

Sommario

Lo scopo di questo lavoro è fare un'analisi di rete usando un campione di antenne telefoniche cellulari, ipotizzando di voler costruire una ipotetica *mesh network* con esse. Abbiamo scelto di analizzare il sistema delle antenne comprese entro il Raccordo Anulare di Roma. Non sapendo a priori di quale tipo di rete si tratta, e che tipo di grafo si costruisce da essa, sono state studiate alcune proprietà topologiche e comportamentali dell'insieme di antenne studiate. Dopo aver definito il criterio con cui dare un link a due nodi, sono state calcolate le matrici di adiacenza della rete complessiva e di quelle composte solo dalle antenne dei singoli gestori presenti in Italia. La prima parte dell'analisi consiste nell'estrapolare dai grafi ottenuti le distribuzioni dei gradi e nell'affidare all'istogramma dei gradi la forma funzionale che meglio gli si adatti. Avendo così ottenuto indicazioni significative sui tipi di rete in analisi, le abbiamo verificate studiando il comportamento percolativo della rete, mediante rimozione di nodi in due diversi scenari: un attacco intenzionale, che rimuovesse le antenne a partire da quelle con maggior grado, e una caduta random del sistema. I risultati ottenuti sono stati confrontati con modelli di rete esponenziali e scale-free. L'esposizione è articolata nel seguente modo:

Nella sezione 1 verranno esposte le basi teoriche dello studio effettuato, con particolare accento ai modelli di rete utilizzati come riferimento e ai concetti di percolazione e soglia percolativa.

Nella sezione 2 si spiega come sono stati raccolti e organizzati i dati, come è stata definita la rete e quindi con quali criteri è stata definita la matrice di adiacenza. Successivamente, accennando ai problemi computazionali avuti nel gestire la rete complessiva da circa 7000 nodi, verranno calcolate le matrici di adiacenza e le distribuzioni del grado.

Nella sezione 3 viene effettuato lo studio percolativo mediante i due scenari di rimozione di nodi, analizzando l'andamento del diametro e delle dimensioni del cluster più grande con la progressiva caduta delle antenne. I risultati sono stati confrontati con quelli ottenuti dai modelli, i quali hanno rivelato delle ambiguità. Nel caso di attacco random è stato testato infine un ipotetico effetto cascata causato dall'overload delle antenne rimaste (previa rimozione dei grossi hub che servono proprio a scongiurare tale evenienza).

Indice

1 Reti <i>scale-free</i>	4
1.1 Proprietà e grandezze caratteristiche	4
1.2 Modelli di rete esponenziale	5
1.3 Reti scale-free	8
1.4 Percolazione	11
2 Costruzione e analisi delle reti	13
3 Network breakdown	13
3.1 Simulazione	13
4 Conclusioni	13
Riferimenti bibliografici	14

Elenco delle figure

1	Reticolo ciclico.	7
2	Generazione con modello Watts-Strogatz.	8
3	Confronto grafi random.	9
4	Confronto clustering.	10
5	Confronto $P(k)$	11

1 Reti *scale-free*

Le reti sono insiemi di oggetti per i quali è possibile una connessione. Gli ultimi 60 anni hanno visto un crescente interesse per esse, perché con le reti è possibile descrivere sistemi dei più disparati tipi nei campi più vari, dalla biologia all'economia, passando per reti elettriche, informatiche, ecosistemi, e altro ancora. Ancora più importante, negli ultimi vent'anni è stato scoperto che una classe di reti ha proprietà del tutto comuni nonostante l'intrinseca diversità tra sistemi; la modellizzazione di queste reti, come crescita con caratteristiche preferenziali, è dovuta a **Barabási 1999** e vengono dette scale-free.

Per maggior semplicità, nell'esposizione del lavoro svolto verrà sempre assunto che le reti siano dirette e non pesate.

1.1 Proprietà e grandezze caratteristiche

Dal punto di vista matematico una rete è rappresentata da un grafo. Gli oggetti costitutivi di un grafo sono i **nodi**, i quali vengono collegati opportunamente tra loro con dei link secondo dei criteri che dipendono dal modello (nella costruzione di un grafo teorico) o dalla natura del sistema (per le reti reali). Il numero di nodi della rete, o di una sua parte, è ovviamente la grandezza fondamentale per definirne le dimensioni; il numero e la distribuzione dei link ne descrivono la connettività.

