverso tipo che possono intercorrere fra essi (ad esempio amicizia).

I concetti principali che riguardano le reti sociali sono:

- Proprietà del mondo piccolo.
- Indici di clustering.
- Misure di centralità (importanza) dei nodi.

Le problematiche principali che caratterizzano le reti sociali sono:

- Caratterizzazione del nodo sulla base delle sue relazioni.
- Ruolo o equivalenza di individui.

Per quanto riguarda il concetto di struttura, relativamente alle reti sociali, per la loro descrizione, si adotta la teoria dei grafi.

Un esempio tipico di reti sociali sono le reti di collaborazione. Una tipica rete di collaborazione è la rete che descrive le collaborazioni tra autori di articoli scientifici. Queste reti sono costituite connettendo fra loro gli scienziati (nodi) se essi sono stati co-autori di almeno un manoscritto scientifico. Molte di queste reti sono costruite per specifiche discipline e mostrano delle proprietà non banali sia in termini di mondo piccolo che in termini di distribuzione di grado.

Un altro esempio di rete di collaborazione è la rete delle citazioni: anche in questo caso tipicamente questa tipologia di rete è riferita ad articoli scientifici. In una rete di questo tipo viene creato un arco diretto dal nodo A al nodo B se l'articolo (nodo) B è citato da A .

Sono importanti anche le cosiddette reti di interazione sociale: un esempio tipico è la rete del linguaggio naturale. In questo le sillabe e le parole costituiscono i nodi, mentre i sinonimi e la sintassi sono rappresentate dagli archi (collegamenti).

Un classico esempio di rete sociale riportato in letteratura è la rete delle relazioni politiche e di amicizia che intercorrono tra le famiglie più importanti della città di Firenze nel Rinascimento. In tale rete si evince che, in termini di importanza (quindi di centralità), la famiglia Medici ricopre il ruolo principale.

Anche lo studio della trasmissione di agenti infetti può essere modellato facendo ricorso alle reti sociali: esempi tipici riguardano l'epatite B, l'HIV, la sifilide, etc.

Una proprietà che, per esempio, contraddistingue la rete dei contatti sessuali tra individui è la distribuzione di grado che obbedisce alla legge di potenza.

Per quanto riguarda invece le problematiche relative alle dinamiche delle reti sociali, sono stati proposti diversi modelli. Tali modelli possono essere suddivisi in due categorie principali:

- Modelli relativi alla formazione delle opinioni.

- Modelli relativi ai giochi di strategia.

9.2 Formazione delle opinioni

L'idea di base dei modelli di formazioni delle opinioni è che gli individui, detti agenti, hanno opinioni che possono cambiare a causa dell'influenza di altri individui. Le dinamiche di queste tipologie di interazione sono governate da semplici regole deterministiche.

Vediamo alcuni modelli proposti.

Modello S. Nel modello S l'opinione dell'individuo i , indicata con o_i ($i = 1, 2, \dots, N$) è trattata come una variabile binarie che può assumere valori $+1$ o -1 . Ad ogni istante di tempo due agenti vicini scelti casualmente trasferiscono la loro opinione ai loro vicini se e solo se essi condividono la stessa opinione.

Modello D. Nel modello D le opinioni o_i sono numeri reali $0 \leq o_i \leq 1$. Ad ogni istante di tempo due agenti vicini scelti casualmente controllano le loro opinioni o_i e o_j per vedere se è possibile un cambio di opinione. Se le loro opinioni differiscono più di un certo valore di soglia fissato $0 \leq \epsilon < 1$ le opinioni rimangono invariate, altrimenti se $|o_i - o_j| < \epsilon$ ciascuna delle due opinioni varia in direzione dell'altra di $\mu \cdot |o_i - o_j|$ con μ un parametro tale che $0 \leq \mu \leq 1/2$.

Modello KH. Nel modello KH le opinioni sono valori reali. Ad ogni istante di tempo un individuo i scelto casualmente cambia la propria opinione assumendo la media aritmetica delle opinioni di tutti i vicini con cui vie è una margine di confidenza fissato ϵ .

In generale, qualunque modello si scelga di utilizzare, per mostrare come si evolve la formazione del consenso, inizialmente si distribuiscono opinioni casuali a tutti i nodi. Il sistema raggiungerà uno stato di equilibrio caratterizzato dalla formazione di gruppi che condividono una certa opinione, in accordo con i valori di controllo fissati.

I modelli visti godono di alcune proprietà. Il modello S dipende dalla distribuzione iniziale delle opinioni: se essa è equilibrata nella configurazione finale avremo circa metà popolazione con un'opinione e l'altra metà con l'altra. Nel modello D se $\epsilon > 1/2$ tutte le opinioni convergeranno verso una, altrimenti se $\epsilon < 1/2$ le differenti opinioni sopravviveranno. Nel modello KH ϵ può assumere come valore minimo 0: per valori bassi di ϵ si formeranno

molti gruppi, mentre man mano che ϵ diventa più grande si formeranno meno gruppi.

È possibile mostrare come nel modello S, se applicato a reti che godono della proprietà del mondo piccolo e che sono ad invarianza di scala, che più alto è il coefficiente di clustering, più facile è la formazione del consenso. Nelle reti BA il consenso completo non sarà mai possibile. Inoltre, sempre nelle reti BA, il modello D produce consenso completo per valori grandi di ϵ e molte opinioni differenti per suoi valori piccoli.

9.3 Giochi di strategia

Un gioco di strategia è un modello che consiste delle seguenti componenti:

- Almeno due *decision-maker*, ovvero persone responsabili di prendere una decisione. Sono anche detti giocatori o agenti.

I giocatori (agenti):

- Seguono le regole del gioco.
- Hanno una loro strategia.
- Giocano per massimizzare il loro guadagno.

