

Report promessi sposi e miserabili

Ginevra Maria Cepparulo

Corso: Network Analytics

Anno 2022-2023

Introduzione

Il seguente report analizza le reti 'I Promessi sposi' e 'I miserabili', facendo un confronto tra di loro e con i rispettivi modelli random e scale-free.

La rete i promessi sposi è basata sull'omonimo celebre romanzo della letteratura italiana scritto da Alessandro Manzoni e reputato il più letto tra le opere in lingua italiana.

La rete i miserabili è basata sul romanzo storico 'Les Miserables' di Victor Hugo, questo considerato uno dei più eccelsi romanzi del XIX secolo.

In entrambe le reti, i nodi rappresentano i personaggi del romanzo e gli indiretti archi indicano interazioni tra i nodi coinvolti. La presenza dell'arco indica quindi che i due personaggi hanno interagito almeno una volta. Il peso dell'arco indica quante volte hanno interagito nel romanzo.

I promessi sposi



Figura 1: Rete 'I Promessi Sposi'

Il layout della rete rappresentato nella *Figura 1* è stato ottenuto tramite gli algoritmi Fruchterman Reingold e Expansion. In questo layout il colore del nodo rappresenta il grado, più scuro è il colore più alto è il grado del nodo. Invece, la grandezza del nodo rappresenta il clustering coefficient, più grande è il nodo più alto è il clustering coefficient.

Il random graph creato con lo stesso **average degree** della rete 'I Promessi Sposi' è rappresentato nella *Figura 2*. Questo grafo è nel regime connected perché $k > \ln(N)$. Il numero di connected components è 1 con dimensione N , come rappresentato dalla tabella

indicata nella *Figura 7*. La *Figura 7*, inoltre, dimostra che il **average clustering coefficient** del modello random è molto più basso della rete 'I Promessi Sposi', e anche il **diametro** e l'**average path length** sono più bassi. Giungiamo alla conclusione che il modello random non è adeguato per rappresentare questo network.

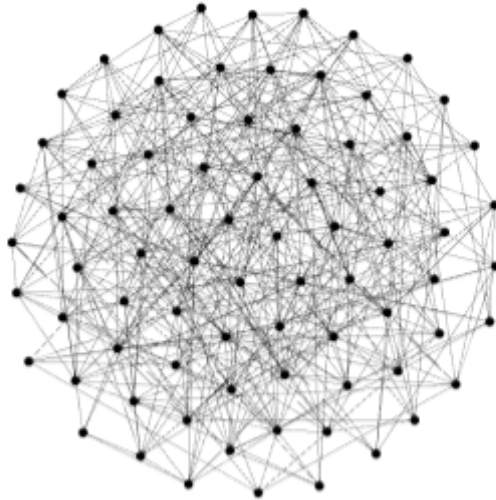


Figura 2: Modello random 'I Promessi Sposi'

Il modello Watts-Strogatz è stato ottenuto ed è rappresentato nella *Figura 3*. Per ottenere un **average clustering coefficient** e **average path length** simile alla rete 'I Promessi Sposi', il rewiring probability è 0.15.

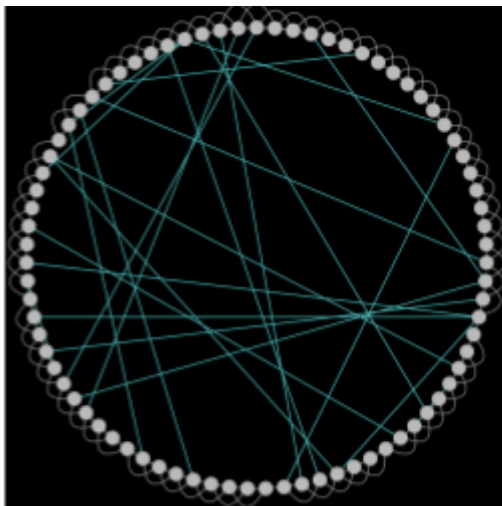


Figura 3: Modello Waltz-Strogatz 'I Promessi Sposi'

