# Analisi e ricerca di componenti fortemente connesse (CFC)

Leonardo Gori

January 24, 2021

**Abstract**

I grafi sono strutture dati che rappresentano un insieme di entità, dette 'nodi' messe in relazione tra loro da collegamenti, detti archi. Essi hanno vari ambiti applicativi nel mondo dell'informatica quali la gestione della rete, dei processi di un sistema operativo o della rappresentazione di una mappa geografica. Fondamentale tra le caratteristiche di un grafo è il concetto di componente fortemente connessa che permette di gestire in modo ottimale le varie connessioni che interessano i nodi che le compongono. In questo articolo verrà analizzata la correlazione tra una moltitudine di grafi creati in modo casuale per numero di nodi e archi, e il numero di componenti fortemente connesse presenti al loro interno.

## 1 Introduzione

Una componente fortemente connessa all'interno di un grafo è un sottografo massimale i quali nodi sono tutti comunicanti tra di loro, ovvero esiste un cammino che collega ogni nodo agli altri all'interno della stessa componente. L'algoritmo usato per la ricerca di una componente fortementa connessa si basa su nodi con caratteristche particolari, quali:

- tempo di scoperta
- tempo di fine scoperta
- puntatore al padre
- colore:
  - bianco se non ancora scoperto
  - grigio se scoperto ma l'analisi dei nodi adiacenti non è completa
  - nero se completamente analizzato

Le componenti fortemente connesse vengono utilizzate per trovare gruppi di entità correlate in un vasto insieme di dati. Ne è un esempio applicativo il suggerimento di amicizia sui social media, dove gli utenti rappresentano i nodi e il rapporto di amicizia gli archi che li collegano.

## 2 Algoritmo di ricerca di CFC

L'algoritmo di ricerca delle componenti fortemente connesse di un grafo è basato sul tipo di ricerca DFS e la struttura dati nota come lista di adiacenza.

### 2.1 Depth-First Search (DFS)

Depth-First Search è un algoritmo per la ricerca di nodi all'interno di un grafo che opera partendo da un vertice casuale del grafo e visita i nodi adiacenti in modo ricorsivo. Si contrappone all'algoritmo di ricerca BFS (Breadth-First Search) per la caratteristica di essere ricorsivo e per il fatto di adottare una modalità di esplorazione dei nodi che visita in profondità piuttosto che in ampiezza. Come suggerisce il nome infatti DFS se applicato sulla radice di un albero, prima dei vertici in prossimità del nodo di partenza, può visitare prima quelli più lontani dalla radice. Il risultato finale di DFS è un albero che è un sottografo del grafo sul quale viene applicato.

### 2.2 Lista di adiacenza

Una lista di adiacenza è una lista che associa a ogni nodo del grafo i vertici che sono collegati a esso da un arco. Si contrappone alla matrice di adiacenza per il fatto di essere più efficiente nella lettura dei nodi di adiacenza, ma meno ottimizzata per la ricerca della presenza di un determinato arco. In ogni caso, ai fini di questa relazione, l'uso di una delle due strutture dati è indifferente. Entrambe vengono utilizzate per definire in modo semplice il grafo trasposto su cui verrà applicata la DFS per contare il numero di CFC.

### 2.3 Strongly-Connected-Components

Questo algoritmo applica una DFS sul grafo G preso in considerazione, dopo di che crea un grafo trasposto $G^T$ sul quale applica una seconda DFS che analizza i vertici in ordine di fine analisi decrescente. Ai fini di questa relazione non è necessario conoscere da quali nodi sono composte le CFC, l'informazione fondamentale è scoprirne il numero. Per fare ciò l'algoritmo considera i nodi i quali il puntatore al padre è nullo, poichè saranno i primi ad essere visitati nella seconda DFS, tutti gli altri faranno parte di una CFC.

## 3 Test

Il progetto allegato alla relazione si occupa di analizzare il numero di componenti fortemente connesse contenute in una successione di grafi che varia per numero di nodi e connessioni. In particolare sono stati eseguiti vari test fissando a priori i range relativi a dimensione e probabilità di presenza di archi della successione di grafi. Il numero di connessioni tra nodi è determinato da un parametro di probabilità (e.g. con un parametro di probabilità del 20%, per ogni nodo i,j del grafo esiste il 20% di probabilità che venga generato l' arco i⇒j). Inoltre per ogni dimensione e parametro di probabilità fissato, vengono generati n grafi (di default n=100), calcolate le relative presenze di CFC e successivamente calcolata la media. I risultati dei test rappresentano quindi il numero medio di componenti connesse in rapporto alla dimensione del grafo e alla probabilità dell'esistenza di archi tra i nodi.

