Who Shall Survive?

Misure di centralità su reti sociali

Una **rete sociale** (*social network*) è una struttura fatta di persone e relazioni tra le persone. I sociologi chiamano **attori** (*actors*) le persone della rete e **legami** (*ties*) le relazioni tra gli attori. In realtà gli attori non sono necessariamente persone: sono anche state studiate **reti sociali tra animali** (scimmie, canguri e delfini).

In una rete sociale le relazioni tra gli attori possono rappresentare diverse cose. Per esempio amicizia, relazioni professionali, relazioni criminali, relazioni amorose o sessuali. Per esempio in **Facebook** la relazione tra le persone è quella dell'amicizia (*friendship*). Notate che questa è una relazione **simmetrica**: se A è a amico di B allora B è amico di A.

In **Twitter** la relazione rappresenta il seguire (*follow*) altri utenti della rete perché ci interessano i loro brevi pensieri o cinguetii (*tweets*). Ogni utente Twitter ha dunque un insieme di **seguitori** (*followers*) che seguono l'utente, e un insieme di **seguiti** (*followings*) a cui l'utente è interessato. In tal caso la relazione tra utenti non è necessariamente simmetrica ma è **direzionale**: posso seguire i pensieri di Britney Spears ma non è detto che lei segua i miei.

Per la maggior parte delle persone una rete sociale significa un servizio quale *Facebook* o *Twitter*. In realtà, lo studio delle reti sociali è molto più antico di questi servizi. I sociologi sono probabilmente i ricercatori che hanno la tradizione di studio delle reti sociali più longeva e consolidata.

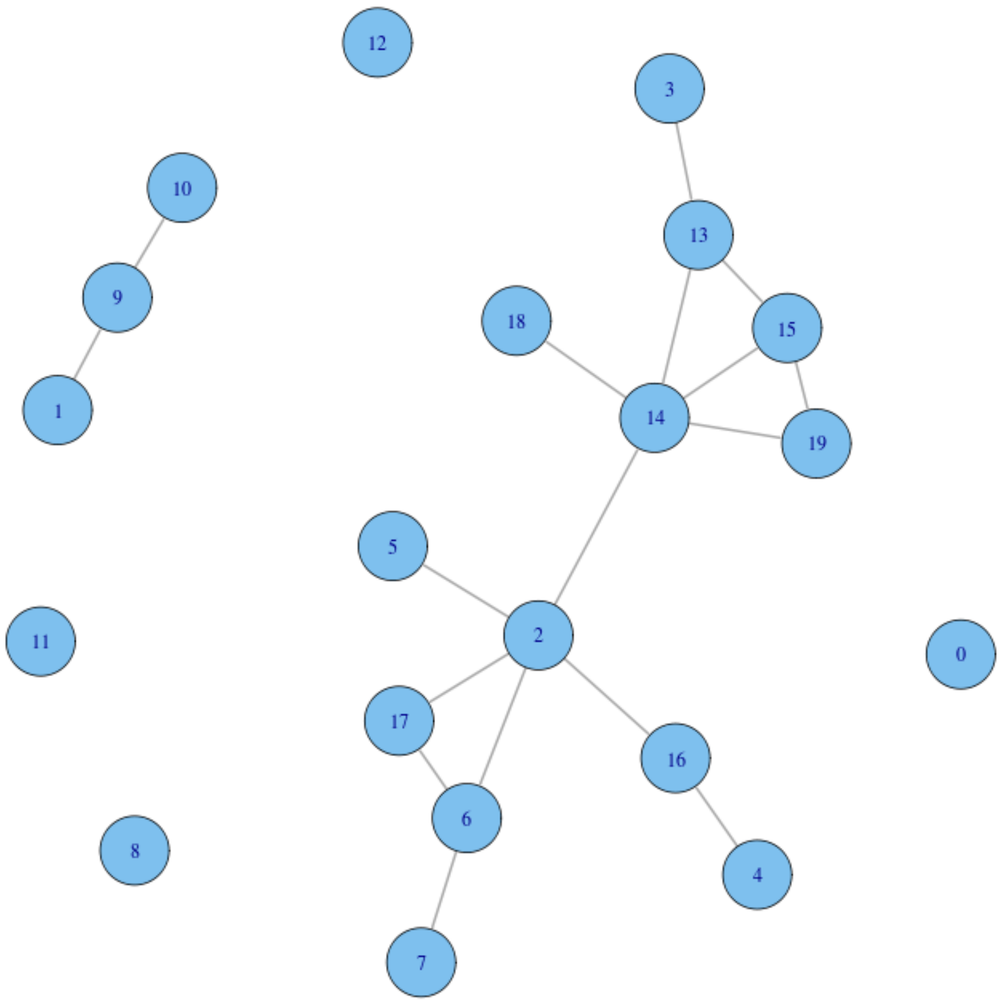
Il fondatore della **sociometria**, cioè della disciplina che si occupa dello studio delle reti sociali, è in realtà lo psichiatra **Jacob Moreno**, un rumeno immigrato negli Stati Uniti che negli anni '30 si appassiona alle dinamiche delle interazioni sociali tra gruppi di persone. Nel 1934 Moreno pubblica un libro dal titolo emblematico *Who Shall Survive?* che getta le fondamenta della sociometria.

