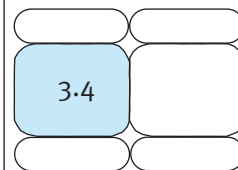


LA STRUTTURA DELLE RETI IN UN MONDO INTERCONNESSO



Andrea Bobbio

Tutti sono d'accordo nell'ammettere che ciò che caratterizza il mondo moderno è l'interdipendenza. Le strutture sociali, tecnologiche, economiche tendono sempre più ad aggregarsi e a connettersi. È possibile individuare leggi generali che regolano queste interconnessioni e ne determinano il comportamento? Questa domanda ha aperto una nuova scienza la "teoria delle reti complesse" che ha avuto uno sviluppo prepotente e impensabile negli ultimi anni.



1. INTRODUZIONE

Una caratteristica peculiare e generale dell'odierna organizzazione del vivere umano è data dalla interconnessione. Le strutture sociali, tecnologiche, economiche tendono sempre più ad aggregarsi, a connettersi. Componendo un numero sul nostro apparecchio telefonico possiamo parlare con qualunque altro utente del servizio telefonico in qualunque parte del mondo. L'energia elettrica viene capillarmente distribuita attraverso una rete, così come il gas e l'acqua. I trasporti sono basati su reti: reti stradali, ferroviarie, aeree. Esistono reti sociali di legami personali e di gruppo, reti di relazioni economiche. I sistemi biologici sono costituiti da cellule interagenti attraverso complesse reti di reazioni biochimiche. Ma forse l'evento che ha avuto di più influenza sulla nostra vita è stata l'apparizione di Internet e del *World Wide Web* (www). Da pochi anni i computer di tutto il mondo possono collegarsi fra di loro attraverso la rete fisica di Internet, mentre il www è una rete virtuale i cui nodi sono pagine o documenti che possono contenere

qualunque informazione: testi, immagini, filmati e suoni. La rete Internet sta diventando sempre più veloce e potente e il www sta crescendo a ritmi vertiginosi: centinaia o migliaia di nuove pagine vengono prodotte ogni giorno e sta diventando forse la creazione dell'uomo più complessa e multiforme. Come studiare questi sistemi complessi? È possibile trovare leggi che ne regolano la struttura, le forme di aggregazione, le proprietà? Nei secoli scorsi lo scopo primo dell'indagine scientifica era stato quello di decifrare i componenti base della natura perché si pensava che la comprensione del mondo passasse attraverso la conoscenza dei suoi costituenti primi: le particelle elementari, gli atomi, le molecole, le proteine, le cellule. Questo approccio ha portato a risultati di enorme valore, ma si è forse perso di vista l'insieme: come i mattoni elementari si combinano in complesse strutture coese. Si è perso di vista il fatto che il comportamento e le proprietà di sistemi costituiti da aggregati di parti elementari dipendono dai legami e dalle relazioni fra le parti. I neu-

rosenziati ci spiegano che la dicotomia fra cervello e mente, cioè fra l'ammasso di cellule che costituisce il cervello e il pensiero dipende dai legami fra le cellule. Scrive Rita Levi Montalcini [1] "Nessuna cellula nervosa, includendo tra queste anche quelle della corteccia cerebrale, è in grado di percepire, né essere capace di reazioni emotive, di pensare, ma tali proprietà sarebbero una funzione emergente dell'intero sistema corticale e dei sistemi neuronali del sistema stesso. Per questo motivo soltanto l'analisi di complessi cellulari e dei circuiti che li interconnettono, può portare un contributo allo studio dei processi mentali".

La consapevolezza dell'importanza dei legami fra le parti, l'impatto determinato dalla presenza, ubiquità e irrinunciabilità delle reti di calcolatori e la spinta derivante dal crescente interesse per i sistemi biologici e per le neuroscienze hanno portato all'insorgere di una nuova disciplina che viene normalmente identificata come "teoria delle reti complesse".

Gli studi per la comprensione delle strutture a rete complessa prendono le mosse da due lavori quasi contemporanei di Watts e Strogatz [2] sulle reti a "piccolo mondo" e di Barabasi e Albert [3] sulle reti a legge di potenza e vedono uno sviluppo incredibilmente fertile e rapido che coinvolge in uno sforzo congiunto fisici, matematici, informatici, biologi. Il presente lavoro intende fornire una prima lettura sugli indirizzi che hanno assunto questi recenti studi e sui principali risultati raggiunti. Per un avvicinamento più profondo alle pro-

blematiche legate alla teoria delle reti, si consiglia la lettura del libro di uno degli iniziatori di questo filone di ricerca A.L Barabasi [4]. Il libro è disponibile in una edizione italiana, è di facile comprensione per un lettore di normale cultura scientifica ed è letterariamente molto gradevole e attraente. Per chi fosse interessato ad approfondimenti più tecnici e dettagliati, riguardanti anche le metodologie di analisi e le tecniche matematiche di indagine, si rinvia a estesi lavori scientifici di rassegna sull'argomento [5, 6, 7, 8] e alle loro voluminose bibliografie. L'ultimo e più recente lavoro citato [8] rimanda a quasi mille riferimenti bibliografici!

2. LE GRANDEZZE CHE CARATTERIZZANO UNA RETE

Una rete è rappresentabile mediante un grafo $G = (V, E)$ dove V è l'insieme dei nodi (o vertici) e E è l'insieme degli archi che connettono i nodi (Figura 1 A). Gli archi possono essere non orientati e quindi possono essere percorsi nelle due direzioni (come ad esempio un collegamento fra due calcolatori sulla rete internet, o un collegamento telefonico fra due apparecchi, o i legami di amicizia fra due soggetti), oppure possono essere orientati.

Gli archi orientati possono essere percorsi in una sola direzione indicata normalmente da una freccia (Figura 1 B). Sono esempi di grafi orientati il reticolo fluviale di una regione (l'acqua scorre in una sola direzione), le reti di citazioni bibliografiche e il www. Se una pagina a del www contiene un puntatore

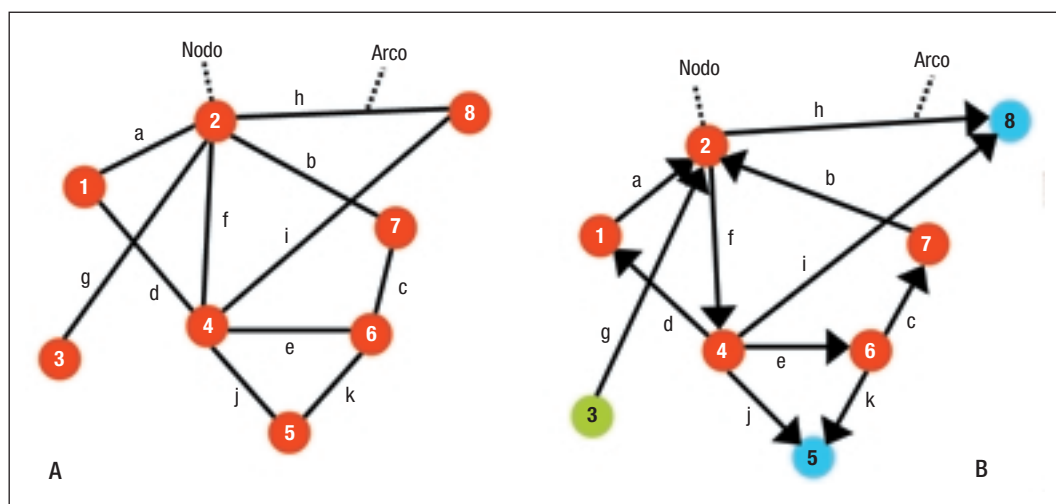


FIGURA 1
A Esempio di grafo non orientato;
B Esempio di grafo orientato

(link) ad una pagina b vuol dire che da a si può raggiungere b con un click del mouse ma non viceversa (a meno che nella pagina b non ci sia un link esplicito alla pagina a).

In una rete non orientata basta un solo arco per nodo per rendere il grafo completamente connesso, cioè per poter raggiungere qualsiasi nodo partendo da qualsiasi altro nodo. Al di sotto della soglia limite di un arco per nodo una rete si decompone in aggregati di nodi non raggiungibili reciprocamente. Le reti in natura superano normalmente il valore di soglia limite, in quanto hanno normalmente più di un arco emergente da ogni nodo: questa molteplicità di archi rende le reti molto più strettamente connesse e ridondanti. Per le reti orientate la connessione è più complessa, in quanto si possono identificare tre tipologie distinte di nodi. Nodi che formano insiemi connessi cioè in cui tutti i nodi sono raggiungibili fra loro seguendo la direzione degli archi (per esempio i nodi rossi nella Figura 1 B). I nodi sorgente che hanno solo archi in uscita e non sono raggiungibili da nessun altro nodo (per esempio il nodo verde in Figura 1 B). I nodi assorbenti che hanno archi in ingresso ma non in uscita: possono essere raggiunti ma una volta raggiunti non possono essere abbandonati (per esempio i nodi blu nella Figura 1 B).

Un grafo, quale quello di figura 1 A o figura 1 B, può avere anche una rappresentazione matematica mediante la cosiddetta matrice di incidenza. Se il grafo ha N nodi numerati da 1 a N , la matrice di incidenza è una matrice quadrata $N \times N$ il cui elemento ij è uguale a 1 se i e j sono direttamente connessi da un arco, altrimenti è uguale a 0. La i -esima riga della matrice di incidenza individua gli archi in uscita dal nodo i , mentre la i -esima colonna individua gli archi in ingresso al nodo i . Se il grafo è non orientato la matrice di incidenza è simmetrica (cioè l'elemento con indice di riga i e indice di colonna j è uguale all'elemento con indice di riga j e indice di colonna i). Se il grafo è orientato la matrice non è simmetrica: i nodi sorgente hanno la corrispondente colonna della matrice di incidenza tutta di zeri; i nodi assorbenti hanno la corrispondente riga della matrice di incidenza tutta di zeri.

