

Analisi di Reti complesse: distribuzione di grado, transizione di fase e robustezza agli attacchi

Freni Salvatore 1037493

Gennaio 2024

1 Introduzione

Un grafo è un insieme di elementi detti nodi o vertici che possono essere collegati fra loro da linee chiamate archi o link. Esistono due tipi di connessioni tra i nodi e si classificano in:

- **Indiretti:** Non c'è una direzione preferenziale nel link.
- **Diretti:** C'è una direzione preferenziale, pertanto non esiste una simmetria nei collegamenti.

Si definisce, invece, grado k di un nodo il numero di archi che il nodo possiede; per cui, un generico grafo può essere descritto dalla sua distribuzione di grado, cioè la probabilità p_k per cui, scegliendo in maniera casuale un nodo della rete, esso possieda esattamente un grado k .

I grafi sono ampiamente utilizzati in una vasta gamma di discipline, dalla teoria dei network informatici alla modellazione delle interazioni sociali. Data, quindi, la loro importanza nella descrizione di sistemi complessi, negli anni sono stati creati vari modelli che potessero replicare le principali caratteristiche di una rete reale, quali distribuzione di grado a legge di potenza, il che comporta presenza di hubs (nodi con un grado molto elevato), effetto small-world (distanza media tra nodi piccola), struttura di scala e robustezza ad attacchi casuali.

1.1 Modello di Erdős–Rényi

Il primo modello realizzato è quello di Erdős–Rényi, il quale costruisce un grafo a partire dagli iperparametri N numero di nodi e p probabilità che tra due nodi esista un link. Viene indicato con $G(N, p)$ e, data la sua definizione, segue la distribuzione binomiale:

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

Essa, per $p \rightarrow 0$ e $N \rightarrow \infty$ con la condizione $\nu = Np$ costante, tende ad una distribuzione di Poisson con parametro ν .

$$P(k) = \frac{\nu^k e^{-\nu}}{k!} \quad (1)$$

Questa caratteristica rappresenta il principale difetto del modello: la distribuzione non è a legge di potenza.

Componente gigante: La più grande componente connessa all'interno del grafo, cioè il sottoinsieme massimo di nodi nel grafo in cui esiste un percorso tra ogni coppia di nodi.

Si può dimostrare che la condizione necessaria affinché ci sia una componente gigante è

$$\langle k \rangle = 1$$

1.2 Modello di Barabási–Albert

Il modello di Barabási–Albert è un modello generativo che si basa sull'attaccamento preferenziale.

Inizialmente si generano m_0 nodi completamente connessi e, successivamente, ad ogni step si aggiunge un nodo con l ($l \leq m_0$) link ai nodi già esistenti. La probabilità che un nodo si attacchi ai nodi i con grado k_i è proporzionale a quest'ultimo:

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

Questo procedimento comporta ad avere una distribuzione di grado a legge di potenza e quindi alla formazione di pochi hubs e un numero elevato di nodi con grado basso.

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

1.3 Attacco ad una rete

Gli attacchi alle reti costituiscono un approccio fondamentale per valutare la robustezza dei grafi mediante la rimozione strategica di nodi.

Per poter condurre un'analisi dettagliata è necessario studiare come cambia la frazione di nodi nella componente connessa rispetto alla frazione di nodi rimossi. Le strategie di attacco sono diverse, ma due approcci comuni sono gli attacchi random e quelli basati sul grado dei nodi. Negli attacchi random, i nodi vengono rimossi casualmente, simulando eventi distruttivi imprevedibili. Gli attacchi basati sul grado, invece, mirano ai nodi con il grado più elevato, che svolgono un ruolo chiave nella struttura della rete.

Si dimostra che, in generale, per attacchi random la frazione critica f_c sopra la quale la rete fallisce può essere espressa come $f_c = 1 - \frac{1}{\frac{\langle k^2 \rangle}{\langle k \rangle} - 1}$, che per il modello di Erdős–Rényi diventa

$$f_c^{ER} = 1 - \frac{1}{\langle k \rangle}$$

Le reti scale-free sono altamente vulnerabili ad attacchi basati sul grado, in quanto gli hubs hanno un ruolo fondamentale nell'intera struttura del grafo. Questa vulnerabilità comporta un valore critico molto basso.