Prima di procedere alla descrizione dei modelli di rete sopraccitati, è utile elencare un glossario di caratteristiche delle reti, spesso determinanti per distinguere un grafo scale-free dagli altri nello studio delle reti reali.

Grado Il grado di un nodo è il numero di altri nodi a cui è connesso. La connettività di una rete è ben rappresentata dalla **distribuzione del grado**, la cui forma funzionale $P(k)$ la cui forma funzionale è il primo criterio per discriminare una rete scale-free.

Distanze alla base di ogni topologia c'è il concetto di distanza. La distanza tra due nodi di un grafo è definita come il numero di link che li separa nel più piccolo cammino possibile lungo la rete. Altre quantità topologiche importanti sono:

- l'**eccentricità di un nodo**, è la massima delle distanze tra un nodo scelto e tutti gli altri nodi della rete;
- il **diametro** è la massima eccentricità tra quelle di tutti i nodi del grafo; detto in termini più generali, il diametro è il minor cammino più grande tra tutte le possibili coppie di nodi della rete;
- **average path length**, cioè la distanza media tra tutte le possibili coppie di nodi della rete.

Custering Per misurare quanto i nodi di un grafo tendono a creare dei clusters, viene definito il coefficiente di clustering C_i . Dal punto di vista di un vertice i di grado k_i , quindi, considerando i nodi a esso collegato, C_i è definito come il rapporto

$$C_i = \frac{2E_i}{k_i(k_i - 1)},$$

dove E_i è il numero di link tra i k_i primi vicini del vertice e $k_i(k_i - 1)/2$ il numero di link possibili tra essi: cioè il numero di collegamenti necessari perché i con i suoi vicini sia una porzione di grafo *completa*, detta anche *clique*. La media su tutti i nodi dei rispettivi C_i dà un'indicazione su quanto la rete sia complessivamente clusterizzata e quindi può essere preso come coefficiente di clustering globale C della rete. Una definizione più recente di C , equivalente alla precedente, lo pone uguale al rapporto tra il numero di triplette di nodi completamente collegate N_Δ e il numero di triplette che vede i tre nodi collegati da almeno due link N_\wedge . In entrambi i casi, più C è vicino a 1, più si ha clusterizzazione.

Il coefficiente di clustering svolge un ruolo importante nel distinguere una rete completamente random da una che presenta caratteristiche di *small-world*: infatti un rete con piccolo diametro tende a avere un $\langle C \rangle$ maggiore di una simile rete puramente casuale costruita con lo stesso numero di nodi e con il medesimo cammino più corto medio. Questo comportamento è stato notato in molte reti reali (**Watts1998**).

Centralità Esistono vari modi per definire quando un nodo è centrale rispetto ad altri.

Il primo è più immediato è il suo grado. Altri due tra i più importanti sono la **betweenness** e il **page-rank**. Preso un nodo i , il primo è definito come il numero di cammini più corti che è possibile tracciare tra due qualsiasi nodi della rete, purché passino per i ; il secondo dà una maggior centralità a i maggiore è il numero di link *diretti* verso di esso, tenendo conto anche del rank dei nodi che si collegano a esso.

In reti dirette, benché non siano *matematicamente* uguali, betweenness e page-rank sono statisticamente molto correlati al grado, pertanto non verranno analizzati.

1.2 Modelli di rete esponenziale

Lo studio delle reti random nasce dagli ungheresi **Erdos1959** i quali per primi modellizzarono la definizione di una rete usando criteri probabilistici. Nel corso dei 40 anni successivi è stato osservato che molte reti reali, a dispetto delle dimensioni, avevano un diametro molto piccolo, portando alla definizione del concetto di *small-worlds* e al modello di **Watts1998**

In entrambi i casi si parte da una configurazione iniziale di nodi per poi randomizzare i link tra essi. Per questo motivo la distribuzione del grado segue una Poissoniana, con una coda destra caratteristica di forma esponenziale; con un numero sufficiente di nodi e link, la distribuzione del grado si approssima a una gaussiana con un grado medio ben definito.