Le reti molto complesse con migliaia, milio-

ni o miliardi di nodi, quali sono quelle con cui abbiamo a che fare quotidianamente, sono difficili da descrivere e impossibili da disegnare. Per definire e caratterizzare le proprietà di queste reti complesse sono state introdotte varie misure. Le misure maggiormente adottate [9], e che verranno riprese in questo lavoro (vedi riquadro a p. 6 per un esempio numerico) sono il *grado di connettività*, la *lunghezza del cammino caratteristico* e il *coefficiente di aggregazione (clustering)*.

2.1. Grado di connettività

Il *grado di connettività* di un nodo (o semplicemente il *grado* di un nodo) è il parametro più studiato perché facilmente discrimina le caratteristiche di una rete complessa. In una rete non orientata il grado di un nodo è dato dal numero di archi emergenti dal nodo (equivalente al numero di nodi direttamente raggiungibili dal nodo in esame, detti anche nodi primi vicini). In una rete orientata si distingue fra il grado in ingresso (in-degree) che conta il numero di archi in ingresso in un nodo e il grado in uscita (out-degree) che conta il numero di archi emergenti da un nodo. Limitandoci ora a reti non orientate si indichi con k_i il grado del nodo i , e con $P(k)$ la distribuzione del grado di connettività cioè la probabilità che un nodo abbia grado k (o più semplicemente la percentuale di nodi di grado k sul numero totale di nodi N). Il valore medio del grado $\langle k \rangle$ è ottenuto mediando il grado su tutti i nodi della rete. Sebbene il grado sia una proprietà locale di un nodo, la distribuzione del grado $P(k)$ è una delle principali caratteristiche che determina, come vedremo, comportamenti complessivi della rete.

2.2. Lunghezza del cammino caratteristico L

La *Lunghezza del cammino caratteristico* (o semplicemente il *cammino caratteristico*) L misura la tipica separazione fra due generici nodi del grafo ed è anche, a volte, indicata come il diametro del grafo. Dati due nodi i e j definiamo d_{ij} la loro distanza minima, cioè il numero minimo di archi che si devono attraversare per raggiungere il nodo j partendo dal nodo i (supponiamo in questo caso che il

Un esempio di misure di rete

Prendendo a riferimento l'esempio di rete non orientata riportata nel grafo di figura 1 A, vengono esplicitamente calcolate le misure caratterizzanti le reti complesse definite nel paragrafo 2 del presente lavoro.

Grado di connettività

Il grado di connettività di un nodo è dato dal numero di archi emergenti dal nodo. Nella tabella A viene riportato, con riferimento al grafo di figura 1 A, il numero di archi emergenti da ogni nodo e il suo grado: poiché ogni arco connette due nodi, il grado è dato dal No. Archi diviso due. Il calcolo della media dei valori della terza colonna della tabella A permette di ottenere il valore medio del grado $\langle k \rangle = 1.375$.

I valori della terza colonna della tabella A forniscono anche la distribuzione del grado di connettività $P(k)$.

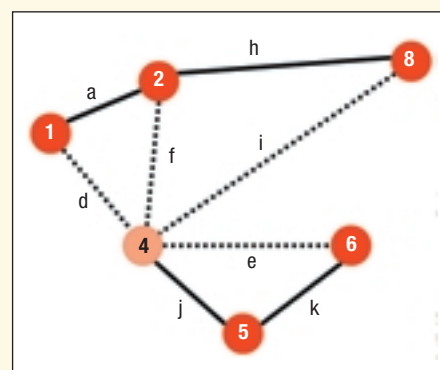
Lunghezza del cammino caratteristico L

Dato che la lunghezza del cammino caratteristico L è definita come la media delle distanze minime fra ogni coppia di nodi, occorre valutare dal grafo la distanza minima, espressa come numero minimo di archi che bisogna attraversare per connettere i due nodi.

Per esempio, prendendo come sorgente il nodo 3 del grafo di figura 1 A, le distanze minime nei confronti di tutti gli altri nodi sono riportate nella tabella B.

TABELLA A		
Nodo	No. Archi	Grado
1	2	1
2	5	2.5
3	1	0.5
4	5	2.5
5	2	1
6	3	1.5
7	2	1
8	2	1

TABELLA B		
Nodo sorgente	Nodo destinazione	Distanza minima
3	1	2
3	2	1
3	4	2
3	5	3
3	6	3
3	7	2
3	8	2



Il sottografo G_4

Calcolando le distanze minime per tutti i nodi e facendo la media si ottiene la *Lunghezza del cammino caratteristico* $L = 1.68$.

Coefficiente di aggregazione (clustering coefficient) C
Nel grafo di figura 1 A, si consideri il nodo 4 e si disegni il sottografo G_4 ottenuto dal grafo completo G prendendo i primi vicini del nodo 4 e eliminando il nodo e gli archi da esso emergenti (vedi Figura).

Poiché i primi vicini del nodo 4 sono i nodi 1, 2, 5, 6 e 8, il sottografo G_4 ha 5 nodi e al massimo (se tutti i nodi fossero connessi) $5 \cdot 4/2 = 10$ archi. Di questi 10 archi potenziali il sottografo G_4 ne contiene 4 da cui il coefficiente di aggregazione relativo al nodo 4 diventa $C_4 = 4/10 = 0.4$.

La tabella C riporta i coefficienti di aggregazione relativi ad ogni nodo da cui si calcola il coefficiente di aggregazione medio $C = 0.492$.

TABELLA C		
Nodo	Primi vicini	Coefficiente aggregazione
1	2	1
2	5	0.2
3	1	0
4	5	0.4
5	2	1
6	3	0.333
7	2	0
8	2	1

grafo sia non orientato e connesso cioè che partendo da ogni nodo sia possibile raggiungere qualunque altro nodo). Quindi, se i e j sono direttamente connessi da un arco avre-

mo $d_{ij} = 1$, altrimenti $d_{ij} > 1$. Si definisce la lunghezza del cammino caratteristico L come la media delle distanze minime fra ogni coppia di nodi. Una rete in cui tutti i nodi sono con-



nessi a tutti gli altri nodi ha una matrice di incidenza con tutti gli elementi ij fuori dalla diagonale principale ($i \neq j$) uguali a uno: in questo caso il cammino caratteristico è $L = 1$, il numero totale di archi è $N(N-1)/2$ e il grado è costante pari a $k = (N-1)/2$. Se $\langle k \rangle$ è il valor medio del grado, da ogni nodo si possono mediamente raggiungere in un solo passo $\langle k \rangle$ nodi, in due passi $\langle k \rangle^2$ nodi ecc... e in L passi si ricopriranno tutti gli N nodi. Per cui avremo $N \sim \langle k \rangle^L$ da cui $L \sim \ln N / \ln \langle k \rangle$. Il valore di L è normalmente un numero piccolo che cresce lentamente con il numero totale di nodi N . Questa proprietà, per cui la distanza media fra nodi è un numero sovente molto basso, viene normalmente riferita come proprietà di “piccolo mondo”.

2.3. Coefficiente di aggregazione (clustering coefficient) C

Dato un nodo i , si definisce il sottografo G_i ottenuto dal grafo G prendendo i primi vicini di i e eliminando il nodo i e gli archi da esso emergenti. Se il nodo i ha grado k_i , allora G_i ha k_i nodi e, al massimo, $k_i(k_i-1)/2$ archi. Il coefficiente di aggregazione C_i relativo al nodo i , misura il numero di archi realmente presenti nel sottografo G_i rispetto al numero di archi teoricamente possibili; misura, cioè, quanto i primi vicini del nodo i sono connessi fra loro rimuovendo i . Se un nodo a è connesso ai nodi b e c , il coefficiente di aggregazione va a considerare se b e c sono anche connessi fra di loro (in termini di reti sociali si usa esprimere questa proprietà dicendo che il coefficiente di aggregazione misura se i miei amici sono anche amici fra di loro). Il coefficiente di aggregazione C è il valor medio di C_i calcolato su tutti i nodi del grafo.

3. LE RETI CASUALI

Le prime reti complesse ad essere state sistematicamente studiate furono le reti casuali introdotte da due matematici ungheresi Erdős e Rényi [4] attorno agli anni '60. In una rete casuale di N nodi, ogni coppia di nodi ha la stessa probabilità p di essere connessa. Quindi quando p è prossimo a zero i nodi sono prevalentemente isolati, quando p tende a 1 i nodi tendono a essere connessi a tutti gli altri nodi. Si può dimostrare che quando N è

grande, il grado medio di connettività dei nodi è $\langle k \rangle = Np$ e la distribuzione del grado $P(k)$ è approssimata da una legge di Poisson.

$$P(k) = (\langle k \rangle^k / k!) e^{-\langle k \rangle} \quad (1)$$

La legge di Poisson ha una forma a campana con un massimo centrato sul valor medio $\langle k \rangle$. Un esempio di rete con distribuzione del grado di tipo Poissoniano (1) è la rete autostradale negli USA, rappresentata nella parte sinistra di figura 4. Gli archi sono le strade e i nodi i loro incroci (prevalentemente nelle città). In una rete Poissoniana, i nodi hanno come grado più probabile il valore $\langle k \rangle$, e la probabilità del valore del grado diminuisce rapidamente (esponenzialmente) al discostarsi di k dal valore medio $\langle k \rangle$ nelle due direzioni. In una rete casuale il cammino caratteristico L tende a essere piccolo e a diminuire con l'aumentare di p ; anche il coefficiente di aggregazione tende ad essere piccolo perchè il fatto che un nodo a sia connesso a due nodi b e c non aumenta la probabilità che b sia connesso a c . In una rete casuale il coefficiente di aggregazione può essere espresso come $C = \langle k \rangle / N$. Tuttavia, le reti reali sono tutt'altro che casuali! In una rete casuale di Erdős e Rényi la probabilità che il mio computer sia connesso al router del mio ufficio sarebbe uguale a quella di essere connesso a qualunque altro computer del mondo oppure la probabilità di un individuo di conoscere, e quindi stabilire un legame con il vicino di casa sarebbe la stessa della probabilità di conoscere uno sciamano della Papuasiasia. Nessuno pensa che ciò sia vero. Poiché lo scopo fondamentale dell'indagine scientifica è quello di identificare modelli per descrivere la realtà, è necessario introdurre nuove ipotesi e nuovi ragionamenti per l'interpretazione e lo studio delle reti reali.

