Network complessi e modelli

Rossano Gaeta
Università degli Studi di Torino

Sommario

- Introduzione
- Proprietà network complessi
- Modelli matematici per la rappresentazione
  - Grafi regolari
  - Grafi random
  - Small World
  - Scale Free
- Conclusioni

Domanda

- Cosa hanno in comune?
  - le comunità di individui
  - le infrastrutture di telecomunicazione
  - il World Wide Web
  - le collaborazioni tra attori cinematografici
  - le citazioni negli articoli scientifici
  - le abitudini alimentari di specie animali
  - i network metabolici
  - proteine interagenti

Risposta

- Sono esempi di *sistemi complessi*
  - composti di molti elementi non identici
  - connessi da interazioni di tipo diverso
- Li chiamiamo *network* dove gli elementi sono detti nodi e le interazioni sono connessioni tra i nodi

Esempi

- Comunità di individui
  - nodi: le persone
  - connessioni: relazioni sociali (famiglia, il lavoro, l'amicizia, ecc.)
- Infrastrutture di telecomunicazione
  - nodi: computer, router, satelliti
  - connessioni: linee telefoniche, cavi TV, onde elettromagnetiche, onde radio, raggi luminosi, ecc)

Esempi

- Il World Wide Web
  - nodi: documenti WEB
  - connessioni: i link alle URL di altri documenti
- Internet
  - nodi: computer, router
  - connessioni: cavi ethernet, cavi fibra ottica, onde radio, linee telefoniche, ecc)

Esempi

- Attori cinematografici
  - nodi: attori
  - connessioni: essere stati nel cast di uno stesso film
- Citazioni in articoli scientifici
  - nodi: articoli
  - connessioni: citazioni

Esempi

- Collaborazioni scientifiche
  - nodi: scienziati
  - connessioni: co-autori dello stesso articolo
- Network metabolici
  - nodi: elementi chimici
  - connessioni: reazioni bio-chimiche

Esempi

- Proteine del lievito
  - nodi: proteine
  - connessioni: interazione fisica (binding)
- Sistema di distribuzione energia elettrica
  - nodi: trasformatori, generatori
  - connessioni: linee ad alta tensione

Modelli di network

- Per capire il funzionamento e le proprietà di un network si sviluppano modelli matematici che li rappresentano
- Rappresentiamo i network con uno modello matematico chiamato *GRAFO* composto da
  - un insieme di nodi (detti anche vertici)
  - un insieme di archi (detti anche collegamenti)

Perché si usano i modelli?

- Per lo studio delle proprietà e del comportamento di un sistema
- Perché si vuole conoscere come la *struttura* del network può influire sulla *dinamica* e sul comportamento del sistema ad esempio, che influenza ha:
  - la struttura delle comunità di individui sulla diffusione delle malattie?
  - la struttura della rete di distribuzione di energia elettrica su guasti a catena (blackout)?
  - la struttura delle interazioni tra proteine sulla resistenza alle mutazioni?

Piccola cronologia

- Lo studio del comportamento dinamico di sistemi complessi è stato condotto con l'ipotesi implicita di un sistema con struttura regolare
  - un grafo completamente connesso (ogni nodo è direttamente connesso con un arco ad ogni altro nodo)
  - un grafo detto reticolo
- Ipotesi irrealistiche (pensate agli esempi precedenti) danno origine a modelli iper-semplificati

Piccola cronologia

- Verso la fine degli anni 50 Erdös e Rényl proposero di rappresentare network complessi usando i *grafi random*
- Osservazioni su network complessi reali (fine anni 90, grazie alla disponibilità di dati e potenza di calcolo):
  - small-world
  - scale-free
- Definizione di modelli matematici (grafi) che siano in grado di catturare queste due proprietà dei network complessi reali

Sommario

- Introduzione
- Proprietà network complessi
- Modelli matematici per la rappresentazione
  - Grafi regolari
  - Grafi random
  - Small World
  - Scale Free
- Conclusioni

Concetti base

- Per capire meglio i diversi modelli proposti per catturare le proprietà di network complessi dobbiamo introdurre i concetti di:
  - lunghezza media di un cammino
  - coefficiente di clustering
  - grado (connettività) di un nodo