1.2.1 1.2.1 Random network

Il modello di Erdős e Rényi parte da un certo numero N di nodi e n di link. Se poniamo che i nodi siano distinguibili, possono esistere (n sopra $N(N-1)/2$) configurazioni equiprobabili tra le quali può esserne presa una in maniera random. Una maniera più articolata per definire una rete random è il criterio binomiale: partire da un set di N nodi

e dare a ogni possibile coppia un link con una certa probabilità p . Il punto più importante di un approccio probabilistico alle reti è lo studio di proprietà dei grafi: ponendo $N \rightarrow \infty$, se la probabilità che una certa proprietà si verifichi tende a 1, si osserva che per reti limitate quella proprietà si verifica con probabilità significativa anche con pochi nodi. Ciò si verifica per molte proprietà, e questo fatto permette di poter trattare le reti per categorie secondo le loro proprietà peculiari.

Nel caso di un grafo di Erdős e Renyi queste sono:

- la distribuzione del grado ha una forma di distribuzione binomiale, la quale tende a una poissoniana per p piccole, e a una gaussiana per $\langle k \rangle$ grandi. Questo implica che la topologia della rete è abbastanza omogenea, con molti nodi che hanno approssimativamente lo stesso grado;
- con N grande, il diametro tende a essere piccolo, come l'average path length. Con p non troppo piccolo il numero di nodi che abbiano una certa distanza una lunghezza l si può approssimare a $\langle k \rangle^l$; uguagliandolo a N deriva che diametro e average path length scalano con buona approssimazione con il logaritmo di N , secondo la relazione

$$l \sim \frac{\ln(N)}{\ln(\langle k \rangle)}. \quad (1)$$

Molte reti reali presentano simili caratteristiche nei gradi di separazione, che hanno portato alla definizione del termine "small world" per esse;

- il clustering di una rete random tende a essere molto basso. Infatti, preso un nodo e i suoi primi vicini, la probabilità una coppia di essi sia connessa è p . Pertanto su tutto il grafo, il coefficiente di clustering medio è proprio p , il quale di solito è abbastanza minore di 1 (con $p = 0.1$ un grafo random diventa già molto connesso, per esempio con 10^4 nodi avrebbe $\langle k \rangle = 10^3$). Questo fatto, al contrario del diametro, si pone in contrasto con le reti reali, le quali hanno quasi sempre un $\langle C \rangle$ sensibilmente più alto.

1.2.2 Small World

Come abbiamo visto il modello di Erdős e Renyi descrive bene il piccolo diametro delle reti reali, ma non il loro elevato grado di clusterizzazione. Inoltre il coefficiente di clustering di esse è simile per reti con numero di nodi molto diverso. Notando per primi ciò, Watts e Strogatz hanno formulato un modello che meglio si adattasse alle caratteristiche reali delle reti.

Il fatto che il coefficiente di clustering non dipende dal numero di nodi è caratteristico dei reticolati, pertanto il punto di partenza del modello di Watts e Strogatz è un reticolo con condizioni al contorno cicliche, i cui N nodi sono collegati ai primi n vicini. Se poniamo, per esempio, $n = 2$, si configura così un anello del tipo:

Successivamente si procede a riarrangiare in maniera random i link tra i nodi, con una probabilità p per ogni link di venire modificato. In questo modo si hanno un certo numero di link (in media pNn) che invece di essere tra nodi in prossimità, saranno tra nodi più lontani, come nel grafo:

Con questo metodo, a seguito del *rewiring* si ha il rischio che il grafo non sia più connesso. Ponendo $n > 1$ ciò può essere evitato, portando la probabilità di avere un