Un caso tipico di gioco di strategia è il *dilemma del prigioniero*, classico esempio riportato nella Teoria dei Giochi. Il dilemma può essere descritto nel seguente modo: due criminali vengono accusati di aver commesso un reato. Gli investigatori li arrestano entrambi e li chiudono in due celle diverse, impedendo loro di comunicare. Ad ognuno di loro vengono date due scelte: confessare l'accaduto, oppure non confessare. Viene inoltre spiegato loro che se solo uno dei due confessa, chi ha confessato evita la pena; l'altro viene però condannato a 7 anni di carcere. Se entrambi confessano, vengono entrambi condannati a 6 anni. se nessuno dei due confessa, entrambi vengono condannati a 1 anno. L'obiettivo del gioco è quello di trovare la migliore strategia che permetta di minimizzare la pena ricevuta: in particolare al fine di minimizzare la pena, conviene ad entrambi i prigionieri confessare.

Del dilemma del prigioniero esiste una versione spaziale caratterizzata da N prigionieri, rappresentabile mediante un reticolo regolare: in questa versione ogni giocatore, ad ogni iterazione, gioca con ognuno dei suoi vicini.

In questo caso il guadagno è dato dalla somma dei guadagni ottenuti dalle interazioni con ogni vicino. In particolare la versione spaziale del dilemma può essere trattata mediante rete complessa: il modello che si può utilizzare è quello di una rete a 2 dimensioni che gode della proprietà del mondo piccolo. È possibile, inoltre, dimostrare che il grado di cooperazione non rimane costante ma tende ad essere instabile.

9.4 Internet e WWW

Internet è una rete in cui i nodi rappresentano gli host e i collegamenti rappresentano le connessioni fisiche tra di essi. Tipicamente Internet è descritta in due scale differenti:

- A livello di sistemi autonomi (AS).
- A livello di router.

I sistemi autonomi sono unità organizzative dette *network service providers* a loro volta suddivisi in sottoreti. I device responsabili dello scambio dei pacchetti all'interno delle sottoreti sono i router. I router sono soggetti a politiche di controllo ed prendono le loro decisioni sulla base di protocolli specifici. Abbiamo, in particolare, due tipi di protocolli:

- Protocolli *intra-domain*.
- Protocolli *inter-domain*.

I protocolli *intra-domain* sono usati per la gestione del passaggio di informazioni all'interno dello stesso sistema autonomo, mentre i protocolli *inter-domain* sono usati per gestire il passaggio di informazioni tra diversi sistemi autonomi.

Come esempio di struttura di Internet possiamo considerare la mappa degli AS dall'anno 1998 all'anno 1999. Durante quest'anno di intervallo è possibile constatare come: N sia aumentato (da 3000 circa a 5000 circa), k sia aumentato (da 5000 circa a 10.000 circa), $\langle k \rangle$ sia rimasto quasi invariato (nonostante sia raddoppiato il numero di nodi, $\langle L \rangle \approx 4$ (costante), mentre il rapporto tra la *betweenness* media e il numero di nodi $\frac{\langle b \rangle}{N}$ è leggermente diminuita. Il coefficiente di clustering, infine, già relativamente alto è aumentato ulteriormente.

Il World Wide Web è una rete formata dall'insieme dei collegamenti ipertestuali tra pagine web. È una rete diretta composta da circa 10^{80} nodi. Per quanto concerne la struttura è stato constatato che esiste una distribuzione a legge di potenza in diverse quantità misurabili, in particolare modo per quanto riguarda il numero di richieste di pagine web, che per quanto riguarda il numero di pagine effettivo in un sito web. Un riscontro pratico di questa valutazione è stato ottenuto nelle pagine web del dominio *.nd.edu.

9.5 Dinamiche

Per quanto concerne le dinamiche occorre fare alcune osservazioni. Internet in particolare è soggetta a problemi di congestione. Il World Wide Web invece ha dinamiche differenti perché ogni nodo contiene informazioni. Una problematica relativa al WWW è proprio la ricerca delle informazioni. I motori di ricerca, ad esempio, si occupano di ottenere informazioni e di categorizzare la maggior parte delle pagine web, cercando di stimare correttamente l'accuratezza dei risultati e il valore delle informazioni. L'esempio più importante è l'algoritmo Page Rank proposto da Google.

9.6 Reti metaboliche, proteiche e genetiche

Obiettivo della moderna biologia è quello di sviluppare un'infrastruttura generale in cui le funzioni biologiche possano essere studiate come parte di una complessa struttura molecolare o cellulare.

Le cellule, in particolare, sono una componente fondamentale della vita. Esse sono costituite da migliaia di molecole con struttura disomogenea. L'attività cellulare è gestita dal DNA che codifica tutte le istruzioni necessarie.

Un esempio significativo è il ruolo che ricopre il gene p53 e la sua proteina corrispondente: esso è un *tumor-suppressor* ed in caso di suo malfunzionamento può essere la causa di tumori.

L'attività delle cellule costituisce un aspetto di studio molto importante. Le cellule hanno il compito di costruire, a partire da molecole complesse, molecole più piccole e semplici: un esempio tipico è il glucosio, a partire dal quale è necessario generare delle molecole più semplici. Le reazioni cellulari che producono una certa quantità di energia sono dette metaboliche.

Le reazioni metaboliche sono gestite e regolate dagli enzimi. Molti enzimi sono proteine che, a loro volta, sono dei polimeri. Un polimero è una macro-molecola (molecola da elevato peso molecolare) costituita da un gran numero di gruppi molecolari (detti unità ripetitive) uguali o diversi (nel caso dei copolimeri), uniti “a catena” mediante la ripetizione dello stesso tipo di legame di tipo covalente. Nello specifico le proteine sono costituite dalla combinazione di 20 tipi di aminoacidi differenti. Le proteine possono avere funzioni strutturali e meccaniche (nei muscoli, nelle ossa o nei tessuti) oppure essere importanti mediatori nella trasmissione di segnali inter ed intra-cellulari.

Un gene, invece, è un’unità discreta che contiene informazioni ereditarie e che consiste in una sequenza di nucleotidi del DNA (quest’ultimi sono le unità ripetitive del DNA).

Un’altra componente fondamentale delle cellule è costituita dai ribosomi. Questi si occupano di sintetizzare una specifica proteina, seguendo le indicazioni riportate nelle sequenze di acidi nucleici trascritti da un gene nell’mRNA (RNA messaggero). L’mRNA è un tipo specifico di RNA che ha proprio il compito di codificare e trasportare le informazioni necessarie alla sintesi proteica dai cromosomi (quindi dal DNA) ai ribosomi. La trascrizione delle informazioni nell’mRNA viene effettuata dall’RNA polimerasi.