Moreno chiama le reti sociali con il nome **sociogramma**. Il primo esempio di sociogramma della storia è una immagine disegnata a mano delle relazioni di amicizia tra bambini e bambine in una scuola elementare. La figura rivela chiaramente come maschi e femmine formino gruppi omogenei con pochissime relazioni miste di amicizia tra i due sessi, e inoltre come vi siano individui isolati dagli altri. Lo studio si merita una colonna del *New York Times* e rapidamente gli altri sociologi danno credito ai sociogrammi inventati da Moreno e ne fanno uso.

Una brevissima introduzione alle teoria dei grafi

In matematica una rete sociale viene rappresentata con una struttura chiamata grafo. Un **grafo** è fatto di **nodi** (o **vertici**), che rappresentano gli attori della rete sociale, e collegamenti tra coppie di nodi chiamati **archi**, che rappresentano i legami tra gli attori della rete.

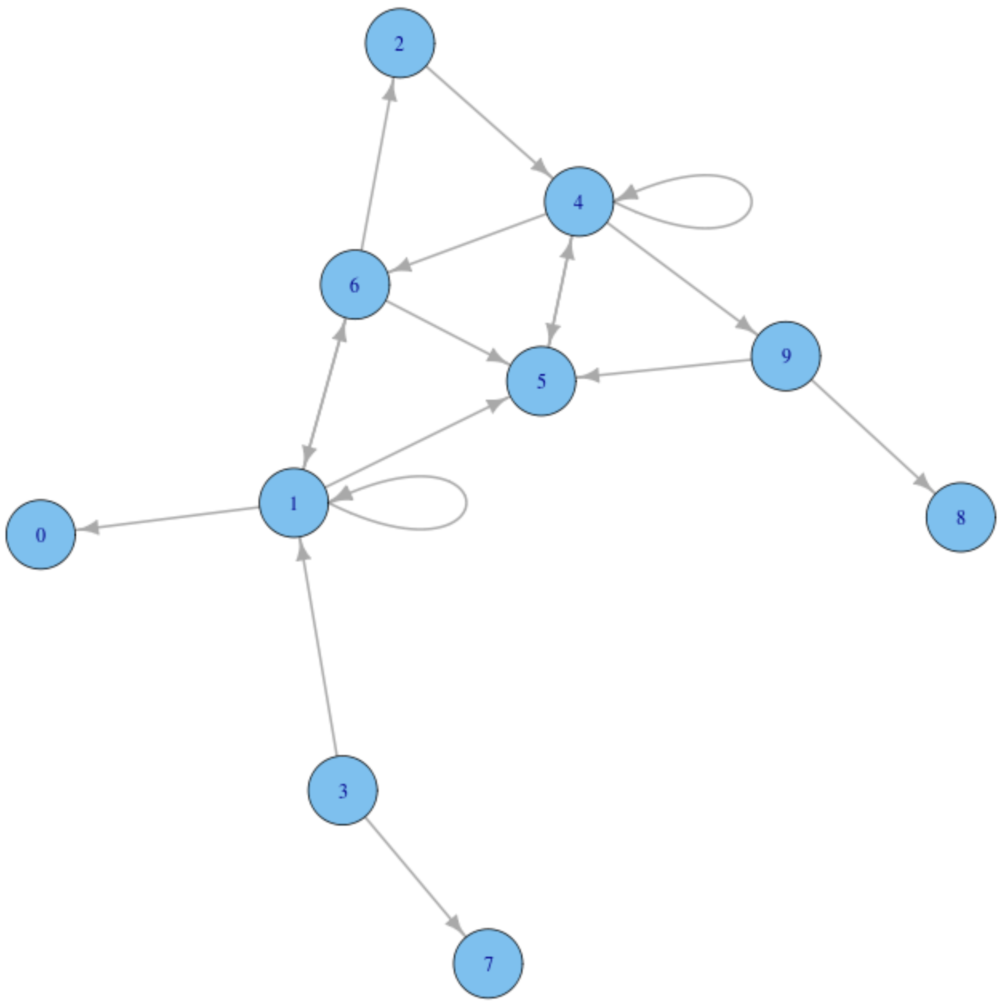
In un **grafo indiretto** gli archi non hanno una direzione. Questo è un semplice esempio di grafo indiretto.