2 Modello

Il modello è composto da tre classi: *Node*, *Graph* e *GraphAnalyzer*.

La classe *Node* istanzia dei nodi, caratterizzati da un Id, dal grado e un vettore di giacenza in cui sono presenti i nodi a cui è connesso. In questa classe si possono creare e rimuovere connessioni.

La classe *Graph* crea dei grafi indiretti che possono essere basati sia sul modello di Erdős–Rényi che sul modello di Barabási–Albert. Essi sono caratterizzati dalla lista di nodi che li compongono. Fornisce metodi per aggiungere o rimuovere archi, ottenere la componente connessa più grande e condurre attacchi random o basati sul grado.

La classe *GraphAnalyzer* si occupa dell'analisi avanzata del grafo. Si può ottenere la generazione della distribuzione di grado e l'esecuzione di un fit basato sulla distribuzione teorica prevista. Inoltre, produce un grafico che visualizza la presenza della componente connessa nelle reti Erdős–Rényi, plottando il grado medio nelle ascisse e la frazione di nodi nella componente

connessa gigante nelle ordinate (lo scopo qui è osservare una transizione di fase per $\langle k \rangle \approx 1$). Infine, il modello consente la visualizzazione grafica degli attacchi, potendo così ottenere una stima della frazione critica.

2.1 Esecuzione del codice

In ROOT si eseguono queste linee di codice:

```
.x compile_my_class.cpp
.L simulation.C+
```

Successivamente si possono eseguire le varie funzioni presenti in *simulation*:

- *simulationER(int seed, unsigned int n, double p)* che analizza la distribuzione di grado di una rete Erdős–Rényi con n nodi e probabilità p e conduce due attacchi, uno random e l'altro basato sul grado.
- *simulationBA(int seed, unsigned int n, unsigned int m_0, unsigned int l)*, in modo analogo, analizza un grafo di Barabási–Albert caratterizzato da n nodi, m_0 nodi iniziali e l link creati ad ogni step.
- *simulationPhaseTransition_2(int seed, unsigned int n_sim)*, produce n_{sim} simulazioni il cui scopo è analizzare la presenza della transizione di fase nella presenza di una componente gigante nel modello di Erdős–Rényi.

3 Risultati

3.1 Fit delle distribuzioni

3.1.1 Erdős–Rényi

Si considera come funzione teorica la distribuzione di Poisson, che, come detto in 1.1, rappresenta l'andamento limite.

Si considera fisso il numero $\nu = Np = 20$ e si fanno variare i due iperparametri. L'ipotesi nulla è che l'istogramma segua la distribuzione teorica di Poisson. Il p-value ottenuto dal fit viene confrontato con un valore critico, ad esempio $\alpha = 0.05$. Se il p-value è maggiore di α , si accetta l'ipotesi che l'istogramma segue la distribuzione ipotizzata.

Come previsto, nel limite $p \rightarrow 0$ e $N \rightarrow \infty$, c'è una miglior adesione dei dati alla funzione teorica, il che implica che, effettivamente, il modello segue