I miserabili

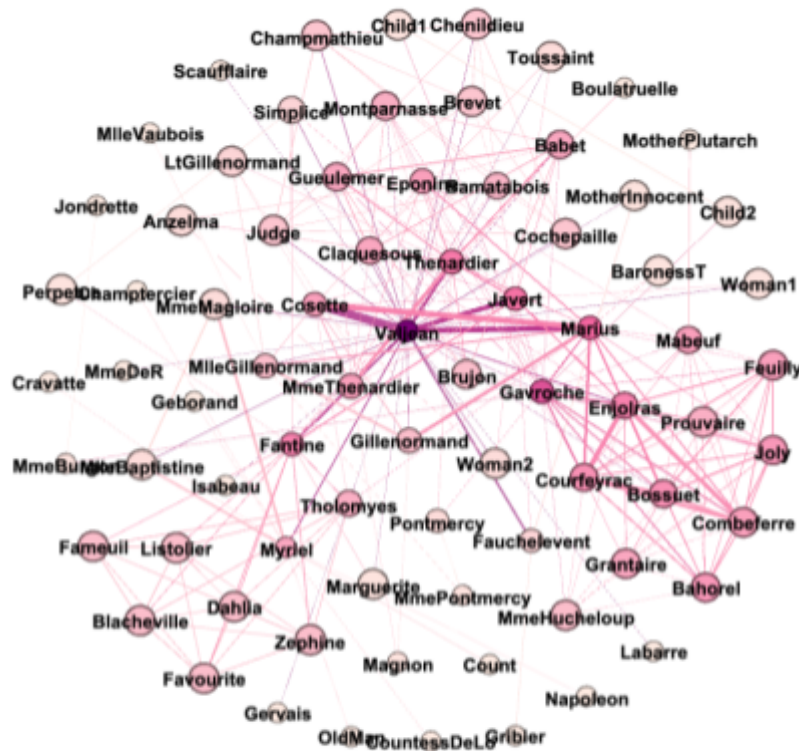


Figura 4: Rete 'I Miserabili'

Il layout della rete 'I Miserabili', rappresentato nella *Figura 4*, è stato ottenuto con la stessa metodologia della rete 'I Promessi Sposi'.

Il random graph creato con lo stesso average degree della rete 'I Miserabili' è rappresentato nella *Figura 5*. Questo grafo è nel regime connected perché $k > \ln(N)$. Il numero di connected components è 1 con dimensione N , come indicato nella tabella in *Figura 7*. La *Figura 7*, inoltre, dimostra che il **clustering coefficient** del modello random è molto più basso della rete 'I miserabili', mentre il **diametro** è uguale. Giungiamo alla conclusione che il modello random non è adeguato per rappresentare questo network però è migliore rispetto al caso dei 'Promessi Sposi'.

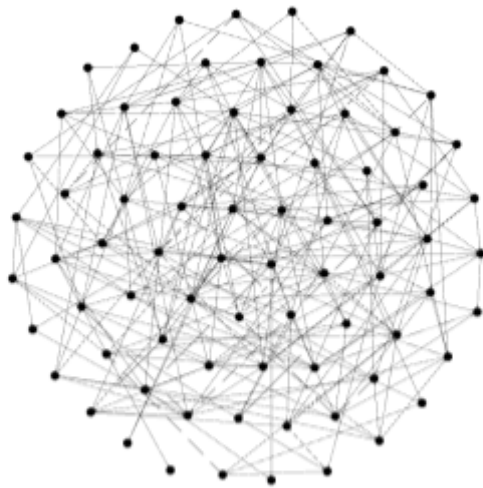


Figura 5: Modello Random 'I Miserabili'

Il modello Watts-Strogatz è stato ottenuto ed è rappresentato nella *Figura 3*. Per ottenere un **average clustering coefficient** e **average path length** simile alla rete 'I Miserabili', il rewiring probability è 0.08.