In un grafo indiretto la relazione degli archi è **simmetrica**: se vi è un collegamento tra il nodo A e il nodo B allora vi è anche un collegamento tra B e A. Ad esempio, la relazione di amicizia di Facebook si può rappresentare con un grafo indiretto.

Dato un nodo A di un grafo indiretto, i nodi collegati ad A mediante un arco sono detti **nodi adiacenti** o **vicini** di A. Nella rete sociale di Facebook i nodi adiacenti di un utente sono i suoi amici.

In un **grafo diretto** gli archi hanno una direzione. Questo è un semplice esempio di grafo diretto.



In un grafo diretto la relazione degli archi è **direzionale**: se vi è un collegamento diretto tra il nodo A e il nodo B non è detto che vi sia anche un collegamento tra B e

A. Ad esempio, la relazione segue di Twitter si può rappresentare con un grafo diretto.

Se il grafo è diretto, distinguiamo tra **nodi successori** di un nodo A, cioè tutti i nodi B per cui esiste un arco diretto da A a B, e **nodi predecessori** di un nodo A, cioè tutti i nodi B per cui esiste un arco diretto da B ad A. Nella rete sociale di Twitter i predecessori di un utente sono i suoi seguitori (coloro che seguono l'utente), mentre i successori di un utente sono i suoi seguiti (coloro che sono seguiti dall'utente).

Un **cammino** tra due nodi A e B è una sequenza di nodi che collega il nodo A al nodo B attraverso gli archi del grafo. Notate che in un grafo indiretto se esiste un cammino da A a B allora esiste anche il cammino inverso da B a A (questo perché la relazione arco è simmetrica). Non è così per grafi diretti. Ad esempio nella rete sociale di Facebook un esempio di cammino è Massimo Franceschet (amico di) Luca Chittaro (amico di) Guido Romeo (amico di) Alessandra Viola.

La **lunghezza di un cammino** è il numero di archi del cammino. Un **cammino di lunghezza minima** (o **geodetica**) tra due nodi è un cammino con la minima lunghezza tra tutti i cammini tra i due nodi. Notate che vi possono essere più cammini di lunghezza minima tra due nodi.

Misure di centralità

Un volume consistente di ricerca in sociometria è dedicato al concetto di **centralità** nelle reti sociali. Questa ricerca affronta la questione:

Quali sono gli attori più centrali (o importanti) della rete?