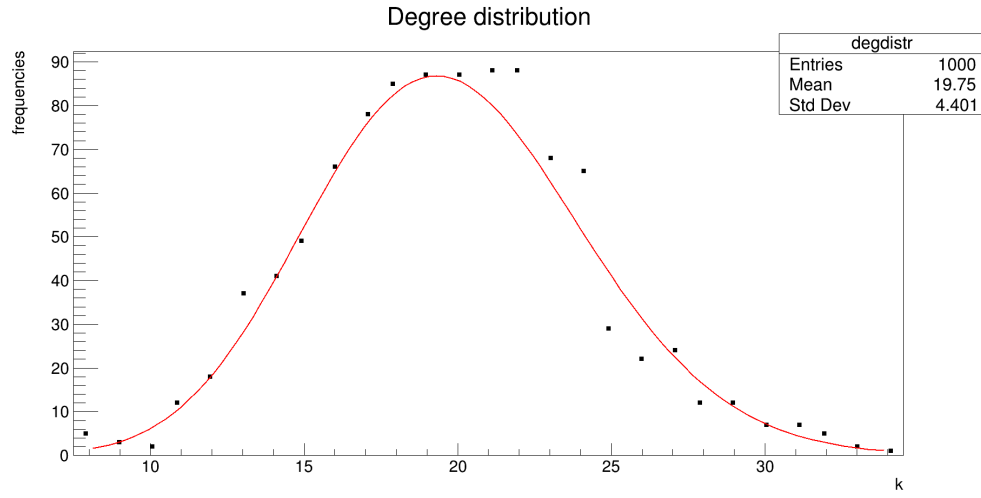


Figure 1: $N = 1000$, $p = 0.02$, $\chi^2 = 32.82$, $ndf = 23$, $p_{value} = 0.08$, $\hat{\nu} = 19.81 \pm 0.14$

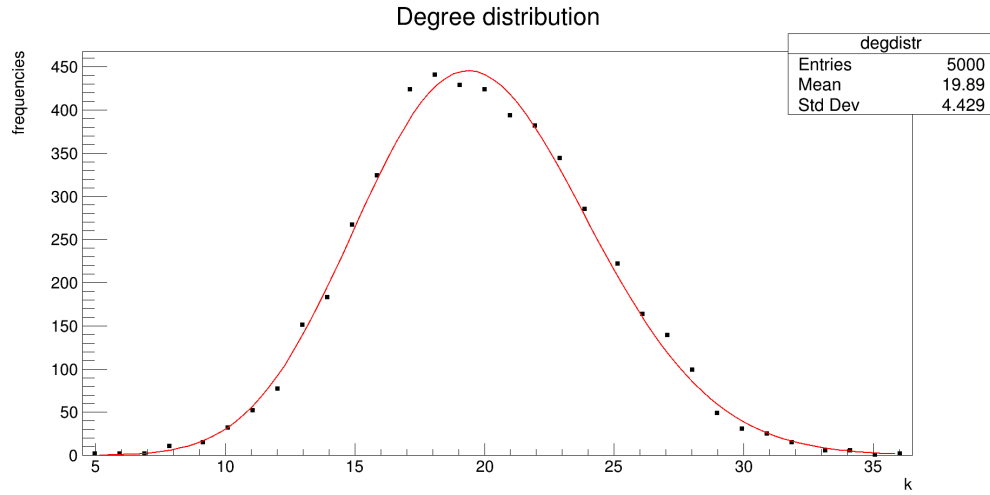


Figure 2: $N = 5000$, $p = 0.004$, $\chi^2 = 31.75$, $ndf = 28$, $p_{value} = 0.28$, $\hat{\nu} = 19.88 \pm 0.06$

la distribuzione di Poisson. Il valore teorico $\nu = 20$, tuttavia, non rientra, in nessuno dei casi considerati, nell'intervallo di confidenza del valore stimato $\hat{\nu}$.