Un’esempio di struttura di rete biologica è la rete delle relazioni metaboliche tra organismi. L’esempio più tipico riportato in letteratura tratta nello specifico 43 specie diverse di individui. In tale rete i nodi rappresentano i metaboliti e gli archi rappresentano le interazioni biochimiche. Un nodo riceve in ingresso un arco quando un metabolita viene prodotto e riceve, invece, un arco in uscita quando un metabolita viene eliminato. Il grafo che ne risulta sarà caratterizzato da invarianza di scala sia per quanto concerne il grado entrante che quello uscente.

In generale nelle reti metaboliche si riscontrano sempre proprietà di invarianza di scale mentre nelle reti delle reazioni no. La proprietà del mondo piccolo è invece tipicamente riscontrata in entrambe le tipologie di reti. Le reti di reazioni, in particolare, sono reti i cui nodi corrispondono alle reazioni metaboliche e due nodi sono collegati tra loro se condividono un metabolita.

È stata studiata la struttura “modulare” delle reti metaboliche: in particolare è stato mostrato che i coefficienti di clustering sono indipendenti dalla dimensione delle rete. Inoltre il coefficiente di clustering gode di proprietà simili a quelle riscontrate in reti gerarchiche e determina che le reti

metaboliche non possono essere partizionate in moduli.

Le reti metaboliche hanno una struttura detta *bow-tie*: questo significa che i metaboliti in una rete metabolica sono molto lontani dall'essere totalmente connessi. Il sottoinsieme più grande totalmente connesso di una rete metabolica è chiamato *giant strong component* ed è caratterizzato da invarianza di scala.

Altre classi di reti cellulari sono le reti di interazione proteina-proteina e le reti proteina-gene. Nella loro rappresentazione sotto forma di grafo un nodo corrisponde ad una proteina e due proteine sono collegate fra loro se interagiscono. Le proprietà di questa classe di reti sono: invarianza di scala, proprietà del mondo piccolo, elevato clustering e modularità gerarchica.

Un altro esempio interessante è la rete delle interazione tra proteine nel lievito: in particolare tale rete risulta essere dissortativa. Una rete si dice dissortativa se i suoi nodi di alto grado preferiscono connettersi con nodi di grado basso. D'altro canto una rete è assortativa se i nodi grado alto preferiscono connettersi a nodi di grado alto.

Una proprietà che caratterizza le reti ad invarianza di scala è l'esistenza di nodi detti *hub* che ricoprono un ruolo fondamentale nell'evitare il riscontarsi di errori casuali. Grazie alla loro presenza le reti di interazione proteica, ad esempio, risultano essere molto tolleranti alla rimozione casuale di nodi.

I cammini molto brevi tra le componenti delle cellule sono, invece, da considerarsi legati alle loro proprietà di risposta veloce e di adattamento rapido a cambiamenti esterni di circostanze e alla loro reazione ad eventuali perturbazioni.

Nelle cellule è stata osservata una struttura modulare: una cellula, infatti, è caratterizzata dalla presenza di moduli separati altamente integrati che svolgono funzioni differenti. Anche in questo caso la struttura modulare influenza il modo con cui le cellule si adattano ai cambiamenti esterni.

Altri studi hanno mostrato come le reti di trascrizione siano caratterizzate dalla composizione di piccoli insiemi di motivi (ad esempio il *feed forward loop*).

Per quanto concerne le dinamiche delle reti biologiche, occorre prima definire i seguenti concetti:

- Operone: insieme dei geni che vengono regolati in modo strettamente coordinato.

- Gene regolatore: gene coinvolto nell'espressione genica di uno o più geni (l'espressione genica è il processo per cui l'informazione contenuta in un gene viene convertita in macro-molecola funzionale).
- Repressione trascrizionale: processo per cui un repressore inibisce l'espressione di uno o più geni legandosi all'operatore.

L'operatore, nella genetica molecolare, è una sequenza di DNA, utilizzata per la regolazione dell'attività dei geni dell'operone. La regolazione genica è il processo che permette ad una cellula di esprimere un determinato gruppo di geni in un contesto e di silenziarne altri.

L'obiettivo è comprendere la relazione tra struttura e funzione e come essa è determinata dall'ambiente biologico. Per esempio è stato analizzato un modello con interazioni attivatorie ed inibitorie che descrive processi biologici, quali le relazioni metaboliche e l'espressione genica, che possono essere usati per esplorare l'esistenza di punti fissi stabili e stati periodici in reti eterogenee. In particolare si evince che l'attività dei nodi e la loro evoluzione nel tempo può essere descritta tramite appositi sistemi di equazioni differenziali di primo ordine.

Un modello per la rappresentazione delle dinamiche fra agenti è stato pensato per descrivere l'interazione virus-cellula nei batteri e, più in generale, per l'attività di regolazione genica. In questo modello ogni nodo rappresenta un gene regolatore, mentre un arco descrive la loro interazione, ovvero: due nodi estremi sono considerati unità trascrizionali se includono un gene regolatore; ciascuno di questi nodi può essere pensato come l'inizio di una interazione.

9.7 Brain networks

I neuroni sono distribuiti e collegati fra loro con una struttura analoga a quella riscontrata nel Web. La rete che ne risulta, in particolare, è estremamente sparsa. Nonostante l'enorme numero di neuroni e delle loro connessioni nel cervello umano, in realtà l'organizzazione stessa del cervello è regolata da principi di ottimizzazione nell'allocazione delle risorse.

Per quanto riguarda la struttura di queste reti occorre sottolineare come, in termini di connettività, occorra distinguere tre categorie principali:

- Connettività neuro-anatomica: connettività fisica tra due punti del cervello. Difficile da formalizzare.

- Connettività funzionale: descrive la correlazione temporale tra due eventi neurofisiologici spazialmente distanti.
- Connettività effettiva: è l'influenza che un sistema neuronale esercita su di un altro, sia per via diretta che indiretta.