Lunghezza media di un cammino

- In un network la distanza tra due nodi i e j di definisce come il numero di archi del cammino più breve che li connette
- Il diametro è definito come il massimo di tutte le distanze tra ogni coppia di nodi del network
- La lunghezza media di un cammino (L) di un network è la media delle distanze tra coppie di nodi ottenuta come media aritmetica su tutte le coppie di nodi

Lunghezza media di un cammino

- L dice qualcosa sulla distanza tipica tra due nodi del network
- Si è osservato che nella maggior parte dei network reali la lunghezza media di un cammino non è molto grande: cresce al più con il logaritmo del numero di nodi
- Questa osservazione è alla base del fenomeno detto *small-world*

Coefficiente di clustering

- Consideriamo il network degli individui e delle relazioni sociali (amicizia) tra loro
- Diventa possibile (e probabile) che l'amico di un vostro amico sia anche vostro amico
- Questa proprietà si riferisce al clustering del network

Coefficiente di clustering

- Supponiamo che il nodo j del network abbia k_j archi che lo connettono ad altrettanti nodi
- Questi nodi sono vicini del nodo j
- Al più $k_j (k_j - 1) / 2$ archi possono esistere tra loro (quando ogni vicino del nodo j è connesso ad ogni altro vicino del nodo j)
- Ne esistono, in generale, di meno (E_j)

Coefficiente di clustering

- Il coefficiente di clustering del nodo j (C_j) si definisce come il rapporto

$$C_j = 2\, E_j \,/\, (k_j\, (k_j - 1))$$

- Il coefficiente di clustering del network (C) si definisce come la media aritmetica dei coefficienti di ogni nodo
- Per definizione $0 \leq C \leq 1$ e C=1 se e solo se il network è completamente connesso (ogni nodo è direttamente connesso con un arco ad ogni altro nodo)

Grado di un nodo

- La caratteristica più semplice e forse più importante di un nodo è il suo *grado* (o connettività)
- Il grado di un nodo j (k_j) è definito come come il numero totale di connessioni con altri nodi
- Intuitivamente, fornisce un'idea dell'importanza di un nodo

Valori per network reali

- Valori calcolati su network reali dimostrano che spesso

Network	C	L	N
WWW	0.1078	3.1	153127
Internet	0.18-0.3	3.7-3.76	3015-6209
Attori	0.79	3.65	225226
Co-autori	0.43	5.9	52909
Metabolico	0.32	2.9	282
Catena del cibo	0.22	2.43	134
C. elegance	0.28	2.65	282

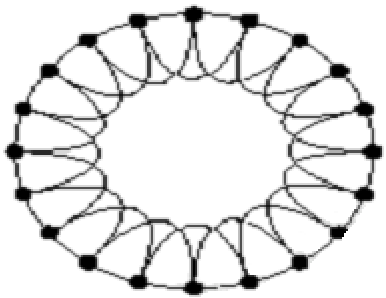
- C non è piccolo ed L non è grande

Sommario

- Introduzione
- Proprietà network complessi
- Modelli matematici per la rappresentazione
  - Grafi regolari
  - Grafi random
  - Small World
  - Scale Free
- Conclusioni

Grafi regolari

- Grafo completamente connesso con N nodi e N (N – 1) / 2 archi
- Reticolo in una dimensione (lattice): ogni nodo è connesso con nodi vicini a distanza 1,2, …,K/2 (K pari). Ogni nodo ha K connessioni.
  - Un anello di N nodi e K=4 dove ogni nodo è connesso ai due immediati vicini e ai due vicini a distanza uno



Grafi regolari

- Un grafo completamente connesso rappresenta network con C alto (C=1) e L piccolo (L=1) ma i network reali non sono completamente connessi
- I reticoli rappresentano network con C alti ma $L \cong N/2K$, ovvero, $L \to \infty$ per $N \to \infty$