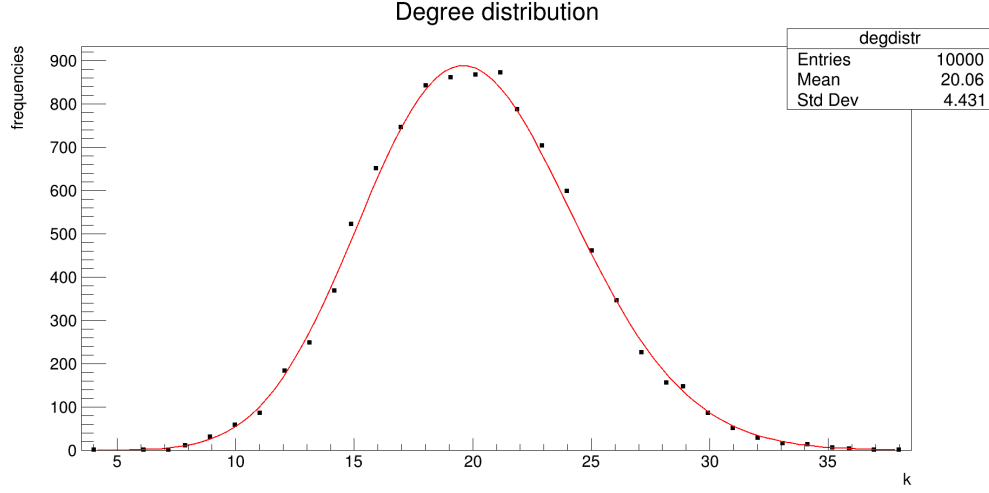


Figure 3: $N = 10000$, $p = 0.002$, $\chi^2 = 35.24$, $ndf = 32$, $p_{value} = 0.32$, $\hat{\nu} = 20.09 \pm 0.04$

3.1.2 Barabási–Albert

Come discusso in [1] Capitolo 4 - Advanced Topic 3.C - Estimating the Degree Exponent, poiché è difficile eseguire il fit sull'intero range di valori di grado osservati (k_{min} e k_{max}), la scelta del cut-off ($K_{min} > k_{min}$) è cruciale per ottenere una stima affidabile dell'esponente γ . Fatto ciò per vari valori di K_{min} , si ottengono le coppie (K_{min}, γ) e si esegue il test di Kormogorov-Smirnov per scegliere la coppia ottimale.

In questa analisi, invece, si procede nel seguente modo: si scelgono i parametri per inizializzare il grafo, si esegue il fit considerando il cutoff $K_{min} = k_{min} + \frac{1}{10}(k_{max} - k_{min})$ (si è scelto questo valore con l'obiettivo di bilanciare la necessità di coprire un range significativo di gradi osservati con la limitazione pratica di evitare valori troppo estremi che potrebbero influenzare negativamente il fit), si ottiene la stima di γ e si valuta il p-value dal valore del χ^2 risultante.

Il modello creato è conforme con l'aspettazione teorica, in quanto $\gamma_t = 3 \in [\gamma - \Delta\gamma, \gamma + \Delta\gamma]$, eccetto il caso $N = 5000$ in cui probabilmente il punto di cut-off è errato e si deve rimodulare.

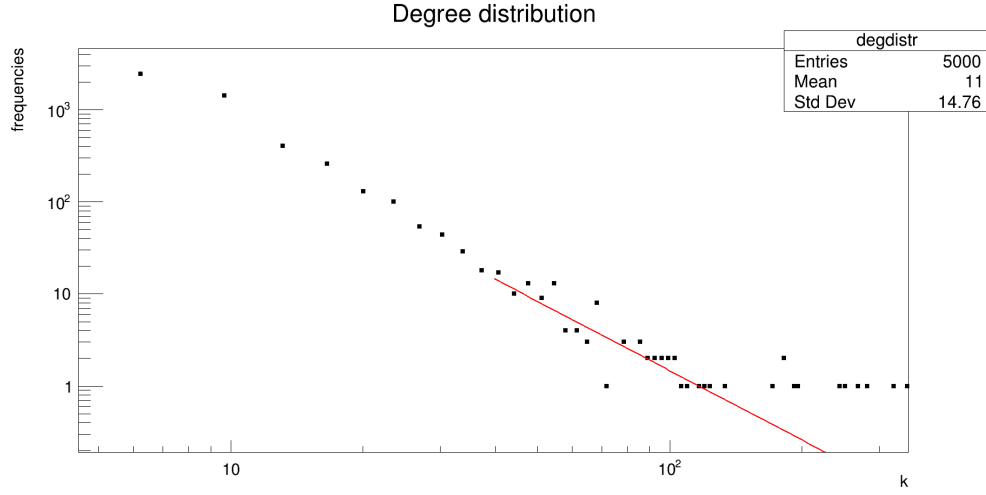


Figure 4: $N = 5000$, $m_0 = 10$, $l = 5$, $\chi^2 = 22.76$, $ndf = 31$, $p_{value} = 0.86$, $\hat{\gamma} = 2.50 \pm 0.39$