Un sistema sperimentale per comprendere la relazione che intercorre tra struttura e funzioni di un sistema nervoso è quello relativo al *Caenorhabditis elegans*, un verme lungo circa 1 mm che vive nel suolo in regioni temperate. Il *C. elegans* ha un sistema nervoso in cui sono stati identificati 302 neuroni la cui connettività è stata interamente mappata.

Sono state anche analizzate le connessioni nella corteccia visiva primaria dei primati (nello specifico nel macaco): tali analisi sono state effettuate mediante un algoritmo basato sul *simulated annealing*. La corteccia visiva primaria è in particolare quell'area del cervello altamente specializzata per il processamento dell'informazione riguardante la forma e la collocazione di oggetti statici e il loro movimento nel campo visivo. È stato provato come nel macaco l'organizzazione locale delle connessioni sembri giocare un ruolo molto importante nell'organizzazione del suo sistema visuale.

In generale, comunque, la rappresentazione della connettività nei sistemi nervosi può essere effettuata mediante metodi di ottimizzazione, quali ricerca dei minimi o dei massimi globali.

In definitiva è stato mostrato che, mediante costruzione di un'apposita matrice detta di connettività neuronale, sia il sistema nervoso del *C. elegans*, sia quello di cortecce cerebrali di altri esseri viventi, così come nelle retine neuronali ricreate in vitro, è possibile mostrare che vale la proprietà del mondo piccolo.

Inoltre in differenti matrici di connettività sono state osservate correlazioni tra lunghezza dei cammini o coefficienti di clustering e il numero dei vicini di ogni regione del cervello individuale. Alcune di queste reti hanno mostrato pure un'organizzazione a multicluster che assume un comportamento *scale-free*.

È stata analizzata anche la connettività del sistema visuale del gatto ed è stato notato che ogni pattern di connettività locale ha caratteristiche univoche. Ciascuno di questi insiemi viene detto *connectional fingerprint*.

Molte famiglie di pattern locali nelle connessioni neuronali sono state scoperte mediante l'analisi dei motivi: in particolare i motivi più frequentemente riscontrati sono il *feed-forward-loop* e il *bi-parallel*.

Per quanto riguarda la vulnerabilità delle connessioni, essa può essere causata da eventuali lesioni del cervello. L'effetto delle modifiche nella connettività può essere studiato mediante appositi modelli: in particolare è stato constatato che la risposta della rete alla rimozione di nodi è strutturalmente correlata alla connettività dei nodi target.

In generale lo studio della struttura neuronale viene effettuato solo da un punto di vista anatomico. Solo pochi studi analizzano la struttura in un'ottica temporale, focalizzando l'attenzione sulle correlazioni temporali fra le attività del cervello. In tal senso le principali tecniche usate sono:

- EEG (elettroencefalografia): registrazione dell'attività elettrica nell'encefalo.
- MEG (magnetoencefalografia): misurazione dei campi magnetici prodotti dall'attività elettromagnetica dell'encefalo.
- fMRI (risonanza magnetica funzionale): imaging a risonanza magnetica per valutare la funzionalità dell'encefalo.

Tuttavia il problema principale nell'analisi statistica della connettività del cervello riguarda la piccola dimensione dei database disponibili, che porta alla necessità di dover far uso di reti surrogate.

Per quanto concerne le dinamiche, al fine di studiare il ruolo che intercorre tra le funzioni del cervello e la connettività neuronale, occorre introdurre una misura, detta complessità, che serve a quantificare come gruppi locali specializzati (nel cervello) siano fra di loro integrati in un unico sistema coerente.

È stato provato, attraverso analisi statistiche, che reti con elevata complessità sono caratterizzate da una coesistenza di alto grado di connettività che risulta integrata in una dinamica globale coerente.

I risultati ottenuti suggeriscono che queste reti tendono a massimizzare l'integrazione fra interazioni locali e globali.

Un'ulteriore misura introdotta è l'indice funzionale di cluster, usata per identificare l'integrazione funzionale tra attività neuronali distribuite ed organizzate.

Le reti neuronali reali mostrano tipicamente una connettività molto sparsa ed il meccanismo di interazione tra distinti assemblaggi neuronali è strettamente collegato al grado di casualità nella topologia delle connessioni. I

sistemi diventano altresì più coerenti se aumenta la connettività per nodo nelle reti.

Molti studi hanno suggerito che la topologia delle connessioni in reti neurali reali presentano una combinazione di proprietà topologiche che non possono essere studiate né con modelli standard (ad es. reticoli), nemmeno mediante reti casuali.

A tal proposito sono stati introdotti tre modelli:

- Un modello neuronale basato su automi cellulari con il quale è stato possibile mostrare come le reti neurali reali tendano a propagare le attività meglio di reti con topologia regolare o reti casuali.
- Il modello HH (Hodgkin-Huxley) che mostra invece comportamenti in contrasto con le proprietà del mondo piccolo, ma produce risposte veloci e coerenti.
- Il modello FN (FitzHug-Nagumo) che mostra come la velocità di propagazione non sia dipendente dalla struttura della rete.

Un altro fenomeno ampiamente studiato è la formazione delle memorie associative. Lo scopo delle memorie associative è quello di memorizzare un insieme di pattern in modo che quando se ne presenta un nuovo, il sistema risponda con uno stato stabile che corrisponde al pattern più simile già memorizzato. È stato mostrato come, a causa della connettività sparsa delle reti neurali, venga modificata la capacità di memorizzare e recuperare i pattern della memoria associativa. Tuttavia i modelli che descrivono questo tipo di memorie risultano essere molto resistenti ad eventuali danni locali.

In alcuni modelli che descrivono le memorie associative sono state introdotte delle topologie che godono della proprietà del mondo piccolo. Altri modelli invece sono caratterizzati da topologie *scale-free*.

Per quanto riguarda le malattie del cervello, come ad esempio l'epilessia, è necessario ricorrere alle teorie relative alla sincronizzazione per descrivere come avviene la loro propagazione nella rete.

10 Algoritmi per la ricerca di comunità

Le comunità sono una caratteristica tipica delle reti sociali: alcuni individui della rete possono far parte infatti di un gruppo strettamente connesso

di un'élite social chiusa oppure essere completamente isolati, mentre altri ancora possono costituire una sorta di ponte di collegamento tra gruppi differenti. Tipicamente questi gruppi sociali sono caratterizzati da loro regole, orientamenti, sottoculture, etc.