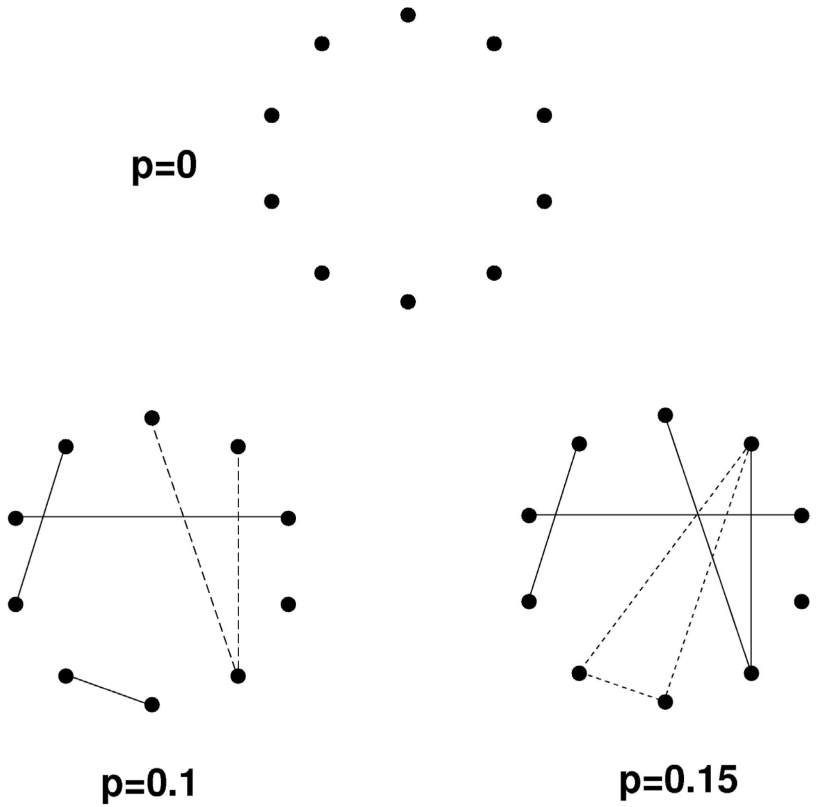
Grafi random

- Sono stati il primo modello matematico diverso da un grafo regolare per la rappresentazione di network complessi
- Si possono vedere come esattamente l'opposto dei grafi regolari
- Si considerano N nodi ed una probabilità p
- Si considera ogni coppia di nodi e per ognuna si definisce un arco con probabilità p
- Il numero totale di archi è circa $pN(N-1)/2$

Grafi random

- Si analizzano in funzione di p
- Si definisce una particolare proprietà del grafo (ad esempio, se il grafo è connesso) e si vuole sapere per quale valore di p la proprietà sarà verificata con alta probabilità

Grafi random



$$G = G_{N,p}$$

Grafi random

- Il grado medio <k> di un grafo random è p(N-1)
- L è circa ln N / <k> per cui cresce lentamente al crescere di N (small world)
- $C=p=\langle k\rangle/N$ ovvero, $C \to 0$ per $N \to \infty$ per cui non rappresenta il fatto che nella maggior parte dei network complessi il coefficiente di clustering non tende a 0

Grafi small-world

- I reticoli in una dimensione esibiscono valori alti di C ma L diverge
- I grafi random esibiscono valori di L che crescono lentamente all'aumentare di N ma C tende a 0 all'aumentare di N
- Scegliamo una via di mezzo!

Grafi small-world – a)

1. Si inizia con un reticolo regolare in una dimensione di N nodi con un certo valore di K
2. Per ogni arco, cambia uno dei due estremi con probabilità p scegliendo a caso il nuovo nodo con il vincolo che:
   - una coppia di nodi non può avere più di un arco che li connette
   - un nodo non può avere un arco che lo connette a se stesso