4. LE RETI A PICCOLO MONDO

Nel 1967 il sociologo americano Stanley Milgram mise in piedi un esperimento per studiare la “distanza” fra due cittadini qualunque degli Stati Uniti [10]. La domanda a cui si voleva dare risposta era: scelte due persone a caso, quanti contatti sono necessari

per connetterle? L'esperimento consisteva nel selezionare casualmente un gruppo di individui e nel consegnare a ciascuno di loro una lettera con indicato un destinatario finale. Ogni soggetto, se non conosceva direttamente il destinatario finale, doveva inoltrare il documento ad una sola persona, a lui personalmente nota, e che riteneva avesse maggiore opportunità di raggiungere il destinatario finale. Si ipotizzava che questa catena di contatti richiedesse almeno un centinaio di passaggi. Alla fine dell'esperimento Milgram registrò che su centosessanta individui iniziatori, quarantaquattro catene erano arrivate alla conclusione e che il numero medio di passaggi era sorprendentemente basso: 5.5 passaggi. È ciò che i sociologi chiamano i sei gradi di separazione. Qualunque persona nel mondo può essere raggiunta mediamente attraverso sei passaggi. Il mondo è un "piccolo mondo". Si noti che nell'esperimento di Milgram i cammini ottenuti non erano necessariamente minimi: un individuo in una catena poteva non sapere che il suo vicino di casa era il cugino del destinatario finale. Tuttavia l'esperimento mostrava un altro aspetto estremamente interessante: gli individui riescono a trovare percorsi brevi (anche se non minimi) orientandosi puramente sulla base delle conoscenze locali della rete e senza nessuna visione globale.

Le reti reali hanno normalmente un coefficiente di aggregazione più elevato di quanto previsto dalla teoria delle reti casuali, in quanto è ragionevole ipotizzare che due nodi che hanno un qualche legame con un terzo nodo possano avere anche un legame fra di loro. È facile, inoltre, immaginare che il cammino caratteristico diminuisca e il grado di aggregazione cresca quando la rete si rivolge a gruppi di persone con interessi omogenei.

Nato da un quiz televisivo americano, è stato sviluppato un curioso sito internet che misura il "numero di Kevin Bacon", cioè la distanza (minima) degli attori cinematografici di tutto il mondo da Kevin Bacon. Si riteneva (supposizione poi dimostratasi non del tutto fondata [4]) che Kevin Bacon fosse uno degli attori di Hollywood che avesse recitato in più film con il maggior numero di colleghi attori, e si intende-

va misurare quanto gli altri attori fossero "distanti" da Kevin Bacon. Indicando il numero di Kevin Bacon con il simbolo NKB , un attore che abbia recitato in un film con Kevin Bacon ha $NKB = 1$; un attore che abbia recitato in un film in cui era presente un attore che avesse recitato in un film con Kevin Bacon ha $NKB = 2$ e così via. Il sito internet che permette di misurare NKB per tutti gli attori cinematografici e che fornisce anche la catena minima di passaggi con l'indicazione degli attori e dei film intermedi è stato sviluppato, e viene costantemente aggiornato, da due studenti dell'Università della Virginia a partire dal 1996 e ha indirizzo <http://www.cs.virginia.edu/oracle/>.

Per due attori italiani sulla cresta dell'onda al momento, Giovanna Mezzogiorno e Roberto Benigni, il sito di Kevin Bacon ha prodotto i seguenti risultati:

Giovanna Mezzogiorno has a Bacon number of 3.

1. Giovanna Mezzogiorno was in *Ilaria Alpi - Il più crudele dei giorni* (2002) with Tony Lo Bianco
2. Tony Lo Bianco was in *City Heat* (1984) with Beau Starr
3. Beau Starr was in *Where the Truth Lies* (2005) with Kevin Bacon

Roberto Benigni has a Bacon number of 2.

1. Roberto Benigni was in *Pinocchio* (2002) with Queen Latifah
2. Queen Latifah was in *Beauty Shop* (2005) with Kevin Bacon

Attivando il sito con qualunque attore si osservano valori di NKB sempre di poche unità. La comunità degli attori è veramente un piccolo mondo, per cui è stata stimata una distanza media di tutti gli attori da Kevin Bacon di $L = 2,8$ e di tutti gli attori fra loro di $L = 3,6$. Le reti reali hanno la proprietà di piccolo mondo riscontrate anche nelle reti casuali di Erdős e Rényi, ma rispetto alle reti casuali hanno coefficiente di aggregazione molto più elevato. Il primo tentativo di spiegare la concomitante presenza di piccolo mondo e grande aggregazione fu proposto da Watts e Strogatz in un lavoro [2] che ebbe un grandissimo impatto sulla comunità scientifica. Per illustrare l'insorgere di una rete a piccolo mondo Watts e Strogatz hanno proposto un modello (che chiamo-

remo modello WS) in cui la variazione di un singolo parametro (p) permette di passare con continuità da una struttura a reticolo regolare ad una struttura a grafo casuale. Nella figura 2 è rappresentato un grafo su un reticolo regolare di N nodi disposti su una circonferenza e ogni nodo ha lo stesso grado k . Nella figura 2 si è rappresentato il caso $N = 16$ e $k = 2$ (ogni nodo è connesso con i suoi due primi vicini sia a destra che a sinistra, ma gli archi sono bidirezionali e vanno contati una volta sola).

Con questa struttura il coefficiente di aggregazione C risulta indipendente dalla dimensione del grafo mentre la lunghezza del cammino caratteristico L risulta proporzionale al numero di nodi N e al grado k , e cresce al crescere di N , perdendo la caratteristica di piccolo mondo. La struttura regolare abbina quindi elevato cammino caratteristico L e elevato grado di aggregazione C .

Il modello WS prevede ora di assegnare un valore di probabilità $p > 0$, chiamata probabilità di riassegnamento; con probabilità p si decide se ritracciare ogni arco, sconnettendolo dalla vecchia locazione e ridisegnandolo in una nuova locazione scelta a caso. Nascono così, in modo casuale, archi di "corto circuito" che attraversano il grafo (Figura 2 B). Con questa procedura di riassegnamento degli archi, quando p cresce la rete tende a diventare sempre più disordinata e per p che tende a 1 il modello WS produce una rete casuale (Figura 2 C). Con questo procedimento, Watts e Strogatz hanno dimostrato che con piccoli valori di p , cioè con un numero anche molto limitato di archi casuali, la rete cambia regime e il suo cammino caratteristico L decade rapidamente pur mantenendo quasi inalterato il valore del coefficiente di aggregazione C . Bastano pochi archi casuali in un reticolo regolare, anche di grandissime dimensioni, per far precipitare il valore del cammino caratteristico L e far apparire l'effetto di piccolo mondo, pur mantenendo un elevato coefficiente di aggregazione caratteristico delle strutture regolari.

Il fatto che pochi archi casuali possano modificare in modo sostanziale la connettività delle reti e ridurre drasticamente il cammino caratteristico può anche essere visto intuitivamente in modo semplice. Siano A , B e C tre reti a piccolo mondo separate fra loro. Tutti i nodi all'interno di A possono essere raggiun-

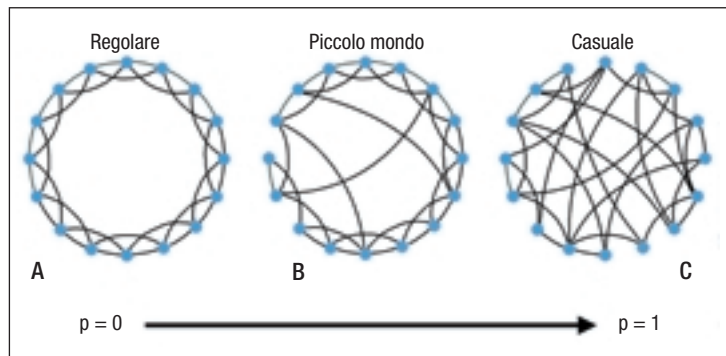


FIGURA 2

Il modello di Watts e Strogatz (WS) al crescere della probabilità di riassegnamento p . A) $p = 0$, reticolo regolare; B) p piccolo, nascita di archi casuali con l'effetto piccolo mondo; C) $p = 1$, grafo casuale