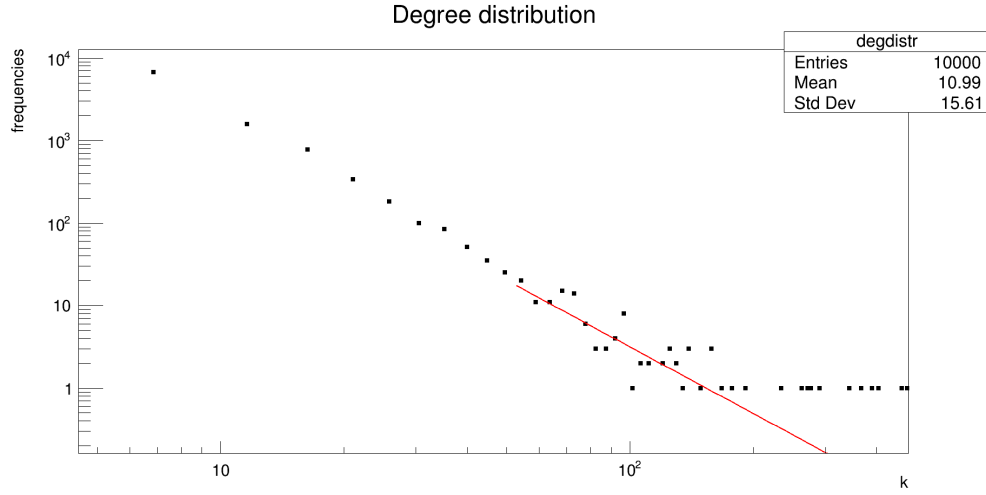


Figure 5: $N = 10000$, $m_0 = 10$, $l = 5$, $\chi^2 = 28.06$, $ndf = 32$, $p_{value} = 0.67$, $\hat{\gamma} = 2.68 \pm 0.37$

3.2 Componente gigante

In questa simulazione si costruiscono 20 grafi Erdős-Rényi con p fissato e N variabile che va da 0 a N_{max} tale per cui $\langle k \rangle_{max} = N_{max}p = 2$. L'obiettivo

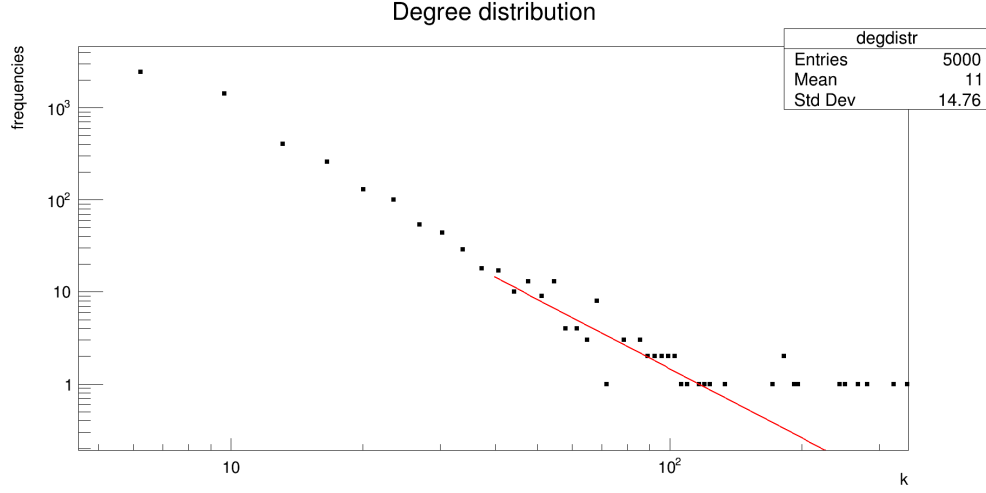


Figure 6: $N = 20000$, $m_0 = 10$, $l = 5$, $\chi^2 = 25.20$, $ndf = 30$, $p_{value} = 0.71$, $\hat{\gamma} = 2.66 \pm 0.34$

è graficare la frazione di nodi nella componente gigante in funzione di $\langle k \rangle$ e verificare che esistono due regioni: sub-critica ($\langle k \rangle < 1$) in cui la componente gigante comprende una frazione irrisoria del numero totale di nodi del grafo e super-critica ($\langle k \rangle > 1$) in cui la componente gigante è predominante, in particolare questo regime ha rilevanza, poiché nelle reti reali la componente gigante ha quasi le dimensioni della rete stessa. La transizione di fase, come detto, avviene per $\langle k \rangle \approx 1$ ed è indipendente dal valore della probabilità p .