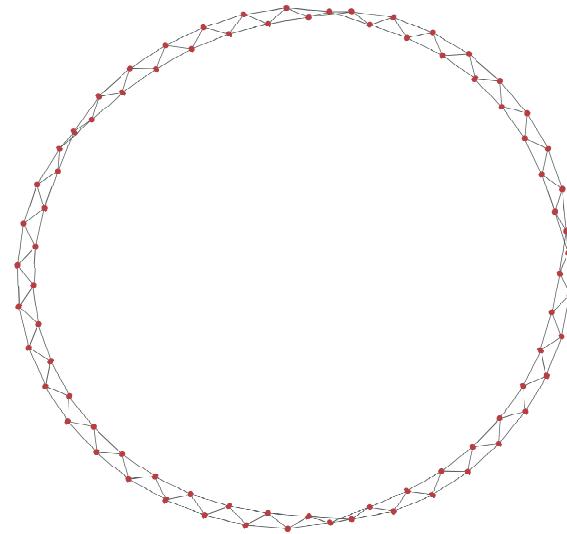


Figura 1: Il punto di partenza del modello Watts-Strogatz: un reticolo con condizioni al contorno cicliche.

grafo non connesso quasi a zero già con $n = 2$. Con $p \rightarrow 1$ il grafo diventa simile a quello di una rete di Erdős-Renyi:

Per avere una rete tipica che abbia un numero di connessioni non troppo elevato, ma non così poco da rischiare da avere un grafo non connesso a seguito dell'operazione di rewiring, possono essere considerati degli N e n tali che $N \gg n \gg \ln(N) \gg 1$. Con queste condizioni la distribuzione del grado $P(k)$ ha una forma gaussiana la cui media coincide con n , con σ più piccole per p basse, tendente a una delta di Dirac per $p \rightarrow 0$. Infatti, mentre con $p = 0$ si ha un reticolo con grado uguale per ogni nodo, l'operazione random di rewiring introduce una casualità sulla $P(K)$ ben descritta da una gaussiana per N grandi, che però ha una larghezza sensibilmente inferiore a quella di una rete random, portando a una rete ancora più omogenea.

Valori tipici di $\langle l \rangle$ per le reti reali sono ben descritti dal modello di Erdős e Renyi secondo l'equazione 1 (**Barabási 2002**), cioè ci si aspetta scali in modo logaritmico fissato $\langle k \rangle$. In una rete generata secondo il modello di Watts-Strogatz le lunghezze dei cammini, fissato $n \equiv \langle k \rangle$, scalano in media in modo diverso al variare di p . Per p molto bassa ($p \ll 1/nN$) le lunghezze caratteristiche sono proporzionali alla dimensione del grafo, il quale è ancora molto simile a una rete regolare; abbastanza presto, per $p \gg 1/nN$,¹ vi è invece un largo intervallo di p che vede già verificarsi il fenomeno small-world, con le lunghezze dei cammini che scalano come $\ln(N)$ in accordo con le reti reali.

L'altra caratteristica fondamentale di uno small-world è che abbia un clustering abbastanza alto in relazione a una rete puramente random. Per $p = 0$ il reticolo ha un $C(0)$ costante al variare di N ; all'aumentare di p , presa una tripletta chiusa, la probabilità che tutti e tre i link rimangano inalterati è $(1 - p)^3$ e i suoi due primi vicini hanno probabilità p^3 di vedere almeno uno dei loro link riarrangiato. Dato che $C = N_\Delta/N_\wedge$ e che N_\wedge rimane costante con il rewiring, si può porre

