

Analisi rete sociale: UCI facebook messaging

DI GINEVRA CEPARULO

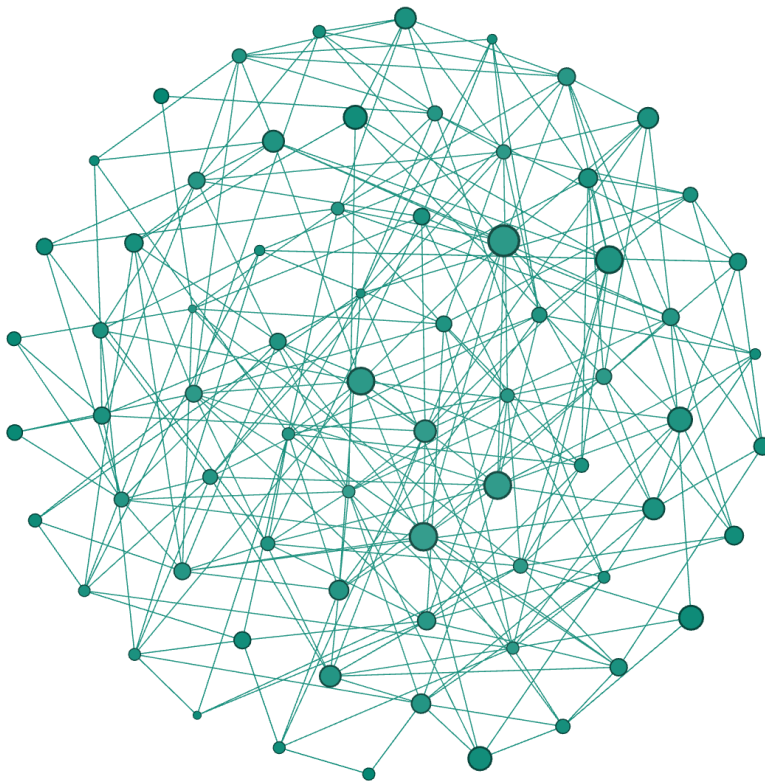
Introduzione

Lo scopo di questa investigazione sarà di analizzare il facebook social network di studenti della UC Irvine (University of California Irvine). Nel network i nodi rappresentano pagine di profilo di studenti e i link rappresentano lo scambio di almeno un messaggio tra gli studenti coinvolti. Purtroppo gli attributi pertinenti alle informazioni personali degli studenti come l'età, il sesso, il corso di laurea, non sono disponibili per motivi relativi a data privacy degli studenti.

Il motivo per cui verrà intrapresa l'analisi di questa rete sociale è che è molto interessante capire quanto è robusta una comunità di studenti sui social network. Questo ci può far capire se le dinamiche sociali universitarie in America sono "sparse" oppure se gli studenti fanno molta amicizia tra di loro. Potrebbe essere interessante comparare un social network simile in Italia e trarre quindi conclusioni su come si differenzia la vita sociale universitaria tra i due paesi.

Analisi descrittiva

La rete studiata è rappresentata nella figura sottostante. Essendo molto grande è difficile capirne la struttura per cui andremo a prendere un subset dei nodi per visualizzarla.



In questo grafo ottenuto usando un sampling dei nodi con il grado tra 30-50, qui appaiono il 5% dei nodi e degli archi. In questa visualizzazione ottenuta usando l'algoritmo Fruchterman Reingold il colore rappresenta il grado. Più i nodi sono tendenti al verde scuro, più il grado è

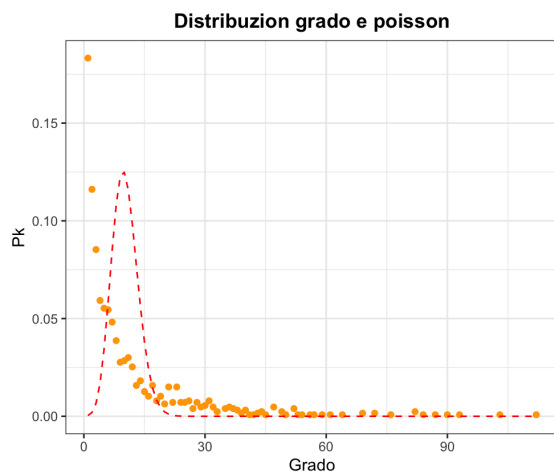
vicino al valore 30, mentre, più sono i nodi tendenti al marrone scuro, più il grado è vicino al valore 50. In questo caso i nodi sono tutti verdi e quindi hanno un grado vicino al valore 30. La grandezza dei nodi è correlata al clustering coefficient.

L'analisi descrittiva iniziale ci consente di capire che la rete è connessa, siccome il numero di componenti connessi è uno con dimensione uguale al numero dei nodi ($\text{num_cc}=1$ e $\text{NG}=\text{N}$). In oltre la rete è triangolata e la dimensione della clique massima è 5 ($\text{dim_clique}=5$).