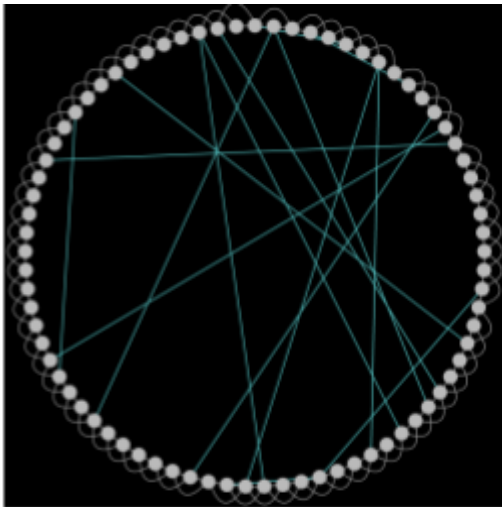


Figura 6: Modello Waltz-Strogatz 'I Miserabili'

Analisi

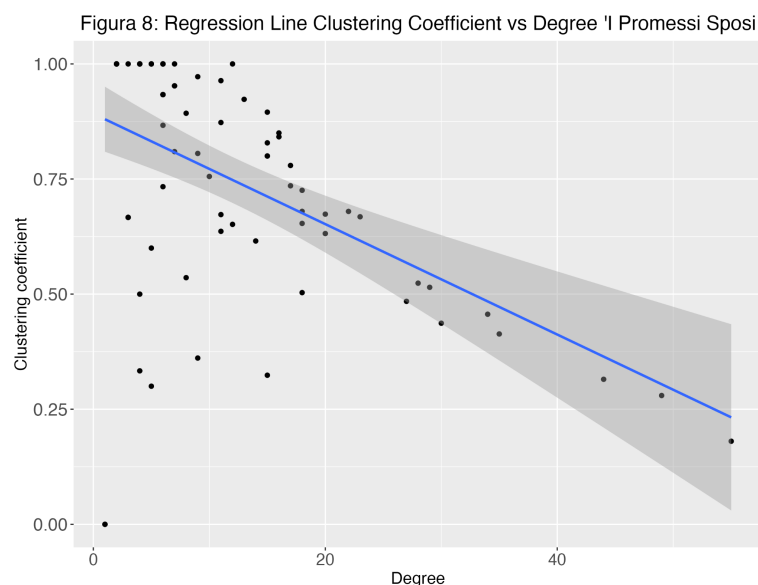
La tabella 1 dimostra alcuni parametri statistici analizzati.

Possiamo notare che la rete di miserabili ha molti meno archi rispetto alla rete dei promessi sposi. Infatti per la rete i miserabili l'average path length e l'average degree sono più alti mentre l'average clustering coefficient più basso.

	L	N	avg PL	avg CC	d	ncc	NG	k	p
promessi sposi	473	79	2,052	0,768	4	NA	NA	11,975	0,154
random sposi	466	79	1,99	0,159	3	1	79	11,797	0,151
miserabili	254	77	2,641	0,736	5	NA	NA	6,597	0,087
random miserabili	271	77	2,434	0,115	5	1	77	7,039	0,093

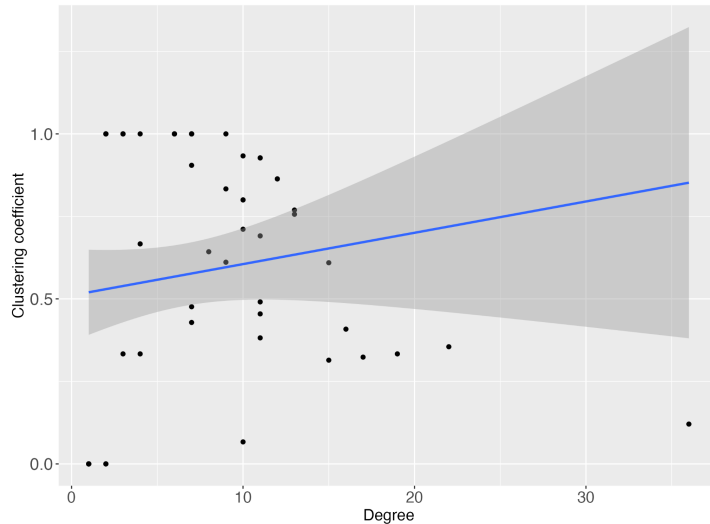
Figura 7: Tabella

Clustering coefficient vs degree



La correlazione tra il Degree e il Clustering coefficient nella rete 'I Promessi Sposi' è raffigurata nel plot in *Figura 8*. Le colorazioni grigie rappresentano i 95% confidence intervals, notiamo quindi che sono più ampi per gradi più alti, questo è dovuto al numero minore di osservazioni con nodi di grado alto.