Le comunità sono importanti anche relativamente alle reti complesse. Per esempio gruppi di nodi strettamente connessi nel World Wide Web corrispondono a pagine web che trattano lo stesso tema.

Gli approcci tradizionali per la ricerca delle comunità sono di due tipi:

- Partizionamento spettrale.
- Clustering gerarchico.

10.1 Partizionamento spettrale

In generale il partizionamento di un grafo consiste, dato un grafo G , nel dividere l'insieme dei vertici V in due sottoinsiemi di vertici di circa la stessa dimensione V_1^* e V_2^* , garantendo di minimizzare il numero di archi tra i due sottoinsiemi.

Sono stati elaborati diversi metodi euristici, principalmente basati su metodi spettrali. Altri metodi sono ottenuti invece sfruttando le proprietà degli autovettori della matrice di adiacenza \mathcal{A} del grafo. Ulteriori metodi sono infine basati sulla matrice Laplaciana Λ e sulla matrice normale \mathcal{N} .

Ricordiamo che \mathcal{D} è la matrice diagonale i cui elementi sono $D_{ii} = \sum_j a_{ij} = k_i$, da cui $\mathcal{N} = \mathcal{D}^{-1}\mathcal{A}$ e $\Lambda = \mathcal{D} - \mathcal{A}$ (quest'ultima è un matrice semi-definita simmetrica positiva).

Dato G definiamo suo spettro l'insieme degli autovalori di \mathcal{A} .

Lo spettro della matrice Laplaciana, inoltre, ricopre un ruolo molto importante in quanto gode di diverse proprietà che possono essere usate per partizionare grafi sparsi molto grandi. In particolare i suoi autovalori sono tutti reali e non negativi. Inoltre Λ ha un insieme completo di N autovettori reali ed ortogonali.

La somma di tutte le righe di Λ è zero, quindi Λ avrà sempre il più piccolo autovalore $\lambda_1 = 0$ a cui corrisponde l'autovettore $v_1 = (1, 1, \dots, 1)$. Di conseguenza quando il grafo è composto da c componenti, la molteplicità dell'autovalore nullo sarà uguale al numero di componenti c . Tuttavia, quando il grafo ha un'apparente struttura con c comunità, ci sarà un solo autovalore uguale a 0 e $c - 1$ autovalori saranno maggiori di 0. In questo caso dunque

consideriamo il secondo autovalore più piccolo λ_2 a cui è associato l'autovettore \mathbf{v}_2 detto autovettore di Siedler. \mathbf{v}_2 può essere usato per approssimare una soluzione al problema del partizionamento del grafo basata sul segno delle componenti di \mathbf{v}_2 . In particolare il primo grafo sarà costituito dai vertici con componenti positive, il secondo da vertici con componenti negative. Questo metodo è detto di bisezione spettrale ed ha complessità $\mathcal{O}(N^3)$ per matrici $N \times N$.

10.2 Clustering gerarchico

I metodi basati su clustering gerarchico sono tipicamente usati nell'analisi delle reti sociali. In questo caso il termine clustering è da intendersi diversamente dal solito: in particolare si intenderà l'organizzazione di N oggetti in gruppi, detti cluster, tali che gli oggetti all'interno di ciascun cluster siano più strettamente correlati rispetto agli oggetti all'interno di altri cluster. In questo caso il concetto di correlazione all'interno di un cluster è assimilabile al concetto di similarità: gli oggetti all'interno di un cluster saranno più *simili*.

Abbiamo due tipologie di metodi:

- Metodi agglomerativi che consistono in una sequenza di fusioni di N oggetti in gruppi.
- Metodi divisivi che consistono in una sequenza di separazione di N oggetti in gruppi.

Una tipica tecnica agglomerativa funziona nel seguente modo. Supponiamo di avere un insieme di N oggetti e una matrice $N \times N$ delle distanze.

I passi saranno i seguenti:

1. Si assegna ogni elemento ad un cluster: quindi inizialmente avremo N cluster contenenti ciascuno un elemento.
2. Si cerca la coppia di cluster più simili e si unisce in un singolo cluster.
3. Si ricalcolano le distanze tra il nuovo cluster e ciascuno degli altri cluster.
4. Si ripetono i passi 2 e 3 fin quando si ottiene un unico cluster di N elementi.

La struttura risultante è rappresentabile mediante un albero gerarchico detto *dendrogramma*.

In questa procedura il passo critico è il terzo in quanto occorre capire come effettuare il ri-calcolo delle distanze. Il ri-calcolo può essere fatto mediante vari metodi:

- Linkaggio singolo: la distanza fra due cluster è uguale alla distanza minima tra un qualsiasi oggetto di un cluster e un qualsiasi oggetto di un altro cluster.
- Linkaggio completo: analogo al linkaggio singolo ma si tiene conto della distanza massima.
- Linkaggio medio: analogo ai precedenti ma si tiene conto della distanza media.

Nel caso in cui gli oggetti da raggruppare siano nodi di un grafo, la similarità è definita in termini di equivalenza strutturale: due nodi saranno strutturalmente equivalenti se avranno lo stesso pattern di relazione con tutti gli altri nodi. Per esempio nelle reti sociali due individui sono strutturalmente equivalenti se hanno gli stessi amici.

Possono essere usate diverse misure per calcolare l'equivalenza strutturale: ad esempio la distanza euclidea tra righe e colonne di due nodi i e j della matrice di adiacenza:

$$D_{ij} = \sqrt{\sum_{k=1, k \neq i, j} (a_{ik} - a_{jk})^2}.$$

Un altro modo consiste nel calcolare la correlazione di Pearson tra righe (o colonne) della matrice di adiacenza:

$$R_{ij} = \frac{\sum_{k=1}^N (a_{ik} - \bar{a}_i)(a_{jk} - \bar{a}_j)}{\sqrt{\sum_{k=1}^N (a_{ik} - \bar{a}_i)^2} \cdot \sqrt{\sum_{k=1}^N (a_{jk} - \bar{a}_j)^2}},$$

dove \bar{a}_i e \bar{a}_j sono i valori medi rispettivamente della i -esima e della j -esima riga di \mathcal{A} .