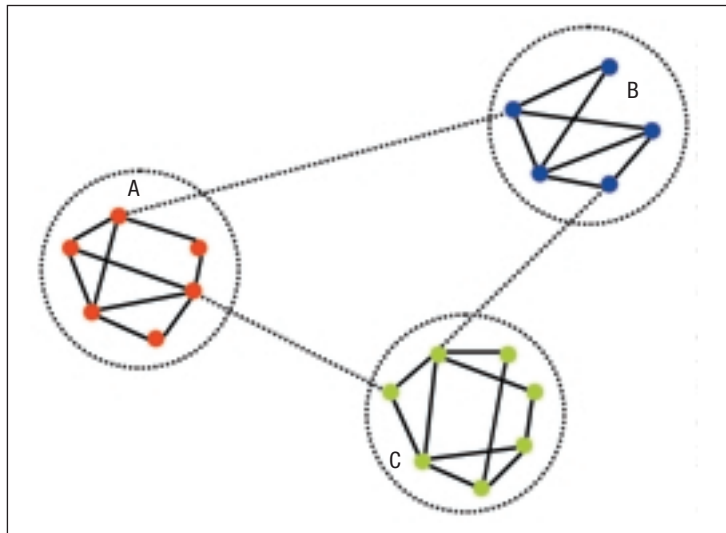


FIGURA 3

Tre reti a piccolo mondo A , B e C connesse da pochi link

ti in pochi passi, tutti i nodi all'interno di B possono essere raggiunti in pochi passi e tutti i nodi all'interno di C possono essere raggiunti in pochi passi (A , B e C possono essere reti locali in una rete di calcolatori, o comunità separate nel caso di reti sociali). È sufficiente che un solo nodo di A sia connesso con un solo nodo di B e B con un solo nodo di C per far diventare un piccolo mondo l'unione delle reti A , B e C e permettere ad ogni nodo di un colore di raggiungere in pochi passi un nodo di qualunque altro colore.

La rete Internet (vedi riquadro a p. 14) presenta una struttura parzialmente ispirata a quella di figura 3, in cui sotto-reti locali sono connesse fra loro mediante collegamenti ad alta capacità.

5. LE RETI IN EVOLUZIONE E LE LEGGI DI POTENZA

Il crescente interesse per le reti complesse ha indotto vari ricercatori ad indagare sulla struttura di varie reti reali in ambiti molto diversi: informatico, tecnologico, sociale, biologico, economico. Questi studi hanno mostrato che molte reti reali presentano una distribuzione del grado di connettività $P(k)$ ben lontana dalla forma a campana prevista dalle legge di Poisson generata dalle reti casuali. I lavori di rassegna sull'argomento [5, 6, 7, 8], già citati, riportano tabelle riepilogative di proprietà di reti da cui emerge che un significativo numero di reti reali (www, internet, citazioni scientifiche nel database ISI, reti metaboliche, reti di proteine) ha una distribuzione del grado $P(k)$ che segue una legge di potenza (almeno per alti valori di k) rappresentabile dalla seguente espressione:

$$P(k) \sim c k^{-\gamma} \quad (2)$$

Nella formula (2) c è una costante di proporzionalità il cui valore è solitamente di scarso interesse pratico (vedi riquadro a p. 11 per maggiori dettagli) e γ è il coefficiente della legge di potenza. Dagli studi citati emerge che il valore di γ è normalmente compreso fra 2 e 3. Le leggi di potenza erano già note da molti anni nella

statistica e nella letteratura tecnica dove erano, a volte, chiamate leggi di Pareto o di Zipf. Già l'economista e sociologo Vilfredo Pareto (1848-1923) aveva scoperto, infatti, che la distribuzione del reddito degli individui segue una legge di potenza. Come illustrato nel quadro esplicativo B, una legge di potenza disegnata su un grafico con scala logaritmica su entrambi gli assi, viene rappresentata da una retta la cui pendenza è legata all'esponente γ .

In estrema sintesi le leggi di potenza sono caratterizzate da una "lunga coda", cioè la funzione $P(k)$ tende a zero lentamente (secondo la potenza $-\gamma$) al crescere di k . Il lento decadimento di $P(k)$ in funzione di k fa sì che ci sia una probabilità non trascurabile di trovare valori di k grandi e molto grandi. Quindi una rete con distribuzione del grado del tipo rappresentato nella formula (2) contiene un gran numero di nodi con grado piccolo insieme a pochi nodi con grado molto alto, cioè con un numero molto elevato di archi. Questi nodi, detti connettori (o *hub*), sono gli elementi che maggiormente influenzano la topologia e le proprietà della rete. Il percorso minimo fra due nodi qualunque passa normalmente attraverso i connettori. Una rappresentazione molto intuitiva, esplicita e significativa delle reti a legge di potenza è data dalle reti delle linee aeree. Nella figura 4 B è sche-

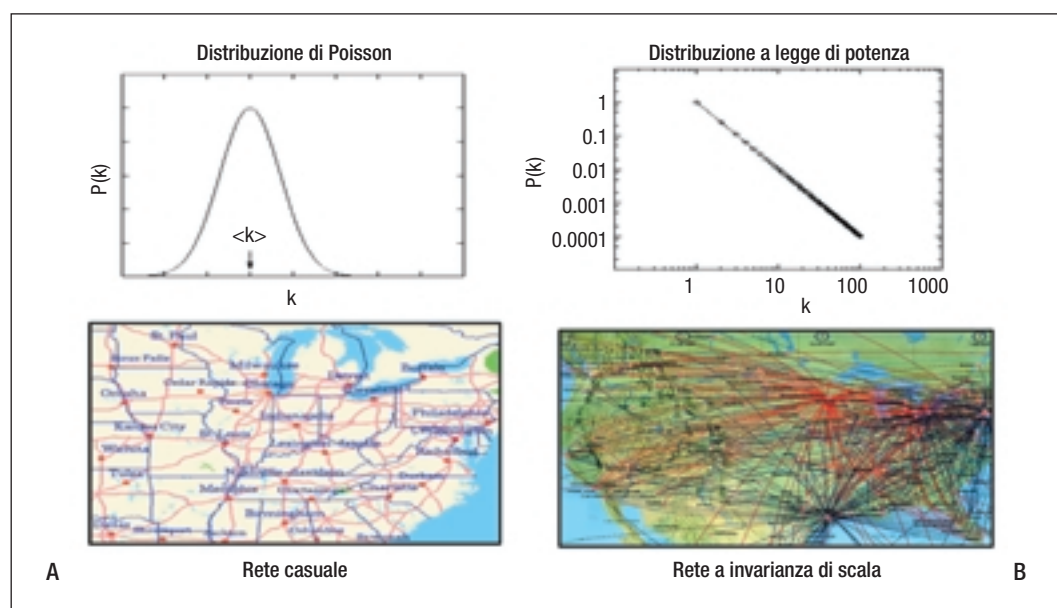


FIGURA 4

A) mappa delle principali autostrade negli USA: un esempio di rete con distribuzione del grado di tipo Poisson.
B) mappa delle principali linee aeree nazionali negli USA: un esempio di rete con distribuzione del grado a legge di potenza

Distribuzioni a legge di potenza

Le distribuzioni a legge di potenza costituiscono una famiglia di distribuzioni caratterizzate dall'aver una "lunga coda" cioè la funzione di probabilità tende a zero lentamente all'aumentare del valore della variabile. Questa caratteristica fa sì che le leggi di potenza forniscano probabilità non trascurabili di riscontrare valori della variabile anche estremamente elevati. Le leggi di potenza pur essendo note e utilizzate nella letteratura tecnica e nella statistica da parecchi decenni, hanno avuto una ripresa di interesse sia per effetto delle ricerche legate alle reti complesse, sia, più in generale, perché si è riscontrato che permettono di interpretare innumerevoli fenomeni fisici e serie di dati sperimentali di natura e origine molto diversa. Per un approfondimento, si rimanda a due recenti lavori di rassegna sull'argomento [1, 2] molto chiari ed esaurienti, e che tracciano anche una illuminante prospettiva storica sull'origine e sull'uso di queste leggi.

Per dimostrare la versatilità delle leggi di potenza, il lavoro [2] cita 12 esempi semplici e reali di grandezze di origine molto diversa fra loro che seguono una legge di potenza, almeno verso la coda della distribuzione. Gli esempi sono: ricorrenza di parole, citazioni in lavori scientifici, accessi a pagine web, copie di libri venduti, chiamate telefoniche, grandezza di terremoti, diametro dei crateri lunari, intensità delle esplosioni solari, intensità delle guerre, reddito netto dei cittadini americani, popolazione delle città negli USA.

Il primo studioso che individuò che l'ammontare dei patrimoni seguiva una legge di potenza fu l'economista e sociologo Vilfredo Pareto (1848-1923) per cui queste leggi sono a volte note come leggi di Pareto, mentre il linguista G.K. Zipf (1902,1950) introdusse metodi statistici nello studio del linguaggio, e trovò leggi di potenza per la ricorrenza delle parole. Una bibliografia aggiornata sulla legge di Zipf (e sulle leggi di potenza) è mantenuta nel sito <http://www.nslj-genetics.org/wli/zipf/index.html>: il primo lavoro citato è del 1881 e la bibliografia contiene ad oggi 555 riferimenti.

Una legge di potenza è caratterizzata dall'aver una densità di probabilità $P(x)$ (dove $P(x)$ rappresenta la probabilità di trovare la variabile compresa fra x e $x + dx$) data da:

$$P(x) = cx^{-\gamma} \quad (1)$$

Dove $\gamma > 0$ è l'esponente che caratterizza la legge di potenza e c è una costante di normalizzazione che non ha rilevanza nel determinare il comportamento della funzione e viene calcolata in modo che l'area sottesa dalla funzione di densità $P(x)$ sia uguale a 1 (essendo $P(x)$ una densità di probabilità la sua area totale deve essere uguale a 1). La legge (1) deve essere definita per un valore di x maggiore di un valore minimo x_{\min} (per $x = 0$ la funzione $P(x)$ tende all'infinito). Si suole quindi affermare che una variabile segue una legge di potenza a partire da un certo valore minimo, cioè quando è la coda della distribuzione a soddisfare la legge (1).