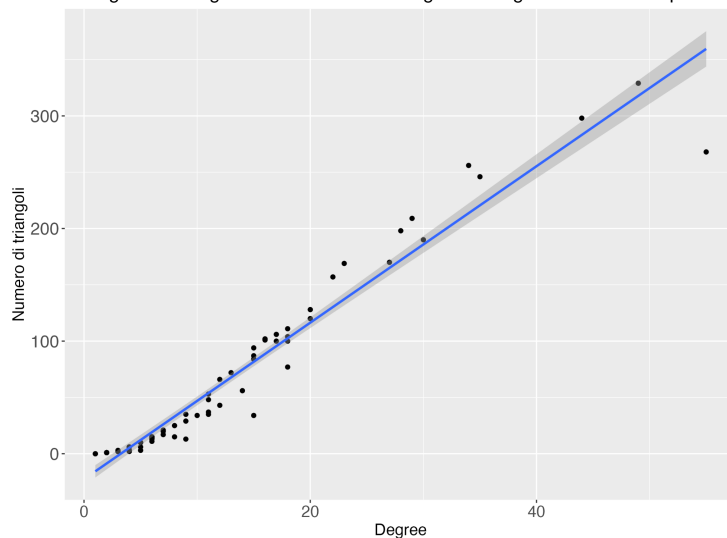
Figura 9: Regression Line Clustering Coefficient vs Degree 'I Miserabili'



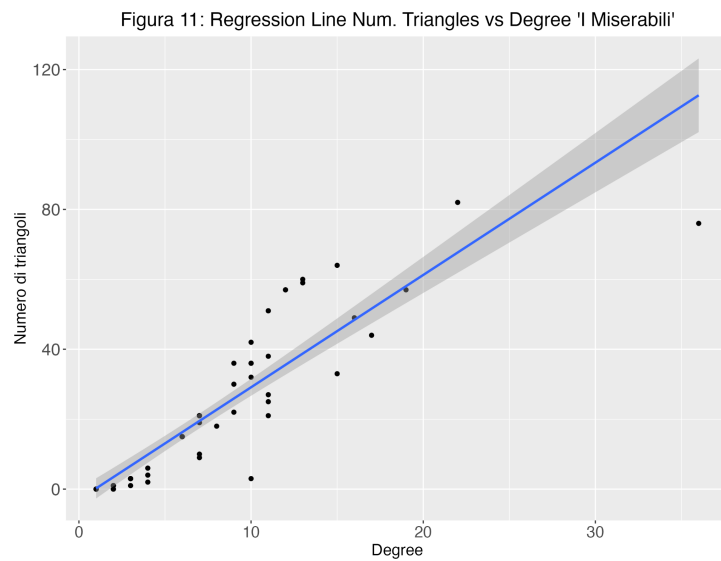
La correlazione tra il Degree e il Clustering coefficient nella rete 'I Miserabili' è raffigurata nel plot in *Figura 9*. Notiamo che i confidence interval sono molto ampi per gradi più alti, questo è dovuto al numero minore di osservazioni con nodi di grado alto. In particolare c'è un outlier con grado 36 che si potrebbe rimuovere.

Numero di triangoli vs degree

Figura 10: Regression Line Num. Triangles vs Degree 'I Promessi Sposi'