10.3 Algoritmo di Girvan-Newman

L'algoritmo di Girvan-Newman è un metodo iterativo basato sulla ricerca e sulla rimozione progressiva degli archi con la più alta *betweenness*. I passi dell'algoritmo sono i seguenti:

1. Calcolare la *betweenness* per ogni arco.
2. Rimuovere l'arco con la *betweenness* più alta.
3. Effettuare un'analisi delle componenti della rete.
4. Tornare al punto 1 fin quando tutti gli archi non siano rimossi e il sistema non sia costituito da N nodi non connessi.

Anche in questo caso l'output dell'algoritmo sarà un dendrogramma.

Le prestazioni dell'algoritmo sono state testate in reti artificiali casuali modulari ovvero in reti generate in modo tale da avere una struttura a comunità ben definita. Tutti i grafi generati hanno 128 nodi e 1024 archi; i nodi sono stati divisi in 4 gruppi e gli archi sono etichettati opportunamente a seconda se colleghino vertici dello stesso gruppo o meno. Ogni generico vertice del grafo ha un numero medio z_{in} di archi che lo collegano a vertici dello stesso gruppo e un numero medio z_{out} di archi che lo collegano a vertici di altri gruppi. z_{in} e z_{out} sono tali che $z_{\text{in}} + z_{\text{out}} = 16$. Lo scopo finale del test è capire quanti vertici sono correttamente classificati in funzione di z_{out} . Con queste ipotesi l'algoritmo produce una gerarchia di suddivisioni delle rete: per capire quale tra le divisioni ottenute sia la migliore occorre però introdurre il concetto di modularità Q . Q è definita come segue. Supponiamo di voler testare la bontà delle suddivisioni di un grafo G in n comunità. Introduciamo una matrice simmetrica $n \times n$ che chiamiamo E in cui ogni elemento e_{ij} è la frazione di tutti gli archi della rete che collegano vertici della comunità i a vertici della comunità j . Sia:

$$\text{Tr}E = \sum_i e_{ii}$$

la frazione di archi nella rete che connettono vertici nella stessa comunità (traccia), mentre sia la somma delle righe:

$$a_i = \sum_j e_{ij}$$

sia la frazione di archi che connettono vertici della stessa comunità i .

Se la rete è tale che la probabilità di avere un arco tra due siti sia la stessa indipendentemente dalla loro eventuale appartenenza alla stessa comunità avremo $e_{ij} = a_{ij}$. A questo punto possiamo definire la modularità come:

$$Q = \sum_i (e_{ii} - a_i^2) = \text{Tr}E - ||E^2||,$$

dove $||E^2||$ è la somma degli elementi della matrice E^2 .

La modularità misura il grado di correlazione tra la probabilità di avere un arco che unisca due siti e il fatto che questi siti possano divenire la stessa comunità. Se $Q = 1$ (valore massimo) allora avremo una struttura a comunità forte, se $Q = 0$ (valore minimo) allora avremo un grafo senza struttura a comunità.

Massimi e minimi locali nella modularità, ottenuti durante i passi dell'algoritmo GN, corrisponderanno a buone suddivisioni del grafo in comunità.

10.4 Variazioni ed estensioni dell'algoritmo GN

Possibili varianti dell'algoritmo GN prevedono l'utilizzo di misure di centralità alternative. Tipicamente sono utilizzate le seguenti misure:

- *Random-walk betweenness*: misura che tiene conto dei cammini casuali tra tutte le coppie di vertici anziché dei cammini minimi.
- *Current-flow betweenness*: misura definita considerando un circuito elettrico associato al grafo e utilizzando le leggi di Kirchoff.

Le leggi di Kirchoff, in particolare, sono due relazioni legate ai principi di conservazione della carica nei circuiti elettrici.

Gli algoritmi basati su queste misure di centralità calcolate opportunamente richiedono tempo d'esecuzione $\mathcal{O}(N^4)$. I risultati empirici ottenuti con le due distinte misure di centralità sono identici.

Un'altra variante dell'algoritmo GN prevede l'uso della *information centrality*, una quantità basata sul concetto di propagazione efficiente delle informazioni sulla rete, così definita:

$$C_k^I = \frac{\Delta E}{E} = \frac{E(G) - E(G'_k)}{E(G)}$$

per ogni arco $k = 1, 2, \dots, K$. G'_k è il grafo con N nodi e $K - 1$ archi ottenuto rimuovendo da G l'arco K , $E(G)$ è l'efficienza del grafo G . I tempi di esecuzione in questo caso saranno in generale $\mathcal{O}(K^3N)$, per grafi sparsi $\mathcal{O}(N^4)$.

Un'ulteriore variante dell'algoritmo GN, che ne migliora la velocità di calcolo a discapito però di una minore accuratezza dei risultati, è basata sulla stima della *betweenness* degli archi effettuata mediante metodi Monte Carlo: in questo caso anziché considerare il contributo della *betweenness* per tutti i possibili cammini, viene considerata solo per uno di essi scelto casualmente. Per grafi relativamente piccoli l'algoritmo funzionerà bene, ma ripetendo più volte i calcoli si potrebbero ottenere risultati ambigui.

Abbiamo pure una variante basata sul conteggio di piccoli cicli all'interno del grafo: poiché la *betweenness* degli archi è una quantità che dipende dalle proprietà di tutto il grafo, un algoritmo più veloce può essere ottenuto usando misure di centralità locali anziché globali. In questo caso viene definito come coefficiente di clustering dell'arco il numero di triangoli in cui un arco arriva diviso per il numero di possibili triangoli che possono essere costruiti dato il grado dei nodi. Formalmente:

$$C_{ij}^{(3)} = \frac{z_{ij}^{(3)} + 1}{\min[(k_i - 1, k_j - 1)]}$$

dove $z_{ij}^{(3)}$ è il numero di triangoli costruiti sull'arco e $\min[(k_i - 1, k_j - 1)]$ è il numero massimo possibile di triangoli. L'algoritmo lavora come l'implementazione classica di Girvan-Newman ma, ad ogni passo, gli archi rimossi saranno quelli con il più basso valore di C_{ij} . In questo caso i tempi saranno $\mathcal{O}(k^2)$ in generale, $\mathcal{O}(N^2)$ su grafi sparsi.