La simulazione svolta, utilizzando tre valori di $p \in [0.001, 0.0001]$ equidistanziati tra loro, è coerente con i risultati teorici. In tutti i casi si registra la transizione di fase e l'andamento sembra essere indipendente dalla scelta del valore di p .

3.3 Attacchi alle reti

Per ogni rete creata vengono condotti due tipi di attacchi, uno basato sull'attacco casuale e l'altro basato sul grado dei nodi. In particolare, ogni attacco è composto da 20 steps, durante i quali viene rimosso un numero costante di nodi. Durante ogni step vengono registrate sia la frequenza totale dei nodi rimossi dal grafo che il rapporto di nodi che compongono la componente gigante.

Poiché i risultati ottenuti sono analoghi, si presenta un solo set di grafici

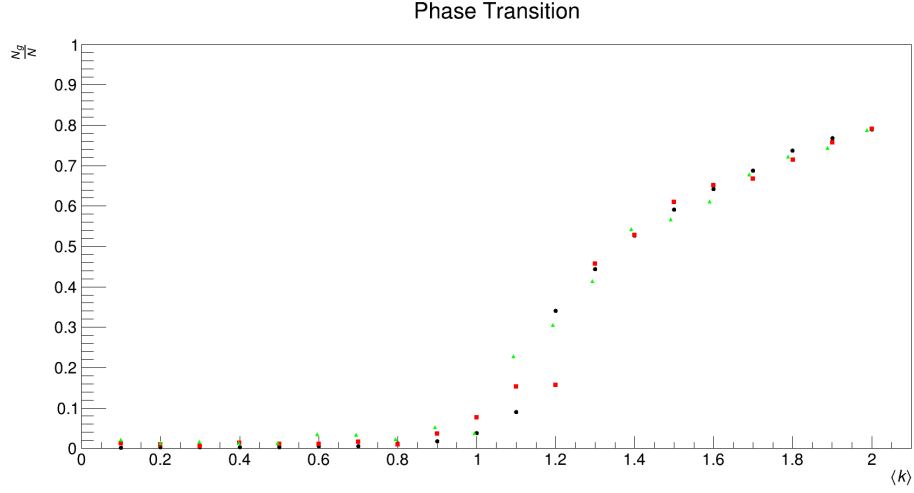
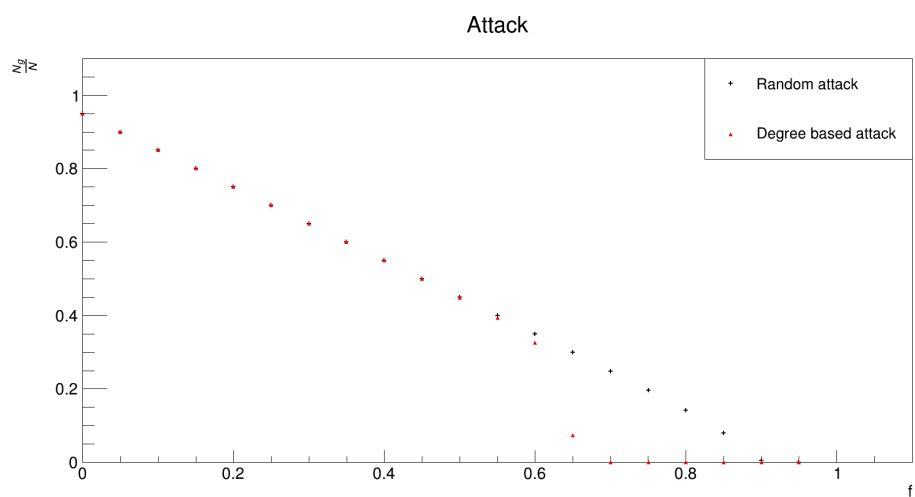