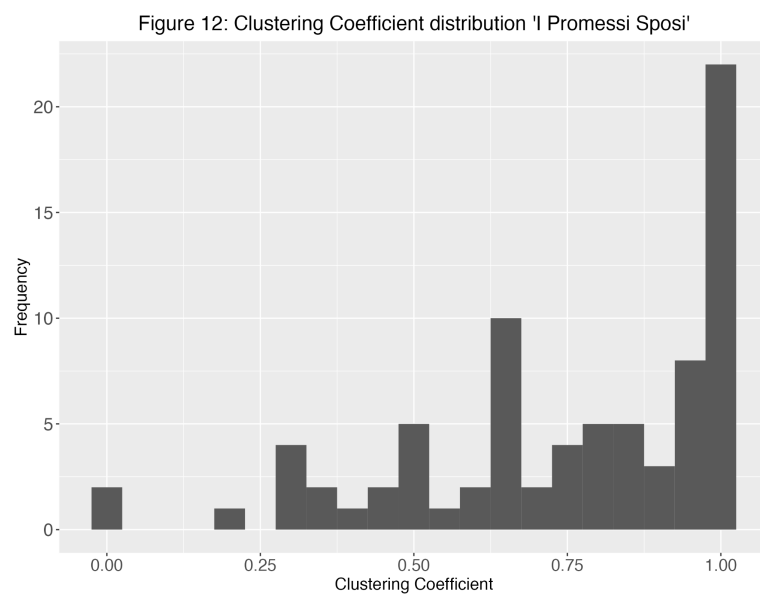


La correlazione tra il Degree e il Numero di triangoli nella rete 'I Promessi Sposi' è raffigurata nel plot in *Figura 10*. Notiamo che in questo caso i confidence interval sono stretti rispetto all'analisi del Clustering coefficient vs degree.

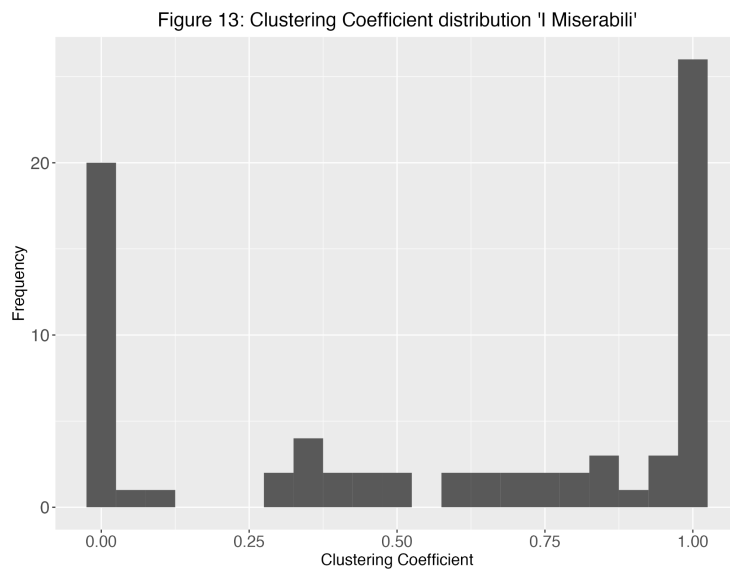


La correlazione tra il Degree e il Numero di triangoli nella rete 'I Miserabili' è raffigurata nel plot in *Figura 11*. Notiamo che in questo caso i confidence interval sono più larghi per gradi alti rispetto alla rete 'I promessi sposi'.

Clustering coefficient distribution



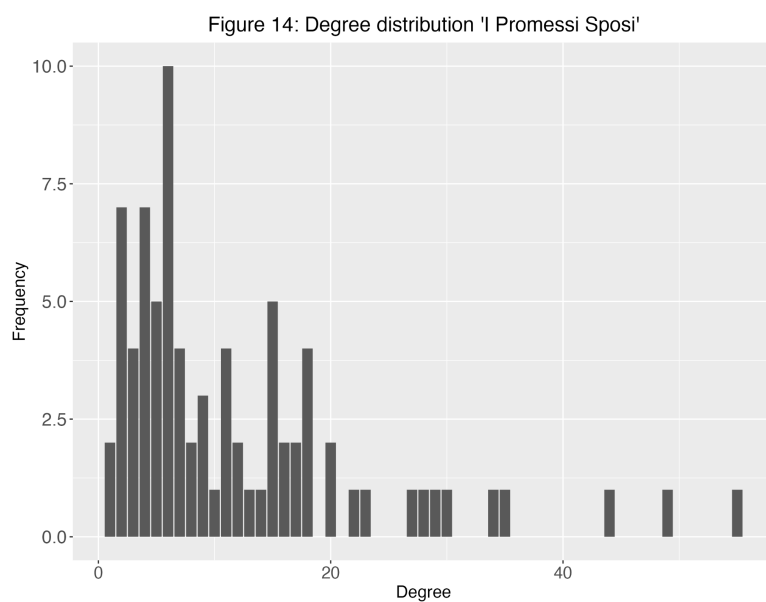
La *Figura 12* mostra la distribuzione del clustering coefficient della rete 'I Promessi Sposi'.