10.5 Metodi veloci basati sulla modularità

Vi è un altro metodo che è basato sull'ottimizzazione della formula relativa alla modularità. Poiché la ricerca esaustiva di tutte le possibili divisioni potrebbe richiedere tempo esponenziale in N , il metodo è un'ottimizzazione *greedy* basata su un'agglomerazione di piccole comunità. L'algoritmo funziona nel seguente modo. Inizialmente abbiamo N comunità contenenti ciascuna un nodo. Ad ogni passo le comunità sono unite a coppie, scegliendo di unire le coppie che hanno avuto il maggior incremento (o il minor decremento) del valore di Q . Vengono considerate solo le coppie tra cui vi sono archi,

poiché le comunità non collegate da archi non producono cambiamenti in Q . La variazione è data da:

$$\Delta Q = e_{ij} + e_{ji} - 2a_i a_j = 2(e_{ij} - a_i a_j).$$

Successivamente, all'unione di due comunità vengono aggiornate le righe e le colonne della matrice E (degli elementi e_{ij}) corrispondenti alle comunità che sono state unite (in tempo $\mathcal{O}(N)$). Ogni passo dell'algoritmo richiede tempo $\mathcal{O}(K + N)$; il dendrogramma è costruito in tempo $\mathcal{O}((K + N) \log N)$ e $\mathcal{O}(N^2)$ su grafi sparsi.

Un'ulteriore ottimizzazione del precedente approccio greedy riesce ad ottenere tempo $\mathcal{O}(KD \log N)$ dove D è la profondità del dendrogramma.

10.6 Metodi basati su analisi spettrale

Un'approccio possibile consiste nell'utilizzare la bisezione spettrale ricorsiva. Il metodo fa ricorso all'autovettore di Fiedler, che viene usato per dividere il grafo in due sottografi. In particolare occorre cercare l'autovettore di Fiedler di ciascuno dei sottografi e continuare ricorsivamente finché non si ottiene il numero di sottografi desiderato.

In molti casi accade che il numero di nodi della rete sia molto alto e le comunità non sempre siano facilmente individuabili guardando, ad ogni interazione, al solo primo autovettore non banale (ovvero quello di Fiedler). Di conseguenza si può pensare di estrarre le comunità a partire dalle correlazioni che intercorrono tra le stesse componenti in differenti autovettori. Una soluzione consiste nell'introdurre una nuova misura per calcolare le correlazioni: tale misura, indicata con R_{ij} è la misura della vicinanza tra il nodo i e il nodo j .

Altri approcci possibili sono:

- Metodo basato sull'associazione al grafo di un circuito elettrico. Ad ogni arco è assegnata un'unità di resistenza; due nodi avranno una certa differenza di potenziale fissata. Il metodo, basato sulle leggi di Kirchoff, produrrà delle comunità in base al voltaggio associato ai nodi in tempo $\mathcal{O}(N + K)$.
- Metodo basato su cammini casuali: si introduce la quantità d_{ij} intesa come il numero medio di passi necessari affinché una particella di

Brownian si muova da i a j . Il moto della particella è detto browniano: esso è un moto disordinato tipico delle particelle nei fluidi.

11 Algoritmi per la navigazione e la ricerca

L'algoritmo *breadth-first* (ampiezza) è un algoritmo per la ricerca del cammino più breve da un nodo p (sorgente) ad un nodo q (destinazione) in un grafo. I passi sono i seguenti: al primo passo ($d = 0$) un agente è collocato nel vertice sorgente p ; l'agente controlla i vicini di \mathcal{N}_p di p ; se il vertice destinazione q non è nessuno di questi vicini, al passo $d = 1$ degli agenti sono distribuiti su tutti gli elementi di \mathcal{N}_p che non siano stati visitati da altri agenti in passi precedenti; nelle successive iterazioni ($d > 1$) ogni nuovo agente ripete la procedura finché tutti i vertici non saranno stati visitati. Successivamente, attraverso un'operazione di back-tracking l'agente che ha individuato il nodo destinazione dovrà ricostruire il processo all'inverso verso il nodo sorgente. Il tempo necessario sarà $\mathcal{O}(N)$ per la costruzione del cammino. Il limite di questo approccio è dato dalla necessità di multi-agenti. Evoluzioni dell'algoritmo prevedono l'utilizzo di un singolo agente.

Un ulteriore approccio prevede l'introduzione di una misura di centralità del cammino casuale C per ciascun nodo del grafo. Tra tutti i possibili cammini tra due nodi, il cammino verso il nodo corrispondente al più grande valore di C è il più veloce rispetto agli altri e, in un certo intervallo di tempo, i nodi con più alti valori di C saranno quelli visitati più frequentemente, a partire da cammini casuali distribuiti uniformemente.

Una strategia possibile, detta strategia del massimo grado, è basata sull'assunzione che un vertice abbia informazioni sul grado dei suoi vicini. In questo caso l'agente si muoverà sempre verso il vicino che avrà il grado più alto. Se il vicino con il grado più alto è già stato visitato in passato, allora si sposterà sul secondo vicino con il grado più alto e così via. In reti altamente eterogenee, come Internet, i nodi con grado più alto sono connessi ad una significativa frazione di nodi della rete e, in questi casi, l'agente trova il percorso verso la destinazione in pochi passi. Con questo approccio si ottengono buoni risultati con reti ad invarianza di scale generate mediante i modelli di Aiello et al. e BA.

Nel modello WS, in cui la posizione dei nodi ha una geometria ad anello indipendentemente dalla sua topologia, è stata testata una strategia di ricerca in cui il nodo successivo da esplorare è selezionato dall'insieme dei nodi vicini

tra quelli che hanno la più piccola distanza dalla destinazione: la distanza è calcolata come la distanza geometrica lungo l'anello tra i due nodi.