Figure 7: Transizione di fase e comparsa della componente gigante

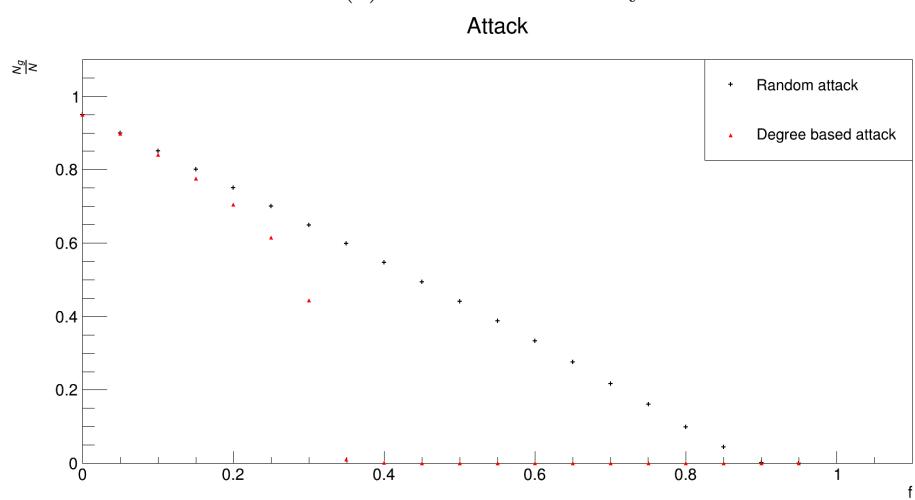
rappresentativi, ognuno per un tipo di grafo creato. In questa sezione, analizziamo i risultati ottenuti da un grafo di Erdős–Rényi in una regione supercritica, caratterizzato da $N = 10000$ nodi e una probabilità di collegamento $p = 0.002$ e quelli ottenuti da un grafo di Barabási–Albert composto da $N = 10000$ nodi, generato con $m_0 = 10$ nodi iniziali e $l = 5$ link creati ad ogni step. Questi valori sono stati scelti per evidenziare al meglio le differenze che sussistono tra i due modelli.

Come evidenziato dalla Figura 8a, si osserva che per l'attacco random, la frequenza critica stimata è $f_c \approx 0.95$, che coincide con il valore teorico ottenuto da $f_c = 1 - \frac{1}{\langle k \rangle}$. D'altra parte, l'attacco basato sul grado risulta più potente e incide in modo significativo sulla struttura della rete.

In generale, si può affermare, con riscontri anche nelle reti reali, che il modello di Barabási–Albert sia più robusto e resiliente ad attacchi puramente random. Mentre, e si può evincere dalla Figura 8b, è più vulnerabile ad attacchi mirati, in quanto gli hubs sono elementi fondamentali della struttura della rete stessa, per cui, una volta rimossi, essa crolla e non presenta più le caratteristiche originali.



(a) Rete di Erdős–Rényi



(b) Rete di Barabási–Albert

Figure 8: Risultati di attacchi

4 Conclusioni

In conclusione, si può asserire che il modello basato su tecniche Monte-Carlo per la simulazione della generazione di grafi presenta le medesime caratteristiche che emergono dallo studio teorico.

Il modello di Erdős–Rényi, nel limite di $N \rightarrow \infty$ e $p \rightarrow 0$, segue una distribuzione di Poisson, rendendolo meno idoneo per rappresentare reti reali. Esso presenta una regione critica in cui si verifica la formazione di una componente gigante per $\langle k \rangle > 1$.

Dall'altra parte, il modello di Barabási–Albert riflette le caratteristiche principali di reti complesse reali, data la distribuzione a legge di potenza, che comporta la formazione di hubs, struttura di scala e robustezza ad attacchi casuali. Tuttavia, ciò comporta anche una minore resistenza ad attacchi basati sul grado. Questa caratteristica può rappresentare uno svantaggio in contesti come reti elettriche o di trasporto aereo, in quanto una rimozione di un hub può portare al collasso dell'intera rete, ma risulta fondamentale nella gestione delle epidemie, dove la vaccinazione mirata degli elementi più sensibili può limitare la diffusione della malattia.

Bibliografia

- [1] Albert-László Barabási. *Network Science*. Cambridge University Press, USA, 1st edition, 2016.