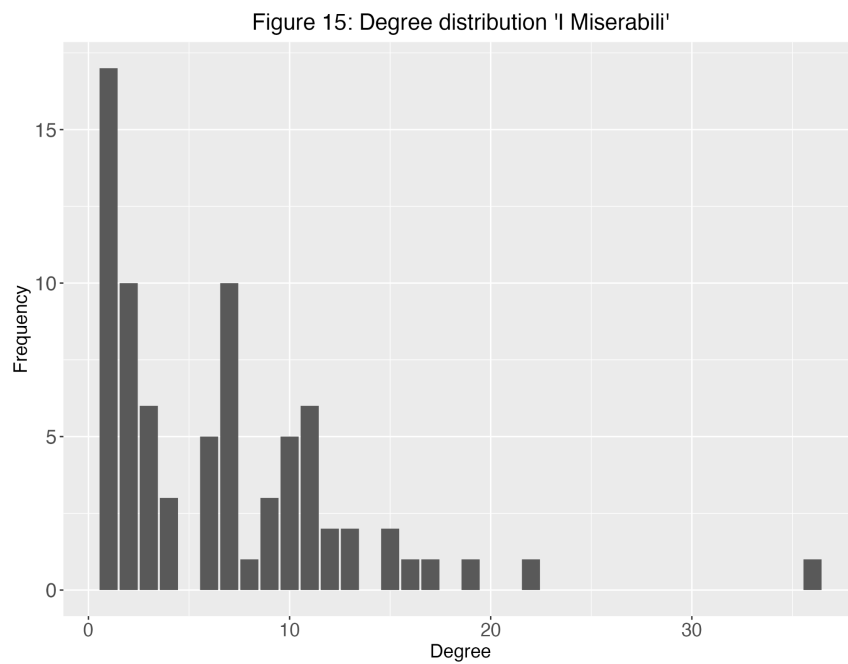
La *Figura 13* mostra la distribuzione del clustering coefficient della rete 'I Miserabili'.

Confrontandole, possiamo notare che nella rete 'I Miserabili' si ha una distribuzione meno uniforme, vi è inoltre presente alta frequenza per clustering coefficient basso.

Degree distribution



La *Figura 14* mostra la distribuzione del grado della rete 'I Promessi Sposi'. Possiamo notare che ha una somiglianza a una poisson, come dovrebbe essere se il grafo fosse modellato da un modello random.



La *Figura 15* mostra la distribuzione del grado della rete 'I Miserabili'. Anche qui possiamo notare una somiglianza ad una poisson.