### 3.1 Aspettative

Secondo la teoria per dei grafi di dimensione fissata, per valori di probabilità maggiori aumenta il numero di connessioni tra i nodi, e quindi la probabilità che esista un cammino che li colleghi tutti; perciò ci si aspetta che il numero di componenti fortemente connesse converga a 1 all' aumentare del parametro di probabilità. Inoltre a parità di probabilità, all'aumentare del numero di nodi ci si aspetta che aumenti anche il numero di cammini possibili che colleghino tutti i nodi, e perciò anche la probabilità che i collegamenti generati formino un'unica componente connessa.

# 4 Risultati degli esperimenti

## 4.1 Al variare delle probabilità

Viene presentata una tabella che mostra i risultati del numero medio di componenti connesse (calcolato su 100 prove) rilevate al variare delle probabilità di presenza di archi (comprese tra il 0,05% e l'1% per un passo di 0.05%), con grafi di dimensioni fissata di 1000 nodi:

### 4.1.1 Dimensione fissata: 1000 nodi

| Probabilità | Numero medio di CFC |
|---|---|
| 0.05% | 999.660 |
| 0.1% | 987.050 |
| 0.15% | 665.520 |
| 0.2% | 367.130 |
| 0.25% | 206.270 |
| 0.3% | 116.310 |
| 0.35% | 66.200 |
| 0.4% | 40.680 |
| 0.45% | 23.490 |
| 0.5% | 15.290 |
| 0.55% | 9.820 |
| 0.60% | 5.540 |
| 0.65% | 4.010 |
| 0.70% | 2.830 |
| 0.75% | 2.020 |
| 0.8% | 1.760 |
| 0.85% | 1.390 |
| 0.90% | 1.170 |
| 0.95% | 1.130 |
| 1.0% | 1.130 |

Table 1: Numero di componenti connesse al variare della probabilità di presenza di archi per grafi di dimensione di 1000 nodi

Come si evince dai risulati descritti nella tabella, si può notare come all'aumentare della probabilità di presenza di archi, aumenta il numero di connessioni e diminuisce quindi il numero di componenti connesse distinte fino a tendere in modo asintotico al valore 1.

## 4.2 Al variare della dimensione

Vengono presentati una tabella e un grafico che mostrano i risultati del numero medio di componenti connesse (calcolato su 100 prove) rilevate al variare delle dimensioni dei grafi (comprese tra 10 nodi e 300 nodi di passo 10), per probabilità di presenza di archi fissata al 5%:

### 4.2.1 Probabilità fissata: 5%

| Dimensione | Numero medio di CFC |
|---|---|
| 10 | 9.830 |
| 20 | 17.900 |
| 30 | 20.450 |
| 40 | 16.960 |
| 50 | 11.150 |
| 60 | 7.740 |
| 70 | 5.340 |
| 80 | 4.190 |
| 90 | 2.750 |
| 100 | 2.220 |
| 110 | 1.780 |
| 120 | 1.570 |
| 130 | 1.290 |
| 140 | 1.230 |
| 150 | 1.090 |
| 160 | 1.060 |
| 170 | 1.100 |
| 180 | 1.040 |
| 190 | 1.030 |
| 200 | 1 |

Table 2: Numero di CFC al variare del numero di nodi per grafi con probabilità di presenza di archi del 5% (i restanti grafi di dimensione maggiore di 200 sono stati omessi in quanto contenevano una media di una componente connessa)