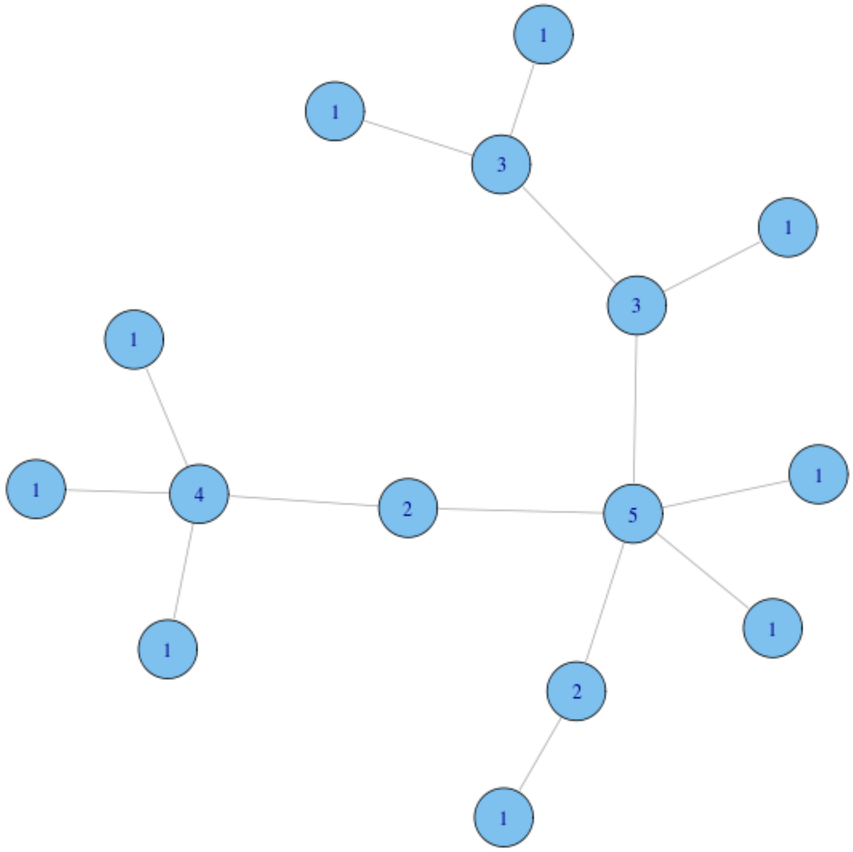
Gli individui più centrali corrispondono generalmente agli individui socialmente più **influenti** e **carismatici**, senza i quali la rete perderebbe una parte significativa del suo valore sociale.

Vi sono varie definizioni di importanza e quindi varie misure di centralità. Ciò che le accomuna tutte è che esse fanno riferimento unicamente alla **struttura** (o topologia) della rete. Per esempio, in una rete sociale di amicizie come Facebook, una misura di centralità non potrebbe essere quanto una persona è ricca o quanto è intelligente, perché queste informazioni non fanno parte della rete sociale. Al contrario, il numero di amici di una persona corrisponde ad una misura di centralità; notate che questa è una informazione che si può dedurre dalla struttura della rete (è il numero di nodi adiacenti ad un nodo).

Centralità per grado

Probabilmente la più semplice e più usata misura di centralità è la **centralità per grado** (*degree centrality*). Questa corrisponde al **grado di un nodo**, cioè al numero di nodi adiacenti a quel nodo. Su Facebook, ad esempio, la centralità per grado è il numero di amici di una persona.

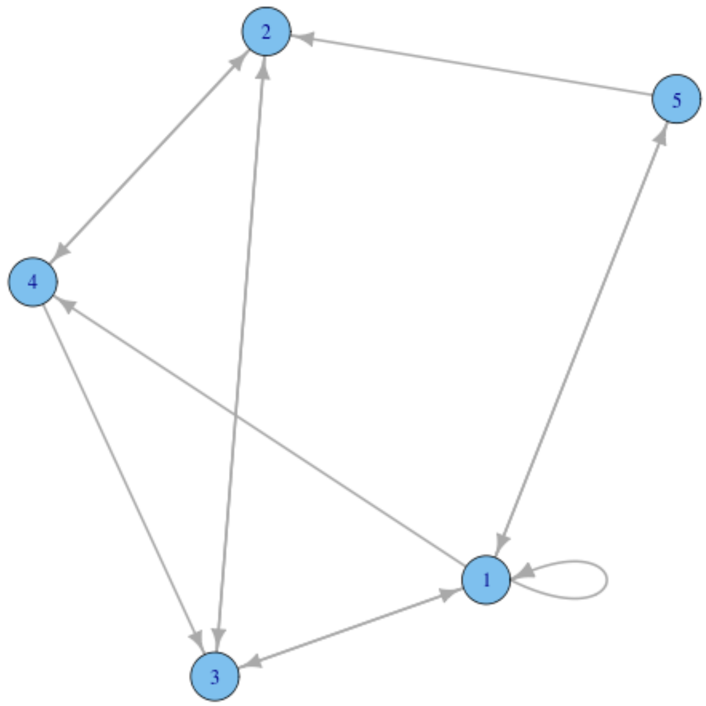
Per esempio, in questo grafo indiretto i nodi sono stati etichettati con il loro grado:



Se la rete sociale è diretta, distinguiamo tra **centralità per grado uscente**, che corrisponde al numero di successori di un nodo, e **centralità per grado entrante**, che corrisponde al numero di predecessori di un nodo. Ad esempio su Twitter la centralità per grado entrante è il numero di seguitori di un utente, mentre quella per grado uscente è il numero di utenti seguiti da una persona. La più importante è la prima, in quanto un utente può seguire chi vuole a sua discrezione, mentre sono gli altri che decidono se seguirlo o meno (a seconda dell'interessa che suscita con i suoi brevi messaggi).

In matematica un grafo è tipicamente rappresentato da una **matrice di adiacenza** $A = (a_{i,j})$ dove i nodi sono numerati con i numeri da 1 in poi e la componente $a_{i,j}$ in posizione (i, j) della matrice vale 1 se esiste un arco dal nodo i al nodo j e vale 0 altrimenti.

Per esempio, il seguente grafo diretto con 5 nodi e 12 archi viene rappresentato dalla seguente matrice 5×5:



$$\begin{pmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Il numero di componenti a 1 è il numero di archi (12), perché ogni argo è rappresentato una volta nella matrice. Notate che il grado uscente out_i del nodo i corrisponde alla somma delle componenti sulla riga i:

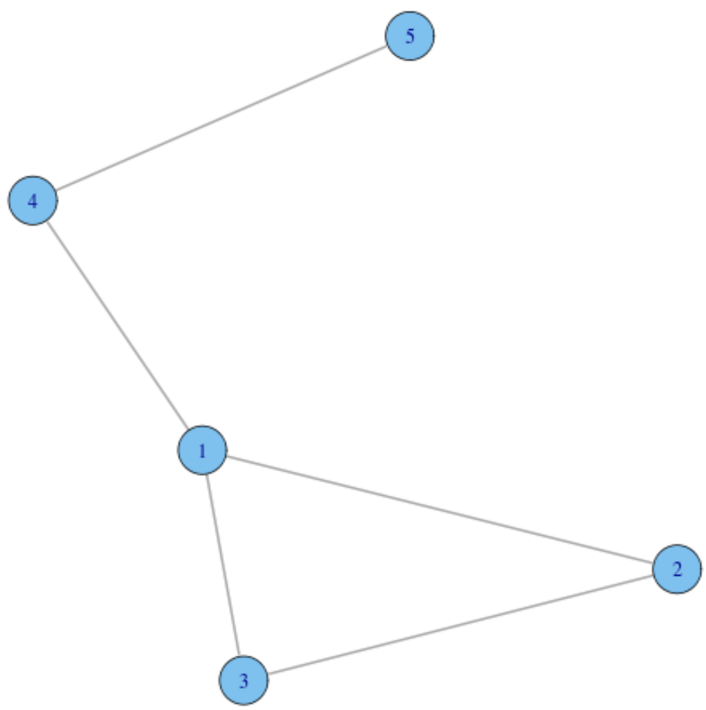
$$out_i = \sum_j a_{i,j}$$

Il grado entrante in_i del nodo i corrisponde alla somma delle componenti sulla colonna i:

$$in_i = \sum_j a_{j,i}$$

Le componenti a 1 sulla diagonale principale corrispondono ai cappi. La matrice non è simmetrica rispetto alla diagonale principale.