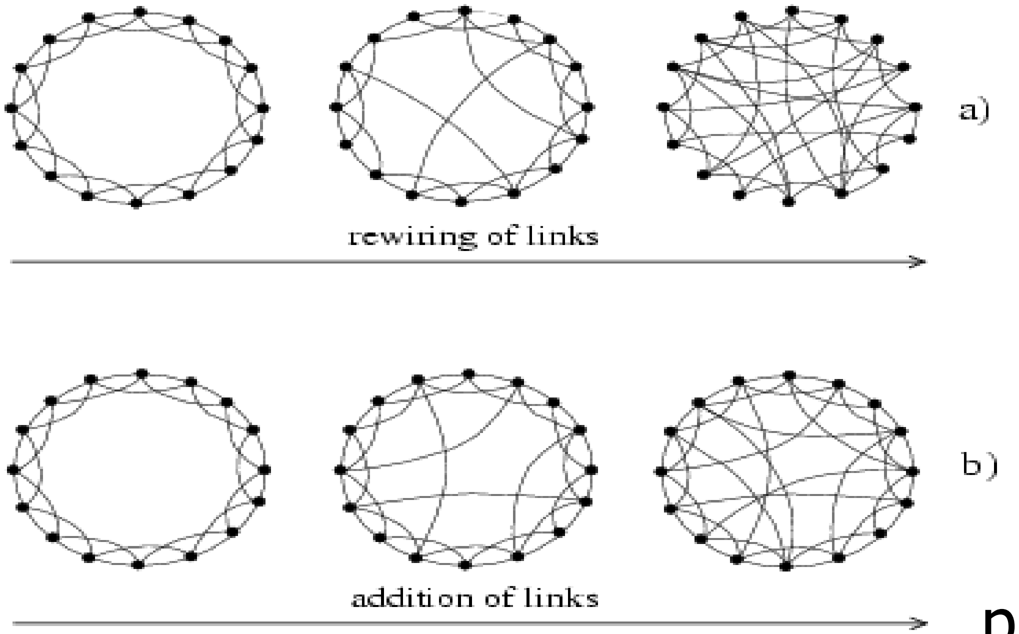
Grafi small-world – b)

1. Si inizia con un reticolo regolare di N nodi in una dimensione con un certo valore di K
2. Per ogni coppia di nodi aggiungi un arco con probabilità p con il vincolo che:
   - una coppia di nodi non può avere più di un arco che li connette
   - un nodo non può avere un arco che lo connette a se stesso

Grafi small-world

- Modulando il valore di p si introduce un certo numero di archi tra nodi distanti