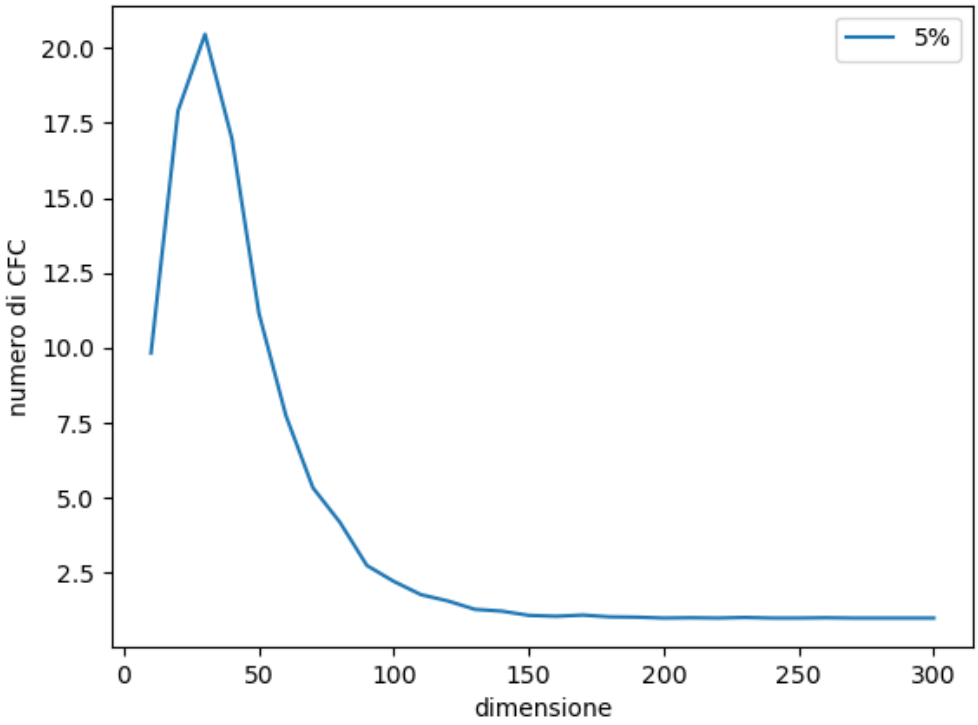


Figure 1: Grafico che descrive il numero medio di CFC (calcolato su 100 prove) al variare del numero di nodi che compongono il grafo e probabilità di presenza di archi fissata a 5%.

Dai risultati precedenti si può notare che a parità del parametro di probabilità, il numero di componenti connesse raggiunge un picco massimo per poi convergere al valore 1. Ciò può essere spiegato dal fatto che grafi troppo piccoli presentino poche possibilità di cammini che colleghino tutti i nodi e perciò la probabilità che collegamenti generati formino un'unica componente connessa. Per grafi di dimensioni maggiori invece il numero di cammini possibili aumenta in modo esponenziale e diminuisce perciò il numero di componenti connesse distinte.

### 4.3 Al variare di probabilità e dimensione

Di seguito un grafico che confronta il numero medio di CFC (calcolato su 100 prove) al variare di entrambi i parametri di dimensione dei grafi(compresa tra 3 e 30 nodi di passo 3) e probabilità di presenza di archi (tra lo 0% e il 100% di passo 5%):

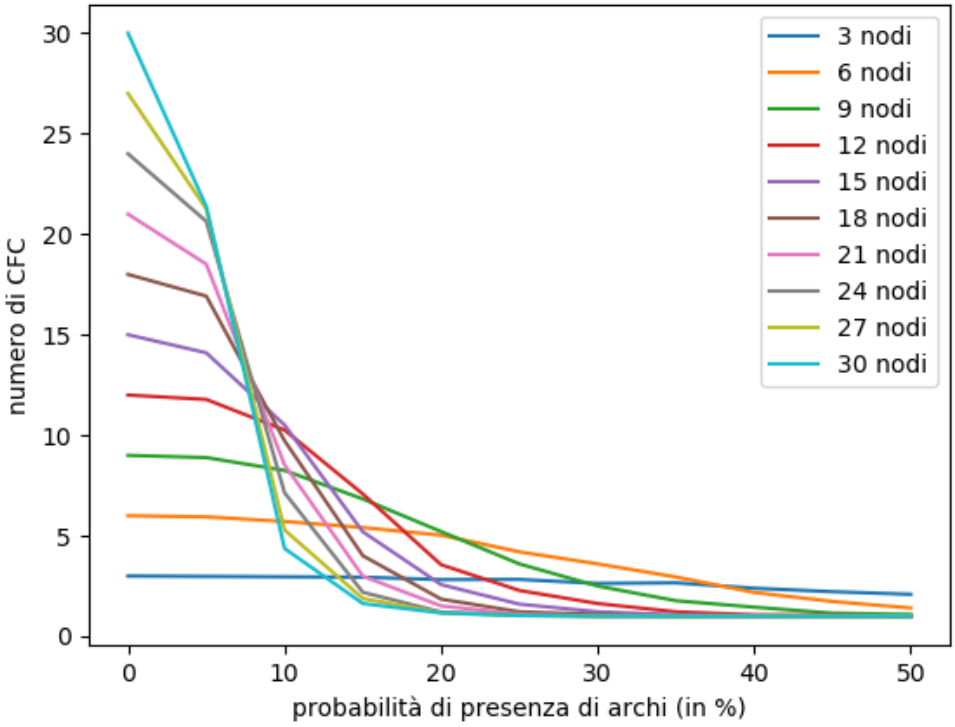


Figure 2: Grafico che descrive il numero medio di CFC al variare del numero di nodi che compongono il grafo e di probabilità di presenza di archi. N.B.: all'aumentare della probabilità di presenza di archi, grafi di dimensione maggiore tendono a contenere meno componenti fortemente connesse rispetto a grafi di dimensione minore.