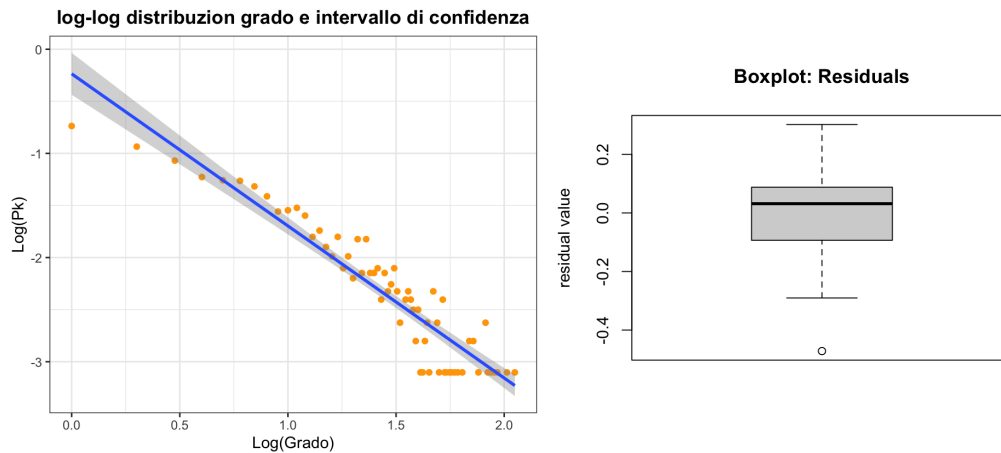
L	N	diameter	k	avg_clustering	NG	num_cc	dim_clique	num_clique
6451	1266	9	10,19	0,0439	1266	1	5	1

Distribuzione del grado dei nodi

Il primo tentativo è comparare la distribuzione del grado dei nodi con quella di un modello Random, una distribuzione di Poisson. Il plot sottostante ci fa intuire che questa distribuzione non è adatta alla nostra rete. Nella distribuzione di Poisson il grado dei nodi è molto più condensato rispetto alla media e inoltre essa sottovaluta il numero di nodi con grado piccolo e il numero di nodi con grado grande.



Possiamo intuire che il grado dei nodi segue una distribuzione di una power law. Il valore stimato del parametro gamma, ottenuto utilizzando un linear regression del log-log plot della distribuzione del grado dei nodi è rappresentato sotto accanto ai rispettivi valori standard error, p-value e R Squared.



estimate_gamma	st_error	p_value	R2
-0.6	0.03	2,00E-16	1.28

Il plot dimostra che l'intervallo di confidenza è molto stretto. Inoltre il p-value, che è molto basso, ci consente di negare l'ipotesi nulla, cioè che il coefficiente angolare è uguale a zero. Dopodichè si può notare il valore R2 è abbastanza alto e poi il plot dei residuals dimostra un valore mediano vicino a zero.

Per cui la stima del parametro della power distribution law è buona.

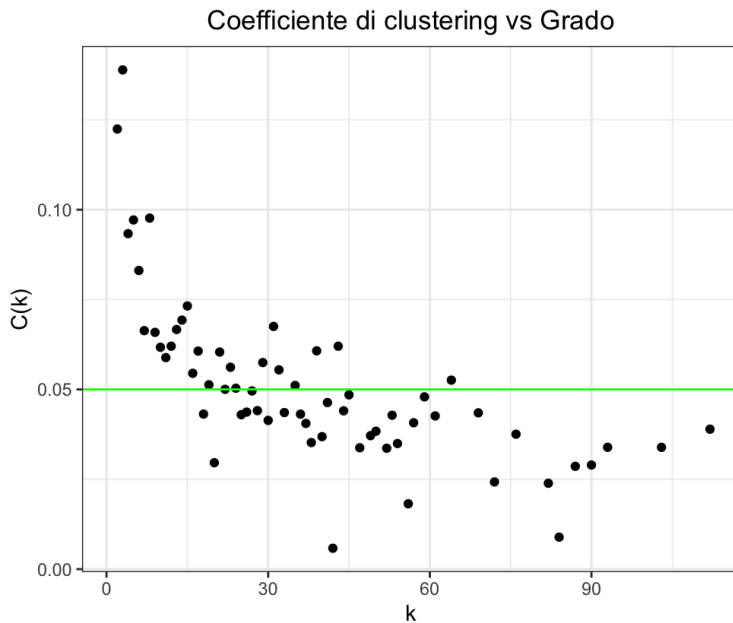
Comparazione modello Random

Il modello Random definito da (N, p) , prevede che la distribuzione del grado dei nodi è una Poisson. Purtroppo come abbiamo mostrato nella sezione 1 la nostra rete sociale non segue una Poisson.

La nostra rete si troverebbe nel regime connesso in quanto $\langle k \rangle > \ln N$ per cui il modello random prevede $NG = N$. Le analisi dimostrano l'esistenza di un singolo componente che contiene tutti i nodi della rete.

Per quanto il fenomeno Small World, la nostra rete sembra obbedirlo. Comparando la distanza media predetta dal modello Random e quella empirica della nostra rete possiamo notare che quella predetta dal modello random si avvicina a quella reale.