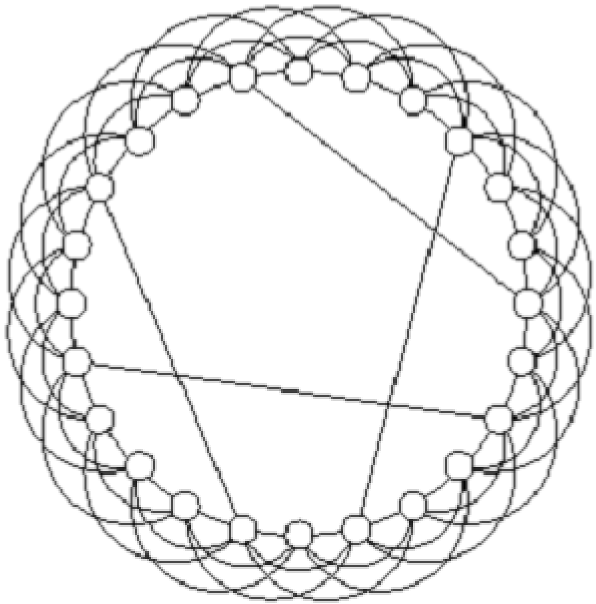
Regolare   Small-World   Random



p=0   p=1

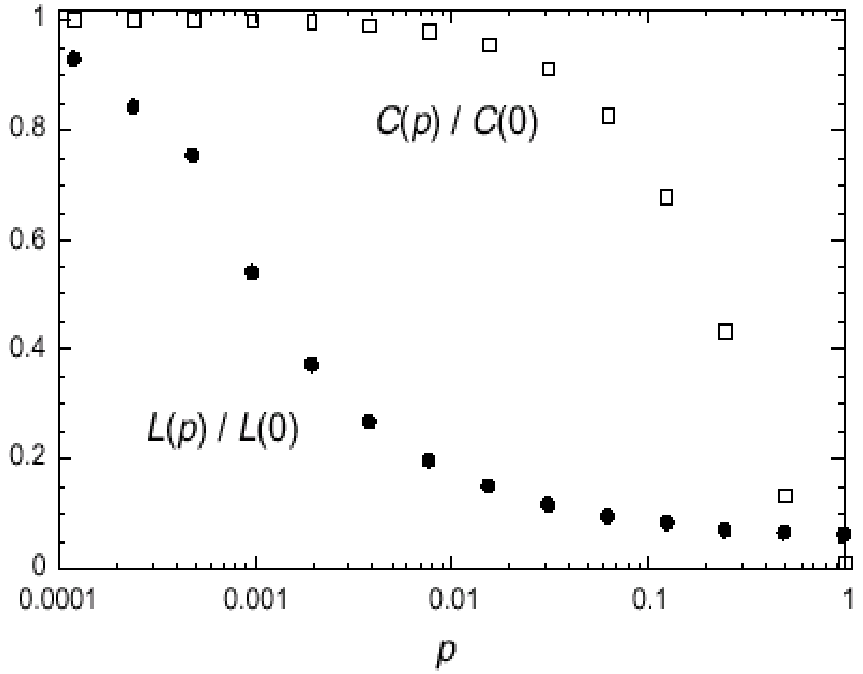
Grafi small-world

- grafo small-world con N=24 e K=6



Grafi small-world

- Sia C sia L sono funzione della probabilità p e per p=0 abbiamo i valori di un lattice



Grafi small-world

- Abbiamo trovato il modello giusto per rappresentare network complessi?
- Esiste un’altra caratteristica dei network complessi che non viene rappresentata adeguatamente dai grafi small-world
- La distribuzione del grado di un nodo

Concetti base

- Per capire meglio i diversi modelli proposti per catturare le proprietà di network complessi dobbiamo introdurre i concetti di:
  - lunghezza media di un cammino
  - coefficiente di clustering
  - distribuzione del grado

Distribuzione del grado di un nodo

- La caratteristica più semplice e forse più importante di un nodo è il suo *grado*
- Il grado di un nodo j (k_j) è definito come come il numero totale di connessioni con altri nodi
- Intuitivamente, fornisce un'idea dell'importanza di un nodo