Il valor medio della variabile $\langle x \rangle$ è finito solo per $\gamma > 2$ (per $\gamma \leq 2$ è infinito); la varianza diverge per $\gamma \leq 3$. La distribuzione mantiene le proporzioni al variare della scala. Se in un disco di un computer i file da 2 KB sono il 25% dei file da 1 KB, la legge predice che i file da 2 MB saranno il 25% dei file da 1 MB. La mancanza di un picco attorno al valore medio e il mantenimento delle proporzioni al variare della scala hanno indotto a definire questa legge come legge a invarianza di scala (*scale free*).

Prendendo il logaritmo naturale di entrambi i membri nella formula (1) si ottiene $\ln P(x) = -\gamma \ln x + \ln c$; quindi in un diagramma log-log la legge è rappresentata con una retta a pendenza negativa e il coefficiente angolare della retta è il valore dell'esponente γ (Figura 4 B).

Quali sono alcune possibili conseguenze pratiche di una legge di potenza.

Esempio 1 - Il reddito netto dei cittadini americani segue una legge di potenza con $\gamma = 2.1$ [2]. Questo comporta che il 50% della popolazione più ricca detenga il 94% del reddito nazionale, e che 80% del reddito nazionale sia nelle mani del 20% degli individui più ricchi. Questa è la regola "80/20" attribuita a Pareto.

Esempio 2 - L'altezza degli esseri umani segue tipicamente una legge a campana con valor medio attorno a $\langle x \rangle = 1.75$ m, quindi una legge molto lontana da una legge di potenza. Se l'altezza degli umani seguisse una legge di potenza con esponente $\gamma = 2.5$ un individuo ogni 100.000 sarebbe alto 15 m, e uno ogni milione sarebbe alto 30 m.

RIFERIMENTI BIBLIOGRAFICI

- [1] Mitzenmacher M: A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, Vol. 1, 2004, p. 226-251.
[2] Newman M.E.: Power laws, Pareto distributions and Zipf's laws. *Contemporary Physics*, Vol. 46, 2005, p. 323-351.

maticamente riportata la rete dei principali collegamenti aerei nazionali negli USA. Alcuni aeroporti fungono da connettori (e sono effettivamente chiamati *hub* in inglese); per viaggiare da un aeroporto ad un altro, in mancanza di un collegamento diretto, si deve effettuare una coincidenza in un *hub*.

Il grafico a destra in alto nella figura 4 riporta la distribuzione del grado $P(k)$ per la rete delle linee aeree in scala bilogarithmica. È evidente, anche ad una prima e superficiale ispezione visiva, la differente struttura della rete autostradale (Figura 4 A) e della rete delle linee aeree (Figura 4 B).

Le reti a legge di potenza sono frequenti in natura e nella realtà. Perché si formano reti

con questa struttura e quale è il meccanismo responsabile della loro generazione? Quali sono le proprietà principali indotte da una topologia a legge di potenza e perché questa struttura organizzativa è spesso privilegiata rispetto ad altre? La risposta a queste domande venne affrontata per la prima volta in maniera semplice e brillante da Barabasi e Albert in [3]. La loro spiegazione parte da una semplice constatazione: le reti complesse sono sistemi aperti che evolvono nel tempo crescendo di dimensione in quanto nuovi nodi si aggiungono man mano a quelli preesistenti. Sia i modelli casuali di Erdős e Rényi che quelli a piccolo mondo di Watts a Strogatz sono modelli statici, pen-

sati per un numero fisso di nodi per i quali si cercano opportune regole per il tracciamento degli archi. Nelle reti in evoluzione nuovi nodi si generano continuamente e i nuovi nodi devono connettersi ai nodi preesistenti sulla base di opportune regole. Gli autori fanno l'ipotesi che il meccanismo di connessione dei nuovi nodi al resto della rete sia ispirato al principio del "aggancio preferenziale (*preferential attachment*)", cioè i nuovi nodi tendono preferenzialmente a connettersi ai nodi con grado più alto. Gli esempi che possono ispirare un meccanismo di aggancio preferenziale sono molteplici. Se si apre un nuovo aeroporto è ragionevole pensare che i primi collegamenti aerei che verranno resi operativi saranno verso gli aeroporti principali della regione (gli *hub*). In un articolo scientifico è facile che venga citato un lavoro che ha avuto notevole impatto sulla comunità scientifica e che ha già avuto molte citazioni. I link che verranno indicati su una pagina *www* di nuova generazione, saranno con grande probabilità rivolti verso i siti di maggiore frequentazione.

Partendo da questi due semplici presupposti (reti in evoluzione e aggancio preferenziale), Barabasi e Albert simularono la crescita di una rete virtuale, e, con ragionevole e legittima soddisfazione, trovarono conforto alle loro supposizioni, in quanto al crescere del numero di nodi immessi nella rete la distribuzione del grado $P(k)$ assumeva la forma di una legge di potenza con esponente $\gamma = 3$.

Le reti generate dal modello di Barabasi e Albert (modello BA) vennero chiamate dagli autori reti a *invarianza di scala* (*scale free*). Nelle reti casuali, in cui la distribuzione del grado segue una curva a campana, è definito un grado medio $\langle k \rangle$ che determina la *scala* della rete, cioè determina il numero di link normalmente presenti nei nodi essendo rare le forti deviazioni da questo valore. Nelle reti a legge di potenza, la curva di distribuzione non presenta un picco, la maggioranza dei nodi ha pochi archi emergenti, e questi innumerevoli piccoli nodi convivono con i grandi *hub*. Nelle distribuzioni a legge di potenza il valore medio (e la varianza) possono anche risultare infiniti per piccoli valori dell'esponente (vedi riquadro esplicativo di p. 11). Sia le reti di Erdős e Rényi sia quelle a piccolo mondo di Watts a Strogatz presu-

ponevano una struttura democratica della rete nel senso che il legame era ripartito in modo sostanzialmente uniforme sui nodi; le reti a invarianza di scala presuppongono un comportamento dei nodi fortemente disomogeneo, i pochi nodi connettori tendono ad aumentare i loro link durante l'evoluzione della rete mentre per gli altri nodi il numero di link rimane basso e sostanzialmente costante. Il sito *www* creato da Barabasi alla Notre Dame University riporta numerosi e suggestivi esempi di reti complesse <http://www.nd.edu/~networks/> con struttura a invarianza di scala.

Una rete BA a invarianza di scala può essere facilmente costruita attraverso il seguente algoritmo di generazione:

1. aggiunta di un nodo - partendo da un numero predefinito (piccolo) di nodi m_0 , ad ogni intervallo di tempo si aggiunge un nuovo nodo con $m \leq m_0$ archi che devono essere connessi ai nodi preesistenti della rete;

2. aggancio preferenziale - la probabilità che un arco del nuovo nodo sia connesso ad un preesistente nodo i dipende dal grado k_i secondo la legge $q_i = k_i / \sum_j k_j$ (vengono favoriti gli agganci ai nodi con grado k_i elevato).

Dopo t passi l'algoritmo ha generato una rete di $N = m_0 + t$ nodi e mt archi. È stato dimostrato che la presenza di entrambi gli ingredienti (crescita e aggancio preferenziale) è necessaria e sufficiente: la mancanza di uno dei due fa perdere la proprietà della legge di potenza. Le reti a invarianza di scala secondo il modello BA hanno piccola lunghezza caratteristica (il passaggio attraverso gli *hub* connette rapidamente tutti i nodi) ma hanno anche piccolo coefficiente di aggregazione normalmente minore di quello riscontrato nelle reti reali: i nodi periferici sono connessi agli *hub* ma non sono connessi fra di loro.

Sebbene il modello BA sia piuttosto semplicistico e rozzo (i nodi possono solo nascere e non morire, e i nodi più vecchi sono favoriti nel diventare gli *hub* della rete), tuttavia ha avuto una grandissima risonanza, in quanto mostrava che era possibile ipotizzare delle leggi di *auto-organizzazione* nei sistemi fisici complessi e queste leggi di auto-organizzazione portavano naturalmente a generare leggi di potenza. Sulla base di questi primi significativi risultati, vari ricercatori e vari gruppi di ricerca si misero a studiare leggi auto-

organizzative più rispondenti al mondo fisico, secondo le seguenti direttrici:

- i nodi e gli archi possono essere ridiretti e possono anche scomparire (si modifica o si elimina un link, si chiude una pagina web);
- i nodi possono invecchiare modificando la loro capacità di attrazione (sono state proposte varie leggi di aggancio preferenziale più complesse di quella BA originaria);
- i nodi competono fra di loro per attrarre i link dei nodi di nuova generazione. I nodi hanno un coefficiente di competitività o di buona salute (*fitness*) che ne esalta il potere di attrazione. Il modello a *fitness* introduce un meccanismo per cui non sempre i nodi più vecchi sono i più favoriti nella formazione degli hub, e si è reso necessario per spiegare fenomeni tipo Microsoft o Google, cioè di attori che entrati sulla scena per ultimi riescono a

imporre molto rapidamente a scapito di nodi più vecchi;

- la rete cresce con un meccanismo di aggancio preferenziale, ma i nodi possono essere attivi o non-attivi e passano da una condizione all'altra attraverso opportune leggi di probabilità; utilizzando questo accorgimento Klemm e Eguíluz [11] hanno ottenuto reti a invarianza di scala con alto un coefficiente di aggregazione;
- il modello BA genera reti non orientate, mentre varie reti del mondo reale sono orientate (ad esempio il www per cui il modello BA era stato inizialmente pensato).