Il coefficiente di clustering è correlato al grado dei nodi, come dimostra il plot sottostante. Il modello random non rispecchia questa correlazione.



Comparazione modello Watts-Strogatz

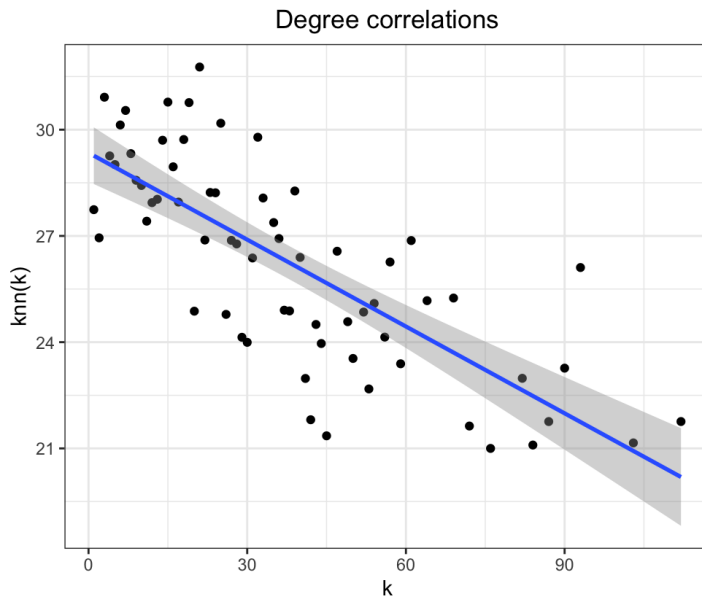
Il modello Watts-Strogatz prevede la Small World property e anche coefficiente di clustering alto. Usano link rewiring il valore di p che consente una buona approssimazione del coefficiente di clustering e della distanza media è $p = 0.35$. Purtroppo però il modello non prevede una correlazione tra $C(k)$ e k (che è presente nella nostra rete) e inoltre prevede una distribuzione del grado dei nodi di tipo Poisson (mentre abbiamo visto che la nostra rete è scale free). Per cui questo modello non è appropriato per descrivere la nostra rete.

Comparazione modello Barabasi-Albert

Il modello Barabasi-Albert differisce dal modello casuale in quanto si basa su due importanti caratteristiche della rete: la crescita e l'attaccamento preferenziale. Il fatto che la nostra rete abbia una scale free property è un'evidenza di attaccamento preferenziale.

Questo modello genera reti scale free con $\gamma = 3$. La rete sociale in questione, ha un valore di $\gamma = 0.6$ come detto precedentemente. Essendo una rete sociale un modello che molto probabilmente descrive la nostra rete è il Copying model.

Un altro aspetto che dimostra l'esistenza di attaccamento preferenziale è correlazione tra il grado dei vicini di un nodo e il suo stesso grado, ossia degree correlations. Le reti che hanno degree associativity cioè una correlazione positiva quindi rappresentano un'evidenza di winner takes all phenomena. Pertanto la nostra rete ha dimostrato la presenza di disassortivity, come potete notare dal punto sottostante.



Pertanto essendo $\gamma < 3$ la ragione per cui la rete presenta structural dissasortivity.

Network Robustness

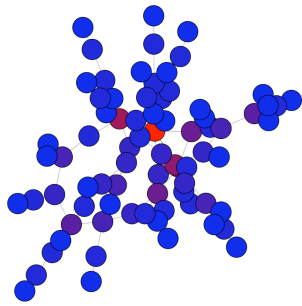
fER	critical_threshold
0.9019	0.9621

Nella rete analizzata, il critical threshold è maggiore del valore predetto per un random network per cui possiamo dire che la rete dimostra enhanced robustness. Il valore si avvicina molto a $f = 1$ per cui è molto robusta. Ciò significa che i legami sociali della rete sono forti.

Comunità della rete

Innanzitutto, notiamo che la rete sociale presenta hierarchical modularity. Come evidenziato, nel plot antecedente il coefficiente di clustering è negativamente correlato al grado dei nodi.

Vari algoritmi sono stati utilizzati per individuare la presenza di comunità nella rete sociale, tra cui hierarchical clustering divisivo usando come centralità edge betweenness oppure random walk oppure l'algoritmo greedy rapido che usa la modularità. Quest'ultimo ha evidenziato la presenza di 18 comunità. Qui sotto vi è un plot di una delle comunità più piccole. Nel sottografo il colore rappresenta il grado dei nodi: blu più basso e rosso più alto.



Tutto sommato l'algoritmo migliore per community detection era l'ultimo soprattutto perchè trattandosi di un network senza attributi o classificazioni decidere il numero di comunità è decisamente più difficile.

Conclusione

Nell'insieme abbiamo scoperto che questa rete sociale è scale free, è molto robusta, ed è composta da molte comunità.

Sources

Rossi , Ryan A., and Nesreen K. Ahmed. "FB-Messages: Interaction Networks." Network Data Repository, networkrepository.com/ia-fb-messages.php. Accessed 28 June 2023.