$$C(p) \propto N_\Delta(p) \Rightarrow C(p) \sim C(0)(1 - p)^3,$$

¹ $p \gg 10^{-4}$ per una rete con 10^3 nodi e $P(k)$ centrata in 10

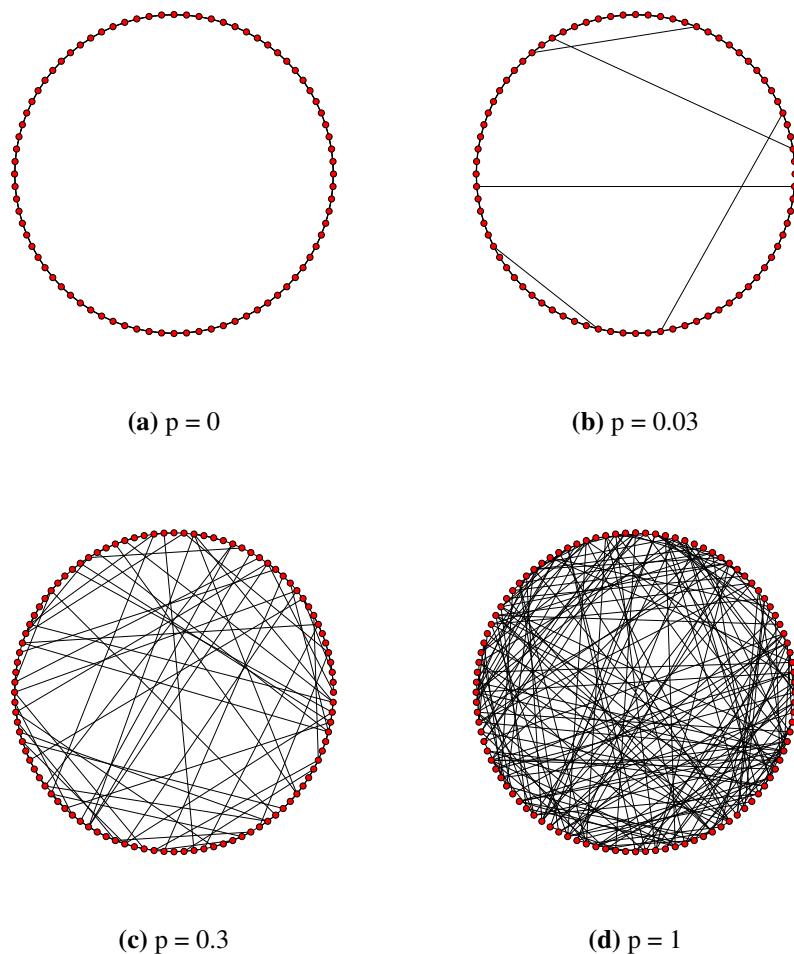


Figura 2: Generazione di grafi con il modello Watts-Strogatz a differenti p di rewiring.

con $C(0) \sim 0.75$ per N grande. Ricordando che per il modello di Erdős-Renyi $C_{ER} = p$, e che per modellizzare una rete reale p solitamente è abbastanza piccola dato che $\langle k \rangle = Np$, risulta che per valori di $p \lesssim 0.25$ si hanno velocemente C_{WS} sensibilmente maggiori di C_{ER} .

Concludendo, un grafo è considerato small-world se ha contemporaneamente piccole distanze medie e, a differenza dei grafi random, una clusterizzazione relativamente elevata. Pertanto il modello di Watts e Strogatz descrive in maniera soddisfacente reti che abbiano queste caratteristiche, purché non siano scale-free.

1.3 Reti scale-free

Molte importanti reti reali di grandi dimensioni hanno la notevole caratteristica di avere un certo livello di invarianza di scala. Questa è manifestata da una distribuzione del grado che segue una funzione a legge di potenza $P(k) \sim k^{-\gamma}$, con γ compreso quasi sempre tra 2 e 3, in maniera sempre più esatta al crescere di k . Mentre le reti random possono riprodurre una $P(k)$ arbitrario, e quindi anche una *power-law*, ma non riescono a generare grafi connessi e con $\langle l \rangle$ non definibile (**Barabási 2002**), le reti small-world