Un esempio tipico di rete con crescita auto-evolutiva, e che anzi è stata quella che più di altre ha spinto gli studiosi ad affrontare lo studio delle reti complesse è il www. La struttura del www come rete complessa è richiamato nel riquadro. La rete Internet è pure una

La struttura della rete World Wide Web

La rapida diffusione del World Wide Web ha già fortemente inciso sull'organizzazione della società moderna e il suo impatto è certamente destinato a crescere. L'immagazzinamento, l'acquisizione e la trasmissione dell'informazione tende sempre più ad essere delegata al www soppiantando tutte le forme precedenti, in particolare quelle cartacee. I motori di ricerca hanno accesso ad un numero di documenti enorme e impensabile fino a pochi anni fa; i criteri di selezione dei documenti sono ancora rozzi (la corrispondenza di sequenze di lettere o parole) ma sono comunque incomparabilmente più potenti di qualunque strumento precedente.

Il www ha un'altra particolarità: è una rete totalmente auto-organizzata e con un tasso di crescita straordinariamente veloce: non c'è nessun supervisore centrale che ne controlli lo sviluppo e ne gestisca l'evoluzione. Il www può essere considerata una struttura paradigmatica per reti auto-evolutive e la comprensione della sua struttura ha originato e fortemente influenzato lo studio delle reti complesse. Quale è quindi la struttura della rete che chiamiamo www?

Innanzitutto il www è una rete orientata, in quanto i link inclusi in una pagina sono dei puntatori direzionali: permettono con un click del mouse di passare dalla pagina corrente a quella puntata dal link. Non è quindi definibile un unico grado di connettività ma si dovrà distinguere fra grado di ingresso (in-degree) e grado di uscita (out-degree) caratterizzati da due distinte distribuzioni $P_{in}(k)$ e $P_{out}(k)$. $P_{in}(k)$ fornisce la probabilità che k link puntino ad una certa pagina, mentre $P_{out}(k)$ fornisce la probabilità che una pagina abbia k link in uscita. È stato verificato sperimentalmente che entrambe le distribuzioni appartengono alla classe delle distribuzioni a legge di potenza con un esponente γ_{in} leggermente inferiore a γ_{out} [6]. La discrepanza fra i due valori può essere ragionevolmente spiegata dal diverso processo evolutivo dei gradi di ingresso e uscita. Il grado in ingresso non ha nessun estremo superiore, dipende da quante altre pagine decidono di indirizzare la pagina data. Pagine molto popolari (per esempio i più noti motori di ricerca) sono indirizzati da un numero enorme e crescente di pagine. Il grado in uscita è invece determinato dal numero di link esplicitamente posizionati su una pagina, che sono definiti dal proprietario della pagina stessa; in questo caso c'è anche un limite fisico al numero di link scrivibili in uno spazio limitato.

Non si conosce il numero esatto di pagine del www, ed è praticamente impossibile contarle tutte in modo rigoroso. Attualmente dovrebbe superare abbondantemente parecchi miliardi (una stima del 2000 parlava di 800 milioni di pagine). Ciò nonostante, il www è un piccolo (abbastanza piccolo) mondo: misure di cammino caratteristico fatte sperimentalmente su porzioni ristrette del www e estrapolate all'intera rete portano a stime di 19 passi di separazione. Il numero medio di link per pagina (cioè il grado medio) di uscita è circa 7. Con questi valori, in diciannove passi si potrebbero raggiungere circa $7^{19} \approx 10^{16}$ pagine, un numero sicuramente molto superiore al numero di pagine realmente esistenti.

Riprendendo la classificazione dei nodi per grafi orientati di figura 1 B, si intuisce che anche il www può essere suddiviso in tre aree principali. Un'area centrale che contiene pagine tutte connesse fra loro. In questa zona si possono raggiungere tutte le pagine saltando da una pagina all'altra attraverso un click del mouse. È la zona centrale del web e quella più facilmente raggiungibile dai motori di ricerca e quindi accessibile a tutti i navigatori.

Esiste poi una zona che chiameremo *Zona-in* in cui le pagine sono, in parte connesse fra di loro, e connesse alla zona centrale ma non viceversa. Sono pagine normalmente inaccessibili ai motori di ricerca. Quando un individuo costruisce una nuova pagina web, può indicare tutti i link che vuole sulla sua pagina, ma se nessun link di nessun'altra pagina punta a lui, rimane nella *Zona-in*, invisibile al resto della rete: non basta creare una pagina per entrare nel gran gioco del www. Alcuni motori di ricerca forniscono anche la possibilità di conoscere quanti link puntano ad una data pagina con un dato indirizzo. Per conoscere il numero di pagine che puntano ad un ipotetico indirizzo www.xxx.yyy, bisogna lanciare il seguente comando link: <http://www.xxx.yyy>.

La *Zona-out* è una zona costituita da siti raggiungibili dalla zona centrale, (e quindi anche dalla *Zona-in*) ma una volta entrati non permettono di essere abbandonati. Molti siti commerciali hanno queste caratteristiche: invogliano i navigatori ad entrare ma non li lasciano più scappare (si ripete nel cyberspazio il mito di Ulisse e della Maga Circe).

I motori di ricerca perlustrano continuamente il www alla ricerca di nuove pagine per catalogare e indicizzare nuove informazioni. Si stima, tuttavia, che anche i motori più affermati abbiano una visibilità del www non superiore al 50%.

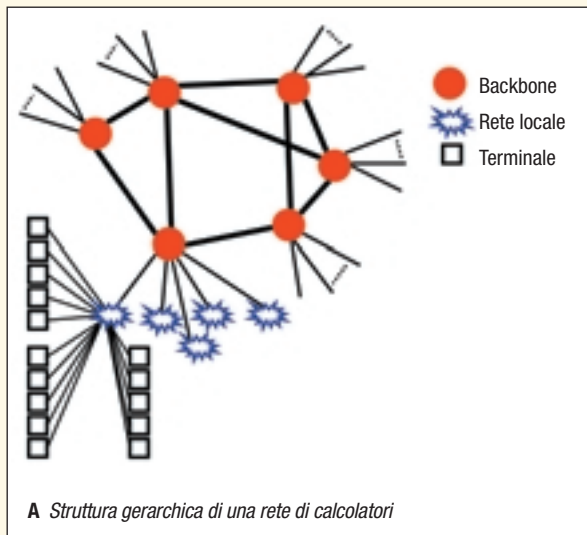
La struttura della rete Internet

La rete Internet è costituita da una collezione di migliaia di reti più piccole, ciascuna sotto il controllo di amministratori locali, e non è possibile avere un quadro esaustivo e completo della sua topologia. Tuttavia, nell'euforia e nell'entusiasmo per la scoperta di fenomeni di diversissima natura che manifestavano un andamento a legge di potenza, nel 1999 venne presentato un articolo nell'importante conferenza SIGCOMM (*Special Interest Group on Data Communication*) dell'ACM (*Association for Computing Machinery*) in cui si affermava di aver sperimentalmente verificato che la rete Internet presentava un andamento ad invarianza di scala con un esponente $\gamma = 2.5$ [1]. Tale risultato fece subito pensare che la topologia di Internet potesse essere spiegata con un modello a crescita del tipo BA, e che di questo modello ereditasse le proprietà.

Successive indagini più accurate [2] hanno messo in luce che la presenza di una legge di potenza nella distribuzione del grado non implica necessariamente la presenza di connettori o *hub* come nelle reti BA, ma sono anche possibili topologie diverse, in cui i nodi a grado più elevato sono localizzati alla periferia della rete. Nel caso della rete Internet i collegamenti fisici fra nodi possono avere ampiezza di banda (capacità di trasmissione) molto disuniforme e quindi non è corretto assegnare a tutti gli archi lo stesso peso (vedi paragrafo 5). È necessario considerare grafi pesati, in cui ad ogni arco viene assegnato un valore proporzionale alla sua capacità trasmissiva. Nei grafi pesati la domanda che è giusto porsi non è quale sia la probabilità che due nodi della rete siano connessi (affidabilità) ma quale sia la probabilità che fra due nodi della rete sia presente una assegnata ampiezza di banda (qualità del servizio).

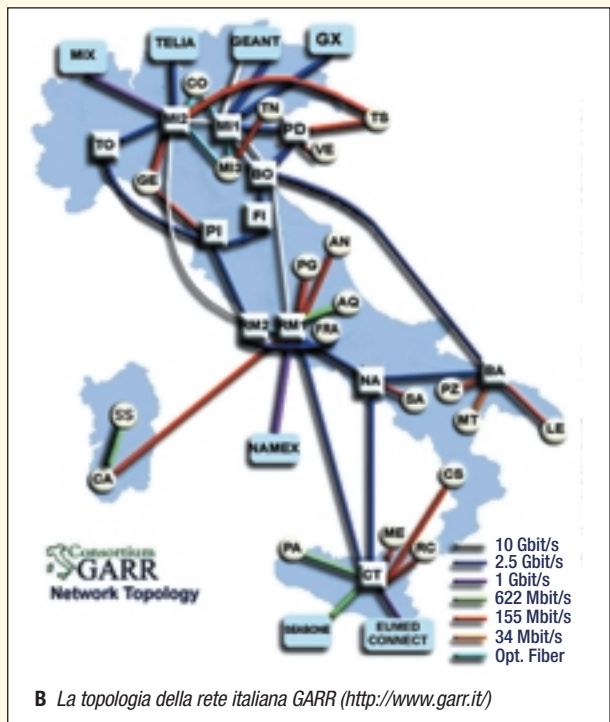
Nella rete internet, inoltre, il grado di connettività di ogni nodo è limitato da considerazioni ingegneristiche e da vincoli tecnologici, dovuti alla capacità massima del nodo stesso di trasmettere flussi di bit. In sostanza un nodo può avere pochi collegamenti a larga banda o un maggior numero di collegamenti a banda stretta, in modo da rispettare il vincolo sul traffico massimo consentito dalla tecnologia del nodo [2]. Queste considerazioni hanno favorito la crescita e lo sviluppo della rete internet con una topologia gerarchica (schematizzata nella Figura A), in cui i nodi in un primo livello (backbone) sono connessi fra loro con linee ad alta capacità; da questi nodi centrali si dipartono reti locali che a loro volta gestiscono i computer terminali. Quindi il grado di connettività dei nodi tende a crescere andando verso la periferia della rete.