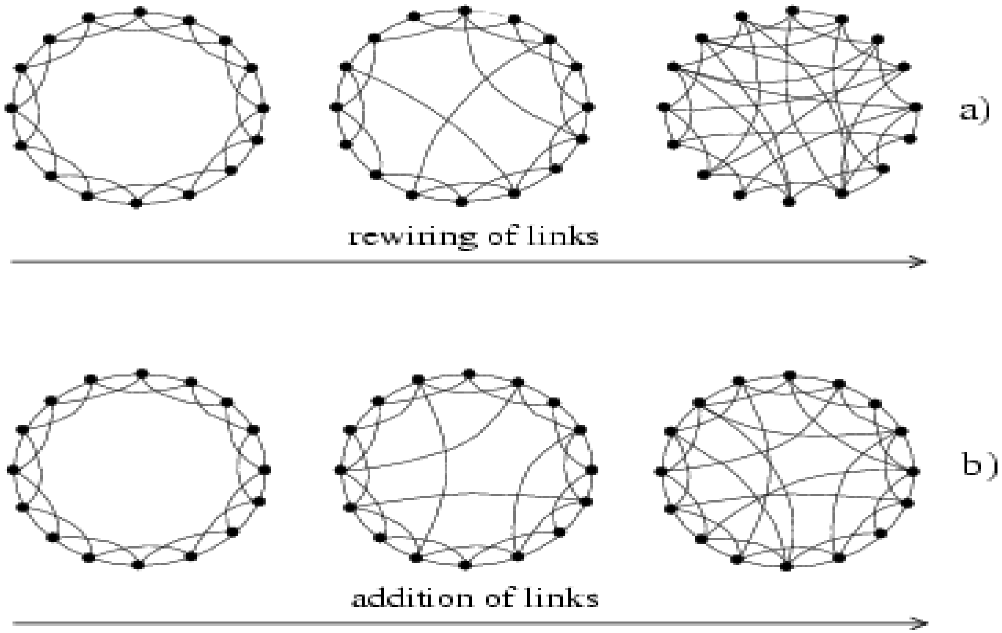
Distribuzione del grado di un nodo

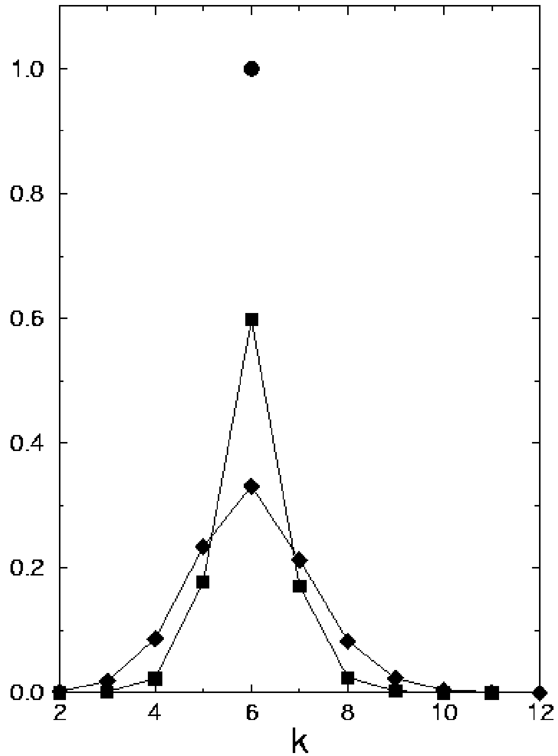
- Il valore della media aritmetica del grado di ogni nodo si definisce come grado medio del network <k>
- In realtà, il valore medio fornisce poche informazioni
- Si definisce la distribuzione del grado di un nodo P(k)
- P(k) è una funzione che fornisce la probabilità che un nodo estratto a caso abbia esattamente k connessioni

Grafi small-world

- Sia i grafi random sia i grafi small-world danno origine a modelli che hanno una distribuzione del grado di un nodo di tipo esponenziale (per N grande, Poisson)
- Questo vuol dire che c'è omogeneità nel numero di archi per ogni nodo
  - la probabilità che un nodo abbia un grado molto più alto o molto più basso del valore medio decresce in maniera esponenziale
- Per i grafi random si ha: $P(k) \approx e^{-pN} \frac{(pN)^k}{k!} = e^{-<k>} \frac{<k>^k}{k!}$

Grafi small-world





$$P(k) \approx e^{-pN} \frac{(pN)^k}{k!} = e^{-<k>} \frac{<k>^k}{k!}$$

Scale-free

- Si è osservato invece che spesso in network reali la distribuzione del grado di un nodo non è Poisson (o esponenziale) ma ha una forma del tipo $P(k) \sim k^{-\gamma}$

- cioè la distribuzione segue una legge di potenza (power law) con un esponente maggiore di 1
- network di questo tipo di dicono <u>*scale-free*</u>

Cosa vuol dire?

- La forma della distribuzione P(k) implica che esistono alcuni nodi con grado altissimo (i cosiddetti hub)
- Infatti la probabilità che un nodo abbia un numero molto alto di archi decresce solo in maniera polinomiale e non esponenziale
- Giocano un ruolo fondamentale nel tenere i nodi del grafo vicini tra loro

Scale-free

- Ad esempio:

WWW (in)	Internet	Actor	Citation index	Sex Web	Cellular network	Phone call network	linguistics
$\gamma = 2.1$	$\gamma = 2.\ 5$	$\gamma = 2.3$	$\gamma = 3$	$\gamma = 3.5$	$\gamma = 2.1$	$\gamma = 2.1$	$\gamma = 2.8$

Scale free

- i modelli presentati in precedenza non catturano due proprietà importanti dei network reali:
  - descrivono network statici ovvero, una volta fissato N il numero di nodi non varia e, al più, si possono aggiungere o modificare archi
  - assumono probabilità uniformi di creazione o modifica di archi mentre queste probabilità possono variare da nodo a nodo e nel tempo