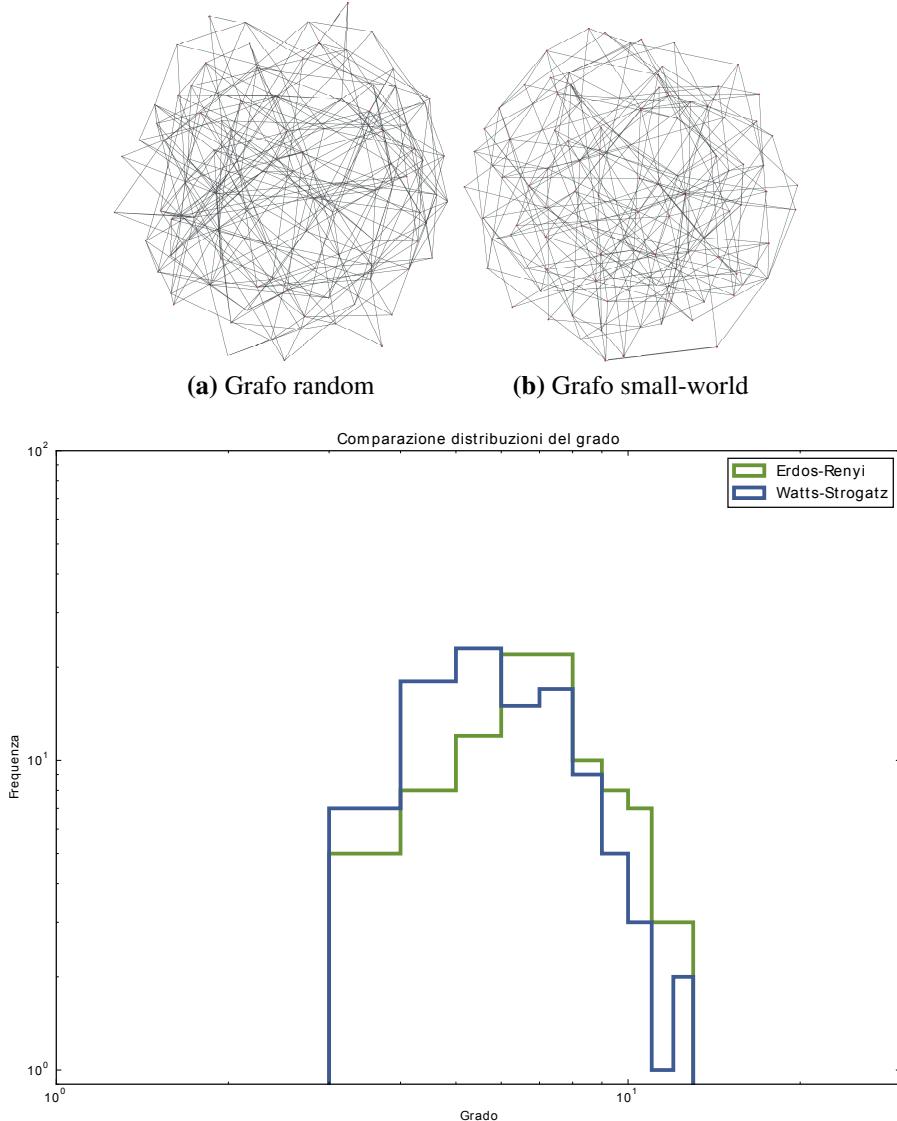


Figura 3: Confronto tra due grafi da 100 nodi: uno generato con il modello Erdos-Renyi ($p = 7\%$) e uno con il modello Watts-Strogatz partendo da un reticolo con nodi connessi ai primi 3 vicini e il 100% di probabilità di rewiring. A parte la somiglianza dei grafi, le distribuzioni del grado hanno simile comportamento.

riproducono bene le proprietà di clustering e cammini medi, ma non possono portare a $P(k)$ power-law. Serve pertanto un modello che riproduca piccoli diametri, grandi clustering e $P(k)$ a legge di potenza.

I modelli di Erdős-Renyi e Watts-Strogatz partono da una configurazione di nodi e poi distribuiscono o riarrangiano i link con una certa probabilità *uniforme* per tutti i link. Le reti reali, tuttavia, spesso partono da un certo numero piccolo di nodi e successivamente crescono. Il primo punto chiave del modello formulato da Barabasi e Albert è proprio il concetto di crescita: la rete parte a $t = 0$ con m_0 nodi e a ogni step temporale si aggiunge alla rete un nodo con un certo numero $m < m_0$ di link da assegnare agli altri nodi esistenti. Il secondo riguarda il fatto che la probabilità di *attachment* di un nuovo nodo agli altri non è uniforme su tutti i nodi ma preferenziale: il *preferential attachment* dà quindi una maggiore probabilità $\Pi(k_i)$ di collegamento di un nodo nuovo a un certo