Esistono nodi con un numero molto elevato di link, che determinano l'osservata legge di potenza, ma che non hanno la posizione e la funzione



di connettori; questi nodi di alto grado gestiscono reti locali periferiche e il loro non funzionamento ha ripercussioni solo locali. La parte centrale della rete (backbone), quella che sostiene la maggior parte del traffico e che garantisce la qualità del servizio anche fra reti locali dislocate geograficamente in zone molto distanti, ha una struttura più vicina ad una rete poissoniana, e come tale possiede una elevata ridondanza dovuta a possibili percorsi alternativi.

Per esempio, la rete italiana GARR (*Gestione Ampliamento Rete Ricerca*) a cui aderiscono tutti gli enti che rappresentano la comunità accademica della ricerca scientifica in Italia ha la topologia rappresentata nella figura B. Nella figura B sono riportati i nodi da cui si dipartono le reti locali gestite dagli enti aderenti al sistema nella zona geografica di pertinenza, e sono anche evidenziate le larghezze di banda dei canali fisici di connessione, che vanno da 10 Gbit/s (Giga bit per secondo) a 34 Mbit/s (Mega bit per secondo).



RIFERIMENTI BIBLIOGRAFICI

[1] Faloutsos M., Faloutsos P., Faloutsos C.: On the Power-law Relationships of the Internet Topology. (*ACM-SIGCOMM99*) *Computer Communication Review*, Vol. 29, 1999, p. 251-262.
 [2] Lun Li, Alderson D., Willinger W., Doyle J.: A First-Principle Approach to Understanding Internet's Router-level Topology. *ACM-SIGCOMM2004*.

rete in evoluzione, ma vincoli ingegneristici sulla capacità trasmissiva dei collegamenti ne condizionano e guidano la crescita deter-

minando una struttura, pure a legge di potenza, ma che si discosta dal semplice modello BA (vedi riquadro).

Le regole di evoluzione dei modelli auto-organizzativi dipendono da comportamenti locali (regole di aggancio di un nodo di nuova generazione) ma determinano comportamenti globali del sistema. Avendo scoperto che l'emergere in natura di reti complesse con distribuzione del grado a legge di potenza può derivare dal meccanismo di crescita della rete accompagnato da semplici principi auto-organizzativi, viene naturale la domanda, perché la natura ha scelto in molti casi questi processi di evoluzione o comunque processi di evoluzione che portano alle leggi di potenza? La risposta a questa domanda non è univoca e inequivocabile, ma alcuni ricercatori la fanno risalire al modo con cui le reti si comportano in presenza di fenomeni di degradazione, invecchiamento o attacco.

6. TOLLERANZA AI GUASTI E AGLI ATTACCHI

I sistemi fisici possono manifestare nel tempo vari meccanismi di degradazione, usura, malfunzionamento o guasto. Fenomeni cioè che tendono in vario modo a ridurre l'efficienza e la connettività del sistema. Si è già osservato che, in una rete non orientata, un arco per nodo è sufficiente per connettere tutti i nodi fra di loro cioè a permettere di raggiungere qualunque nodo partendo da qualunque altro nodo. Le reti reali sono normalmente molto più connesse di questa soglia minima. Le stime riportate da vari autori non sono molto uniformi: seguendo i quadri riepilogativi riportati in [6] si trovano i seguenti valori per il grado medio $\langle k \rangle$: per la rete internet un grado medio di circa 4, per il www un grado medio attorno a 7, per la rete degli attori cinematografici un grado medio di 61. Questo valore di grado medio maggiore di uno, comporta una intrinseca ridondanza nella rete, cioè il fatto che si possano raggiungere due nodi qualunque attraverso una molteplicità di cammini distinti. Nel grafo non-orientato di figura 1 A, partendo dal nodo 1 si può raggiungere il nodo 6 attraverso numerosi cammini: i più brevi sono, per esempio, il cammino *de*, il cammino *abc*, il cammino *afe*, *djk* ecc¹.. Se per qualche motivo il cam-

mino *abc* fosse interrotto, non si pregiudica la possibilità di un collegamento fra 1 e 6 in quanto sono percorribili strade alternative. Ovviamente in questo ragionamento si suppone che qualunque cammino connesso abbia lo stesso valore. Nella pratica i cammini hanno associato un peso (lunghezza, capacità, costo ecc.) che rende preferibili alcuni percorsi (in generale i più brevi o i più veloci) rispetto ad altri. Dovendo andare in treno da Torino a Reggio Calabria, se tutte le altre linee fossero interrotte, si potrebbe trovare un collegamento passante da Bolzano: il collegamento sarebbe garantito ma gli utenti avrebbero comunque qualcosa da ridire! Nelle reti informatiche la larghezza di banda (o capacità trasmissiva) del collegamento determina la velocità della comunicazione e rende quindi preferibili alcuni cammini (non necessariamente i più corti) rispetto ad altri.

La forza e la robustezza delle reti risiede nella loro intrinseca ridondanza. Per indagare questa problematica più da vicino alcuni studiosi hanno cominciato a porsi il problema di come valutare la ridondanza di una rete complessa, e di come le proprietà di robustezza fossero influenzate dalla struttura topologica della rete. Un malfunzionamento in un arco o in un nodo perturba il funzionamento del sistema localmente, ma può avere scarse ripercussioni sulle proprietà globali del sistema; solo all'aumentare dei nodi o degli archi interrotti cominciano a manifestarsi ripercussioni globali che incidono sui valori dei parametri caratteristici della rete. Per reti di piccole dimensioni è possibile calcolare in modo accurato l'affidabilità della rete, cioè la probabilità che due nodi siano connessi date le probabilità di fallimento di archi e nodi, attraverso algoritmi esaustivi che cercano tutti i possibili cammini fra due nodi [12]. Per reti di grandi o grandissime dimensioni è necessario ricorrere ad argomenti statistici.

Il guasto di un elemento della rete, sia esso nodo o arco, comporta la perdita di operatività dell'elemento stesso, e può essere emulato rimuovendo l'elemento guasto dalla rete. La rimozione di un arco ha influenza solo sul legame fra i due nodi connessi e su tutti i

¹ Si noti, incidentalmente, che la direzionalità degli archi modifica sostanzialmente la connettività della rete. Nel grafo orientato di figura 1 B, l'unico cammino che connette 1 con 6 è il cammino *afe*.

cammini che avrebbero attraversato quell'arco. La rimozione di un nodo implica, invece, la disconnessione di tutti gli archi in ingresso e in uscita dal nodo.

Gli studi in questa direzione hanno cercato di valutare la reazione di una rete alla rimozione progressiva di suoi elementi, cioè la tolleranza della rete ai guasti in funzione della topologia della rete stessa. Prendendo esempio dalle reti informatiche, sono state individuate e studiate due principali categorie di guasti:

□ *guasti casuali*: ogni elemento della rete ha la stessa probabilità di guastarsi indipendentemente dalla sua posizione e dalle sue proprietà (un nodo della rete internet può andare fuori servizio per innumerevoli motivi, un fulmine può cadere su una linea ferroviaria in qualunque punto della rete nazionale, un black-out può causare la sconnessione della rete elettrica). Per emulare questo tipo di guasti, i nodi della rete vengono via via rimossi con probabilità uniforme;

□ *guasti maliziosi o attacchi preferenziali*: sono guasti provocati intenzionalmente allo scopo di indurre il maggior danno sulla connettività della rete. Per i sistemi informatici sono queste le minacce più gravi alla sicurezza del sistema. Emulare gli attacchi di questo tipo, comporta iniziare a rimuovere i nodi partendo da quelli di grado più alto.

Le prime ricerche furono indirizzate a esplo-

rare le proprietà di connettività della rete in funzione della frazione di nodi rimossi, e in funzione della topologia della rete. Indicando con f la frazione di nodi rimossi, in [13] si è indagato sull'esistenza di una frazione critica f_c per cui la rete perde la sua coesione.

6.1. Reti casuali

Nel caso in cui i guasti indotti siano di tipo casuale, all'aumentare della frazione f di nodi rimossi, il cammino caratteristico L cresce lentamente (Figura 5, tratta da [9]). All'aumentare di f si può identificare una frazione critica f_c per cui, quando $f > f_c$, la rete si disgrega in piccole sottoreti fra loro non connesse. Tale effetto si verifica per valori di f_c dell'ordine del 3% [13]. La rete manifesta un comportamento molto simile anche in presenza di attacchi mirati in cui i nodi sono rimossi a partire da quelli con grado più alto. La figura 5 mostra che nel caso di reti casuali (EXP) la crescita di L dovuta a guasti casuali (*failure*) o attacchi (*attack*) è molto simile.