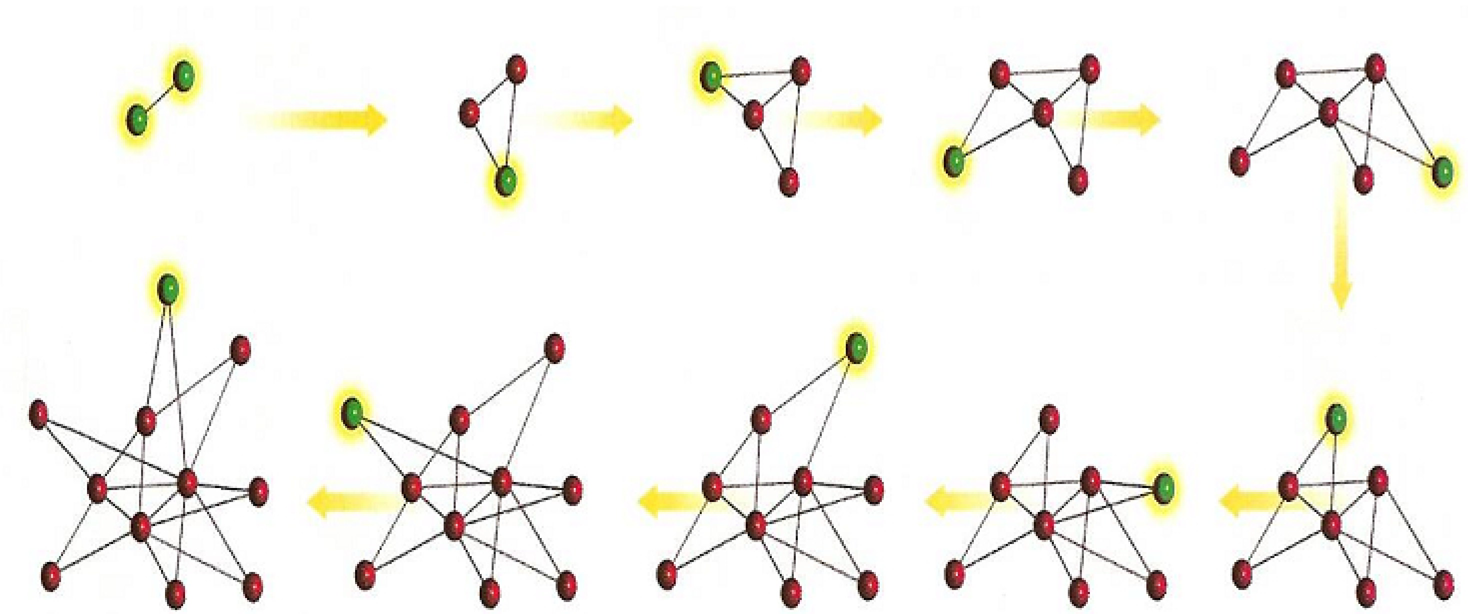
Scale-free

- Barábasi ed Albert hanno proposto una possibile spiegazione del fenomeno che origina una distribuzione del grado di tipo power law
  - crescita del network: buona parte dei network reali evolve mediante l'aggiunta di nuovi nodi
  - connessioni preferenziali: un nuovo nodo tende a connettersi a nodi che hanno già un alto numero di connessioni
- i due fattori devono essere contemporaneamente presenti, non basta solo uno dei due per avere network scale-free

Scale free

- Hanno proposto un algoritmo per la generazione di modelli scale-free
  1. Crescita: si inizia con un numero (piccolo) m_0 di nodi
  2. Ad ogni passo si aggiunge un nuovo nodo e lo si connette ad altri $m \leq m_0$ nodi già esistenti
  3. La probabilità che un nuovo nodo venga connesso al nodo j (uno degli m già esistenti) dipende dal suo grado k_j nel seguente modo: $\Pi_j = k_j / \sum_j k_j$

Scale free



Scale-free

- Dopo t passi l’algoritmo genera un modello con $N=t+m_0$ nodi e mt archi
- Si dimostra che il modello ha una distribuzione del grado di un nodo del tipo power law con esponente 3 cioè la probabilità che di trovare un nodo con k archi è proporzionale a k^{-3}
- Non dipende da m, l’unico parametro del modello