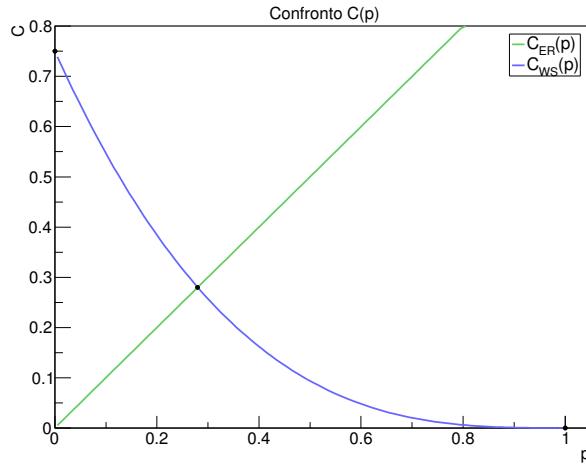


Figura 4: Confronto tra i coefficienti di clustering teorici dei due modelli, al variare del rispettivo parametro p .

nodo i in maniera proporzionale al suo grado, secondo la formula

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}.$$

La costruzione della rete avviene in modo dinamico, pertanto il grado di un nodo i sarà funzione crescente del suo tempo di vita nella rete, con una probabilità di attachment dei nuovi nodi a i anch'essa dipendente da t . Pertanto, con l'andare del tempo il grado di i aumenterà sempre più velocemente. Approssimando k_i a una variabile continua per t grandi, dato che

$$\frac{dk_i}{dt} = m\Pi(k_i) = m \frac{k_i}{\sum_{j=1}^{N-1} k_j} = m \frac{k_i}{2mt - m} = \frac{k_i}{2t - 1} \sim \frac{k_i}{2t},$$

dove m è il numero di link aggiunti per iterazione, ponendo quindi la condizione iniziale per il nodo i , $k_i(t_i) = m$

$$\Rightarrow k_i(t) = m \left(\frac{t}{t_i} \right)^{\frac{1}{2}}.$$

Ovviamente se il grado di un nodo è funzione di t , anche la $P(k)$ lo sarà. Definendo $P(k_i(t) < k)$ la probabilità che un nodo i abbia un grado minore di un certo valore k , la distribuzione del grado $P(k)$ può essere derivata ponendola uguale a $dP(k_i(t) < k)/dk$, ottenendo per $t \rightarrow \infty$:

$$P(k) \sim 2m^2 k^{-3}.$$

Il modello di Barabasi e Albert è un modello minimale, importantissimo perché il primo in grado di riprodurre un grafo che mostrasse un'invarianza di scala, ma che in alcuni aspetti mal si accorda con quelle reti reali con caratteristiche scale-free. Per cominciare la distribuzione del grado è una legge di potenza con esponente 3, ma le reti scale-free solitamente hanno un esponente inferiore, seppur maggiore di 2. Benché non sia possibile calcolare in modo analitico $\langle l \rangle$ e C per una rete di Barabasi-Albert, da simulazioni numeriche è possibile ottenerne quantità caratteristiche.