Tutti gli approcci visti finora non richiedono la conoscenza della topologia della rete. Adesso vedremo quindi quali sono le topologie delle reti che consentono di ottimizzare specifici algoritmi. Prendiamo il caso delle reti godono della proprietà del piccolo mondo. Un modello, che è una variante del modello WS, è costruito in modo che vengano aggiunte delle scorciatoie tra coppie di punti in reticolo bidimensionale con una probabilità $r^{-\alpha}$, dove r è la distanza tra i punti e α è un parametro variabile. La strategia di ricerca adottata è di tipo greedy: l'agente esplora i nodi che sono più vicini alla destinazione. È possibile mostrare che la prestazione migliore dell'algoritmo si ottiene con $\alpha = 2$.

Un modello alternativo consiste nel considerare gli individui della rete gerarchicamente raggruppati in categorie, in accordo con le loro identità sociali. Queste categorie sono poi raggruppate in super-categorie e così via, fino a creare una gerarchia ad albero che definisce la distanza sociale tra due popolazioni, determinata dall'altezza del più basso livelli dell'albero in cui i due nodi sono collegati. La rete viene quindi costruita connettendo due individui con una probabilità maggiore nel caso di distanza sociale più bassa e assumendo che la probabilità decada esponenzialmente per distanze sociali più grandi. L'algoritmo è implementato in modo da consentire lo scambio di messaggi tra gli individui (i nodi). La strategia adottata è di tipo greedy, in particolare lo scambio di messaggi avviene nel seguente modo: ad ogni passo il messaggio viene inviato al vicino che ha la più piccola distanza sociale dal destinatario.

Un altro approccio è così implementato: supponiamo che un agente in un nodo p voglia conoscere il cammino più breve verso un certo nodo q collocato da qualche parte nella rete. Nel momento in cui l'agente si trova in p deve individuare l'arco uscente nella direzione che condurrà a q . L'approccio utilizzato è tipico della situazione in cui, ad esempio, un individuo, all'interno di una città che non conosce, deve recarsi, a partire da un certo punto di partenza, ad una certa destinazione avente un certo indirizzo. L'unica informazione che si conosce è solo quella relativa alla sequenza di nomi di strade che devono essere attraversate per giungere a destinazione. A livello di rete complessa si costruirà quella che viene definita *information city network* in cui le strade sono i nodi e le intersezioni fra strade sono gli archi che collegano due nodi. L'algoritmo si basa sul fatto che al primo passo la scelta dell'arco

uscite è correlata al grado k_p del nodo p ; al passo successivo, una volta arrivati al nodo j , è possibile escludere l'arco precedente cui si è pervenuti; quindi la successiva scelta del nodo uscente dipenderà dal valore $k_j - 1$ e così via.

Altri esempi di modelli di reti sono costruiti appositamente per gestire anche situazioni di congestione.

12 Cablaggio adattivo e dinamico

Supponiamo di considerare le reti come entità dinamiche, ovvero come reti la cui topologia si può evolvere nel tempo a causa di eventi esterni o interni o per via di regole prefissate.

Vediamo alcuni esempi:

- Modello che descrive le interazioni tra le specie. In questo modello sono state introdotte due tipologie di variabili dinamiche: la variabile *fast* che serve per modellare le dinamiche delle specie e la variabile *flow* che serve per modellare le interazioni tra le specie.
- Reti (tipicamente neuronali) che si evolvono tramite un insieme di regole attraverso le quali alcuni nodi guadagnano nuovi link nel corso del tempo, mentre altri li perdono.
- Reti booleane adattive: reti in cui gli agenti hanno associato uno stato binario che può evolversi nel tempo secondo una regola definita *minority game*. In questo caso gli agenti adottano la loro dinamica scegliendo tra diverse funzioni booleane, passando da una all'altra in funzione dell'attuale capacità di predire il gruppo vincente.
- Reti che si evolvono stocasticamente. In questo caso sono stati proposti due approcci: un approccio basato su alberi ricorsivi casuali e uno basato su modelli stocastici.

Alcune ricerche hanno tentato di descrivere come avviene la diffusione dei virus attraverso le email. Molti virus utilizzano un determinato meccanismo per diffondersi: in un primo momento pervengono ad un destinatario, tipicamente come allegato ad un messaggio di posta elettronica, poi, se attivati, inviano delle repliche di se stessi ad altri utenti. Una strategia adottata

dai virus è generalmente quella di usare gli indirizzi di posta elettronica che sono contenuti nella rubrica del destinatario. Per riprodurre le dinamiche della diffusione dei virus sono stati sviluppati appositi modelli. In uno di questi i nodi sono le rubriche degli indirizzi di posta elettronica degli utenti, mentre gli archi sono i record degli indirizzi di posta elettronica di ciascuna rubrica. La peculiarità di questo approccio è che, al trascorrere del tempo, nuovi archi vengono creati ed altri vengono rimossi. Questo è conseguenza del fatto che gli agenti, in genere, riorganizzano periodicamente le loro rubriche. Il bilanciamento tra aggiunta e rimozione di indirizzi dipende da svariati fattori. Possono, in particolare, essere distinti due casi di contatti di posta elettronica:

- Indirizzi equivalenti: indirizzi per cui il numero di messaggi di posta elettronica inviati da ogni utente è indipendente dal grado dell'utente.
- Indirizzi correlati al grado: indirizzi per cui il numero di messaggi di posta elettronica inviati da ogni utente dipende dal grado dell'utente.

In ambedue i casi per grado dell'utente si intende il numero di indirizzi presenti nella propria rubrica. Nel caso di indirizzi equivalenti è stato evidenziato che piccole reti si evolvono esibendo una lunghezza media dei cammini che tende a diminuire nel tempo. La rete tenderà quindi ad acquisire la proprietà del mondo piccolo.

Vi sono poi modelli relativi all'evoluzione della cooperazione: questo fenomeno è tipicamente osservato nelle reti sociali, economiche e biologiche ed è rappresentato dall'interazione tra elementi della rete.

Anche le reti di collaborazione sono soggette ad evoluzione nel tempo: un modello tipico è costruito combinando le proprietà di attaccamento preferenziale e una struttura bipartita imposta.

Bibliografia

- Boccaletti, Stefano, et al. Complex networks: Structure and dynamics. *Physics reports* 424.4 (2006): 175-308.