Goodness of fit with exponential distribution

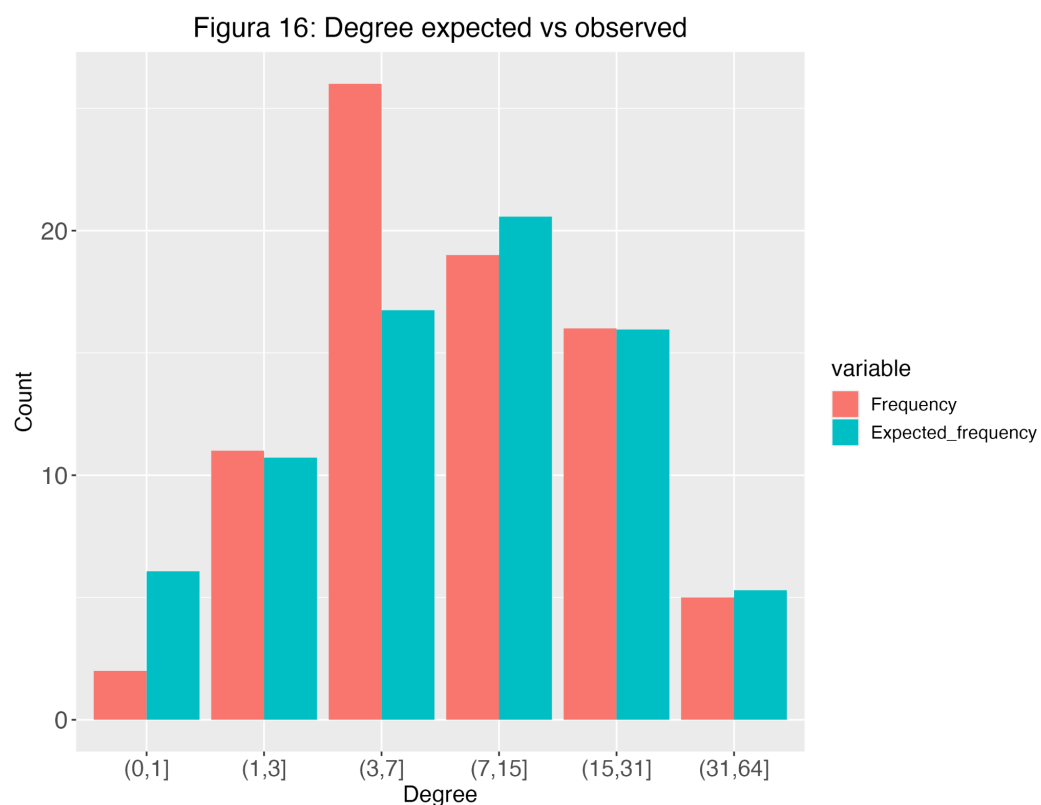


Figura 16 mostra la frequenza attesa di una distribuzione esponenziale per il grado rispetto alla frequenza osservata. Abbiamo anche eseguito una aggregazione logaritmica. Quindi è possibile notare la differenza con il grafico nella *Figura 14*.

Il goodness of fit test ha rilevato un valore $X^2 = 7.989832$ e un valore $p\text{-value} = 0.1138064$. Per cui siamo in grado di non rigettare la null hypothesis e dire con 95% confidence che la distribuzione del grado è esponenziale.

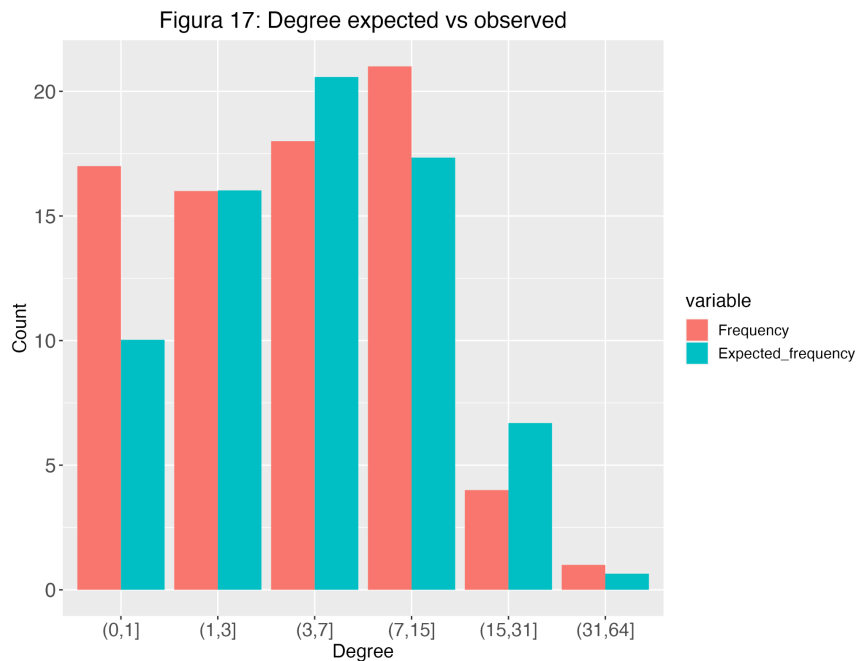


Figura 17 mostra la frequenza attesa di una distribuzione esponenziale per il grado rispetto alla frequenza osservata per la rete 'I Miserabili'. Abbiamo anche eseguito una aggregazione logaritmica. Quindi è possibile notare la differenza con il grafico nella Figura 15.

Il goodness of fit test ha rilevato un valore $X^2 = 6.26298$ e un valore $p\text{-value} = 0.09949358$. Per cui siamo in grado di non rigettare la null hypothesis e dire con 95% confidence che la distribuzione del grado è esponenziale anche per la rete dei 'Miserabili'.

Conclusione

In linea di massima, i modelli random non rappresentano un'adeguata stima delle reti analizzate. Possiamo dire comunque che il grado dei nodi è distribuito esponenzialmente.