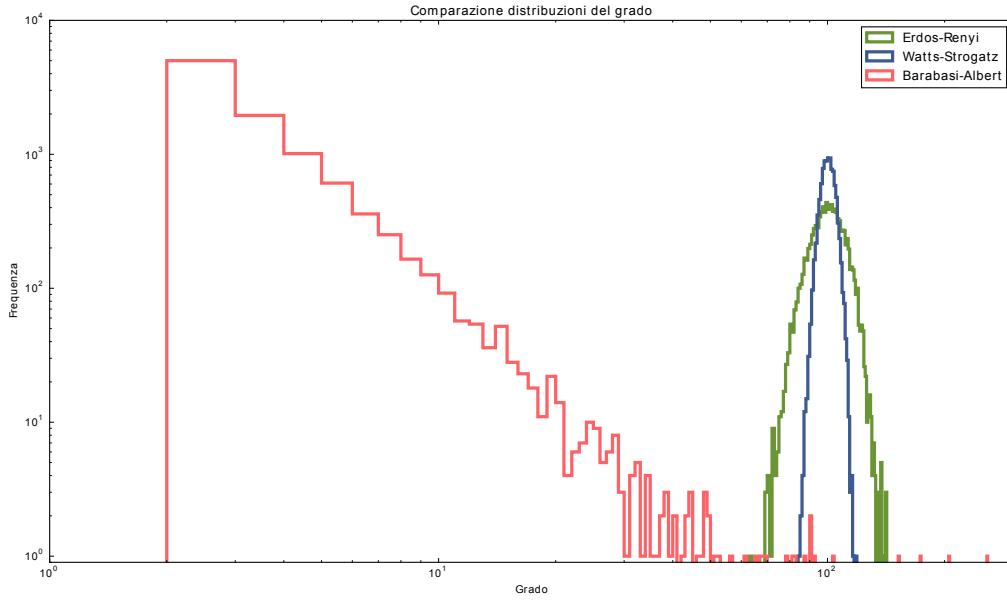


Figura 5: Confronto tra le distribuzioni del grado dei tre modelli esposti.

Il cammino medio risulta sottostimato rispetto alle reti reali e il coefficiente di clustering non è costante con l'aumentare di N come per le reti reali ma diminuisce, anche se più lentamente di una rete random² ([Barbalbert2002](#)). Per questo motivo il modello iniziale è stato integrato e reso più complesso.

1.4 Percolazione

Per percolazione, in analogia con i liquidi, si intende il passaggio di informazione da un lato all'altro di una rete. Nello specifico, quando si fa uno studio percolativo di una rete, si cerca l'esistenza o meno di un cluster che ricopre l'intera rete (*spanning cluster*, o *giant cluster*). Il giant cluster è definito come quella componente di una rete le cui dimensioni sono dello stesso ordine della dimensione della rete complessiva (e quindi gli altri nodi o cluster minori hanno dimensioni di ordini inferiori), cioè il rapporto tra le sue dimensioni e le dimensioni dell'intera rete è ~ 1 .

Tipicamente uno studio percolativo su una rete random consiste nell'individuare, durante la costruzione della rete partendo da N nodi scollegati, la soglia percolativa p_c , cioè la probabilità di creazione link tale che compaia lo spanning cluster. Nel caso di una rete costruita con il modello di Barabasi-Albert la questione non si pone, dato che per definizione questa è composta da un solo cluster in crescita progressiva. Inoltre, dato che una rete di antenne cellulari nasce con una progettazione molto dettagliata in modo che venga costruita già oltre la soglia percolativa, nel caso in esame non è praticabile né interessante individuare p_c in costruzione. In questo caso ha più senso individuare la soglia percolativa per distruzione della rete.

Uno studio per distruzione può avere più approcci che portano a soglie diverse:

²Ovviamente il coefficiente di clustering dipende dal numero m di link generati per ogni nuovo nodo. Con $m = 1$ il clustering è nullo dato che il grafo è un albero, mentre $C \neq 0$ per $m > 1$ e cresce all'aumentare di m , mantenendosi comunque al di sotto dei valori di reti reali con parametri simili.

- se si parte da una rete con uno spanning cluster, rimuovendo link tra tutte le copie di nodi connesse con una certa probabilità, si avrà una p_d oltre la quale la rete non percola più. In questo caso $p_{d_{link}} = p_c$ soltanto se si parte da un grafo completamente connesso.
- in alternativa si possono rimuovere i nodi in maniera random. In questo caso, rimuovendo un nodo si rimuovono in un colpo solo tutti i link a esso collegati, portanto quindi a una $p_{d_{nodo}} \neq p_{d_{link}}$. Si può dare una probabilità maggiore ai nodi con grado più alto per simulare un attacco mirato.
- in maniera opposta a un attachment progressivo, può essere usata un metodo di distruzione progressiva, nel quale si rimuove un nodo per volta e si simula la rimozione successiva con la rete risultante dalla rimozione precedente. Nel caso di una rimozione random, la differenza di questo metodo progressivo con quello diretto è poco significativa, ma risulta molto differente con l'attacco intenzionale, come spiegato in maniera più dettagliata nel paragrafo 3.1.

1.4.1 Soglia percolativa

Vedere le due pubblicazioni di cohen e erez

2 Costruzione e analisi delle reti

3 Network breakdown

3.1 Simulazione

4 Conclusioni

Come