La panacea o solo l'inizio

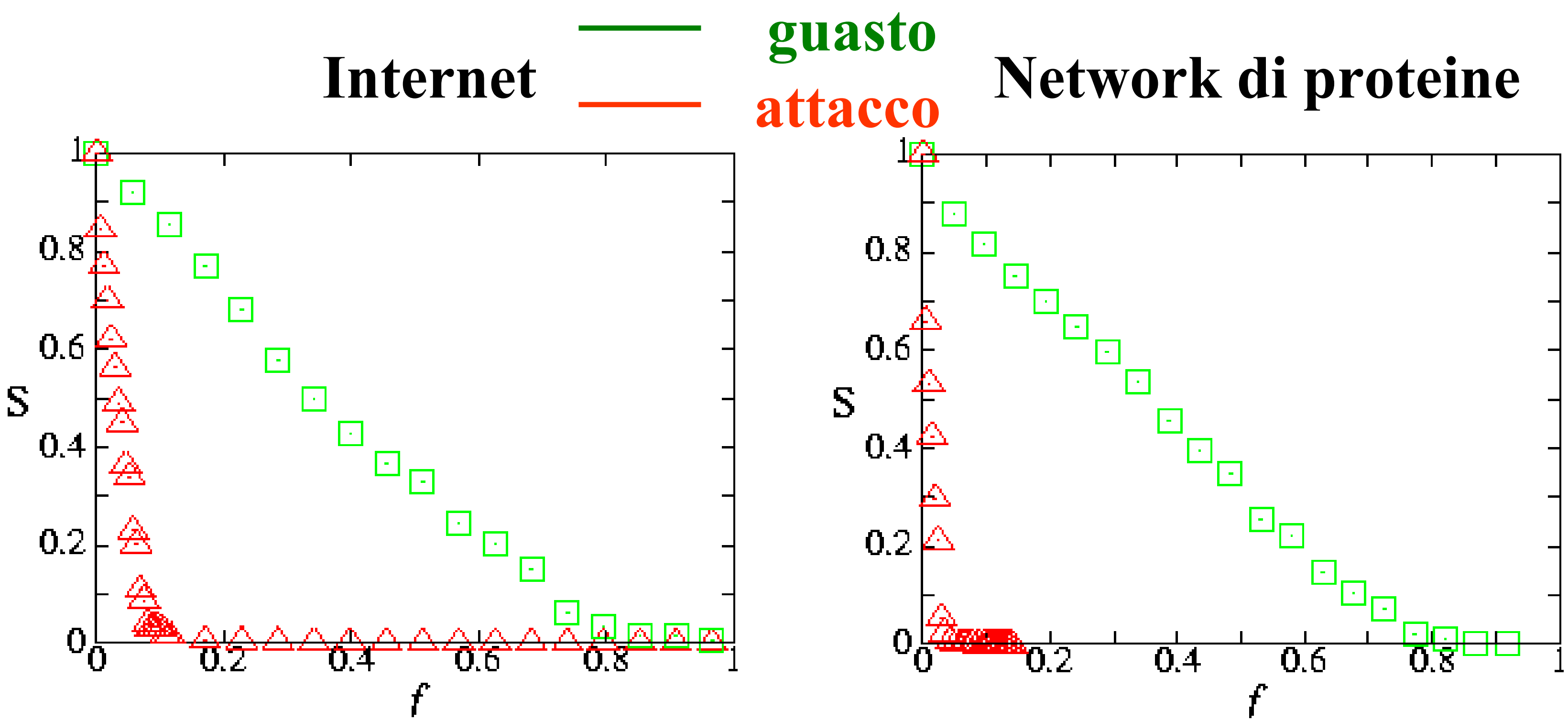
- Non ci sono ancora espressioni analitiche per C ed L ma simulazioni rivelano che:
  - L cresce con il logaritmo di N (small world)
  - C decresce come $C \sim N^{0.75}$
- Il modello presentato cattura solo alcuni aspetti dell'evoluzione dei network reali

Caratteristiche dei network scale-free

- Caratteristiche dei network studiate grazie ai modelli
  - estrema tolleranza a guasti o danni casuali (può spiegare perché tali network sono così diffusi in natura)
  - estrema vulnerabilità a guasti degli hub o a danni mirati (attacchi). Gli hub devono essere attentamente protetti (o attaccati con precisione)

Caratteristiche dei network scale-free

- Andamento della percentuale di nodi connessi in funzione della frazione di nodi guasti (o attaccati)



Caratteristiche dei network scale-free

- Su network scale-free malattie o virus informatici si diffondono in maniera velocissima
- Questo suggerisce di immunizzare gli hub come strategia di prevenzione di malattie infettive (o virus informatici)

Sommario

- Introduzione
- Proprietà network complessi
- Modelli matematici per la rappresentazione
  - Grafi regolari
  - Grafi random
  - Small World
  - Scale Free
- Conclusioni

Conclusioni

- Modelli matematici per rappresentare network reali (di provenienza eterogenea)
- Utili per lo studio di proprietà dei network
- Per saperne di più

  - **[1] "Linked: The new science of Networks"**
    A.L. Barabasi , Perseus Publishing
  - **[2] "Scale free networks"**
    A.L. Barabasi, E. Bonabeau
    Scientific American, May 2003, pp.60-69
  - **[3] "Statistical Mechanics of Complex Networks"**
    R.Albert and A.L. Barabasi
    Review of Modern Physics, Vol.74, January 2002