6.2. Reti a invarianza di scala

In questo caso l'effetto di guasti casuali o attacchi mirati è molto diverso. Nel caso di guasti casuali, il cammino caratteristico L cresce molto lentamente (Figura 5): nodi con grado piccolo sono in numero enormemente alto e una loro rimozione ha scarso effetto sulla con-

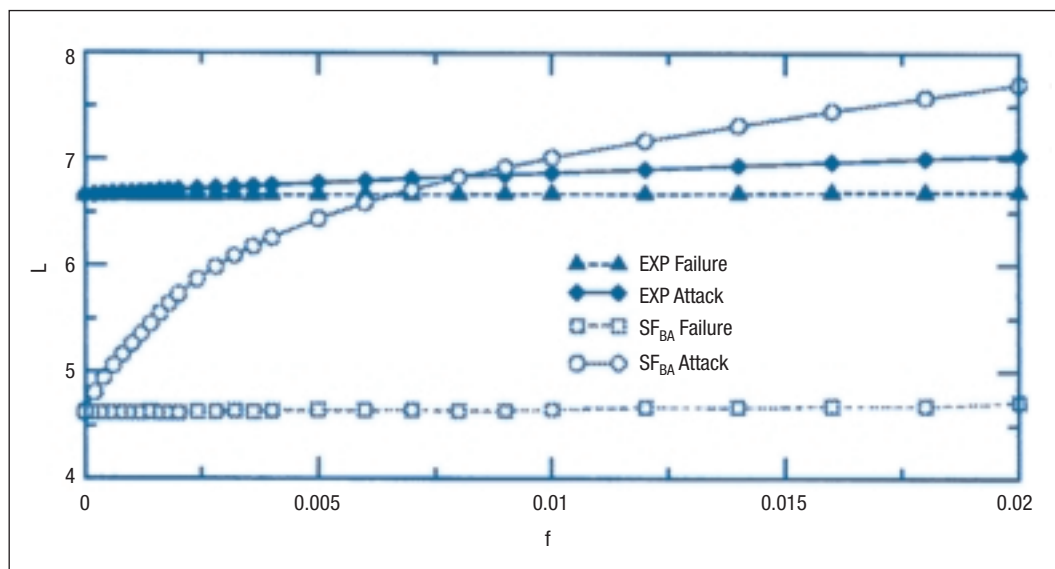


FIGURA 5

Variazione del cammino caratteristico L in funzione della frazione di nodi rimossi f per reti casuali (Exp) e a invarianza di scala con il modello BA (SF) – Figura tratta da [9]

nettività globale della rete. Si è dimostrato che, nel limite di reti che crescono all'infinito, non è possibile identificare una frazione critica f_c , in quanto la rete rimane comunque connessa attraverso gli hub. Ben diverso è il comportamento in presenza di attacchi mirati. Iniziando a colpire gli hub di grado più elevato, la topologia della rete viene drasticamente alterata: il cammino caratteristico L cresce rapidamente (Figura 5) e la rete perde ben presto le sue caratteristiche di connettività.

La conclusione può essere sommariamente formulata con queste parole: le reti casuali hanno un comportamento sostanzialmente simile sia rispetto ai guasti casuali che agli attacchi mirati e si disaggregano al superamento di una soglia critica. Le reti a invarianza di scala sono praticamente immuni rispetto ai guasti casuali ma sono fragilissime rispetto ad attacchi mirati. La fortissima tolleranza rispetto ai guasti casuali delle reti a invarianza di scala viene considerata uno dei fattori principali per giustificare la presenza di questa topologia in reti che evolvono naturalmente.

7. PROPAGAZIONE DELLE EPIDEMIE

La propagazione di malattie o virus (anche informatici) può essere affrontata mediante la teoria delle reti. La rete consiste di nodi (soggetti) sani e nodi (soggetti) infetti. Per modellare la propagazione dell'infezione si può supporre che un soggetto infetto abbia una probabilità v di propagare la malattia ad un soggetto sano a lui direttamente connesso, e che un soggetto sano abbia una probabilità δ di guarire. Indicando con $\lambda = v/\delta$ il tasso di diffusione dell'infezione si sono ottenuti i seguenti risultati principali [5]. In reti casuali l'infezione si propaga a tutti i nodi soltanto se il tasso di diffusione supera una soglia epidemica critica, altrimenti, l'epidemia rimane confinata in zone più o meno ristrette della rete e non riesce ad auto alimentarsi. In reti ad invarianza di scala, quando le dimensioni tendono all'infinito, non esiste soglia epidemica, ma l'epidemia assume le forme di una pandemia e si propaga a tutta la rete, per qualunque valore del tasso di propagazione. Nelle reti finite (anche a invarianza di scala), tuttavia, la soglia epidemica deve esistere altrimenti tutte le reti

biologiche a invarianza di scala sarebbero sparite da tempo, eliminate dalla inarrestabile propagazione di una infezione distruttiva. È vero, comunque, che le reti a invarianza di scala sono più vulnerabili rispetto alla propagazione di virus e malattie rispetto alle reti casuali. Ma esse sono anche più curabili (almeno in teoria). Curando i nodi a partire dagli hub, si può arrestare l'epidemia e ripristinare le condizioni di salute più rapidamente che in una rete casuale. Se invece, in una rete ad invarianza di scala, si applica una strategia di cura che agisce su nodi scelti a caso, la probabilità di arrestare la propagazione dell'infezione diventa molto piccola. Questi studi hanno, come ovvio, un enorme potenziale interesse per arginare e controbattere la distruttività di attacchi maliziosi a sistemi informatici e la propagazione di virus informatici di varia natura, oltre che per debellare epidemie in sistemi biologici (per esempio per sviluppare strategie di confinamento della diffusione dell'AIDS [4]). Per una più approfondita discussione su questi temi si rimanda al lavoro [14].

8. CONCLUSIONI

La teoria delle reti complesse ha avuto negli ultimi anni uno sviluppo travolgente. Ma nonostante centinaia di lavori scientifici siano stati pubblicati sulle riviste di tutto il mondo, essa può ancora considerarsi una scienza nella fase della prima infanzia da cui è lecito aspettarsi altri e più significativi avanzamenti. La speranza è quella di gettare luce su come i sistemi complessi evolvono e si auto-organizzano e su come la topologia della rete (indipendentemente dalle sue caratteristiche sociali, fisiche o tecnologiche) ne determina le proprietà e il comportamento. La comprensione di questi meccanismi guiderà verso la costruzione di reti sempre più efficienti, affidabili e sicure. La sfida alla complessità sarà sicuramente l'obiettivo più impegnativo ma anche il più stimolante per gli scienziati delle prossime generazioni.

Ringraziamenti

Nella compilazione del presente lavoro mi sono avvalso della preziosa collaborazione di Roberta Terruggia.

Bibliografia

- [1] Rita Levi Montalcini: *L'asso nella manica a brandelli*. Baldini & Castoldi 1998, p. 45.
- [2] Watts D.J., Strogatz S.H.: Collective dynamics of small world networks. *Nature*, Vol. 393, 1998, p. 493-592.
- [3] Barabasi A.L., Albert R.: Emergence of scaling in random networks. *Science*, Vol. 286, 1999, p. 509-512.
- [4] Barabasi A.L.: *Linked: the new science of networks*. Perseus, 2002 (traduzione italiana: A. Barabasi, "Link: la scienza delle reti", Einaudi, 2004).
- [5] Dorogovtsev S.N., Mendes J.F.F.: Evolution of networks. *Advances in Physics*, Vol. 51, 2002, p. 1079-1187.
- [6] Albert R., Barabasi A.L.: Statistical mechanics of complex networks. *Review Modern Physics*, Vol. 74, 2002, p. 47-97.
- [7] Newman M.E.: The structure and function of complex networks. *SIAM Review*, Vol. 45, 2003, p. 167-256.
- [8] Boccaletti S., Latora V., Chavez M., Hwang D.: Complex networks: structure and dynamics. *Physics Reports*, Vol. 424, 2006, p 175-308.
- [9] Crucitti P., Latora V., Marchiori M., Rapisarda A.: Efficiency of scale free networks: error and attack tolerance. *Physica A*, Vol. 320, 2003, p. 622-642.
- [10] Milgram S.: The small-world problem. *Physiology Today*, Vol. 2, 1967, p. 60-67.
- [11] Klemm K., Eguíluz V.M.: Highly clustered scale free network. *Physical Review*, Vol. 65, 2002, art. 036123.
- [12] Bobbio A., Ferraris C., Terruggia R.: *New challenges in network reliability analysis*. In: Proceedings of International Workshop on Complex Network and Infrastructure Protection - CNIP06, 2006, p. 554-564.
- [13] Albert R., Jeong H., Barabasi A.L.: Error and attack tolerance of complex networks. *Nature*, Vol. 406, 2002, p. 378-82.
- [14] Lloyd A., May R.M.: How viruses spread among computers and people. *Science*, Vol. 292, 2001, p. 1316.

ANDREA BOBBIO è professore di informatica presso l'Università del Piemonte Orientale A. Avogadro con sede ad Alessandria. Egli si è prevalentemente occupato di metodologie di analisi per sistemi stocastici e di teoria e metodi dell'affidabilità di sistemi. In quest'area di ricerca ha pubblicato numerosi lavori su riviste e conferenze internazionali, ed è stato responsabile di vari progetti di ricerca con enti pubblici e privati. È stato ospite ricercatore in importanti università in Europa, USA e India.

E-mail: bobbio@mfn.unipmn.it

URL: <http://www.mfn.unipmn.it/~bobbio/>