Il seguente grafo indiretto con 5 nodi e 5 archi viene rappresentato dalla seguente matrice 5×5:



$$\begin{pmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Il numero di componenti a 1 è il doppio del numero di archi (10), perché ogni argo è rappresentato due volte nella matrice. La matrice è simmetrica rispetto alla diagonale principale perché la relazione arco è simmetrica. Notate che il grado nodo i corrisponde alla somma delle componenti sulla riga (o colonna) i:

$$deg_i = \sum_j a_{i,j} = \sum_j a_{j,i}$$

Centralità per autovettore

Il principale svantaggio delle centralità per grado è che assume che tutti gli attori della rete abbiano lo stesso peso. E' invece evidente che, per esempio, avere amici importanti fa differenza rispetto ad avere amici di poco peso.

La **centralità per autovettore** (*eigenvector centrality*) assegna ad ogni attore una importanza che è **proporzionale all'importanza degli attori adiacenti**. Su Facebook, un utente è importante secondo questa definizione se ha amici a loro volta importanti, non necessariamente molti.

Per ogni nodo j, la centralità per autovettore x_j del nodo j è definita come segue:

$$x_j = \frac{1}{\lambda} \sum_i a_{i,j} x_i$$

dove $a_{i,j}$ è l'elemento (i, j) della matrice di adiacenza del grafo e $\lambda \neq 0$ è una costante. Dunque, come promesso, la centralità x_j del nodo j è proporzionale alla sommatoria delle centralità dei nodi adiacenti a j (o predecessori di j per grafi diretti).

Si badi bene che la definizione è **ricorsiva** in quanto l'importanza di un attore X è definita in termini dell'importanza degli attori adiacenti Y; ma quest'ultima può essere direttamente o indirettamente definita in termini dell'importanza di X. Questo accade sempre, per esempio, su grafi indiretti, in quanto se Y è adiacente ad X allora X è adiacente a Y, o su grafi diretti con dei cicli (cammini che partono e tornano allo stesso nodo).

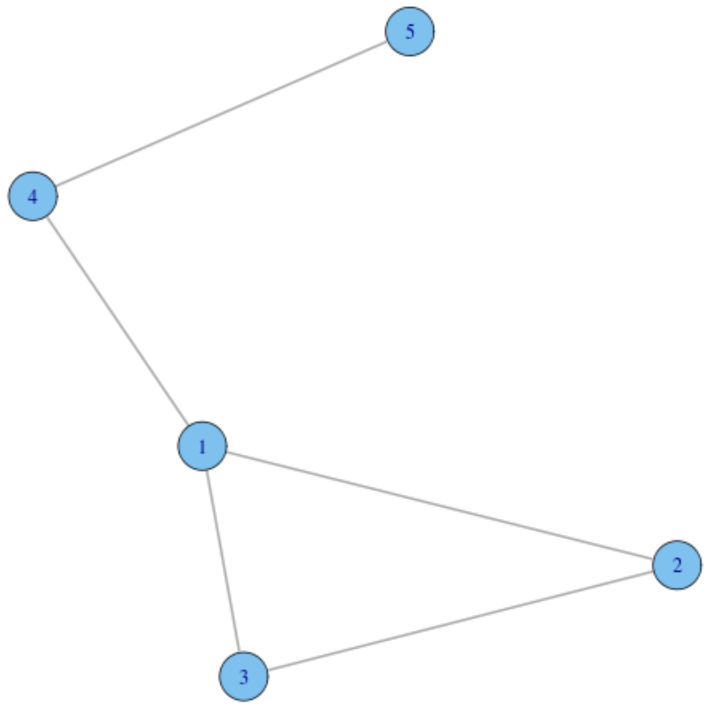
In **notazione matriciale**, l'insieme delle equazioni che definiscono la centralità di un nodo risulta essere definita da questa equazione singola:

$$\lambda x = xA$$

dove λ è una costante, x è un vettore le cui componenti contengono la centralità degli attori, e A è la matrice di adiacenza del grafo. Il problema consiste nel trovare valori per $\lambda \neq 0$ e $x \neq 0$ che soddisfano l'equazione. In algebra, il vettore x viene detto **autovettore** della matrice A associato all'**autovalore** λ. Un autovettore è un vettore che non viene modificato dalla **trasformazione lineare** A, a meno della moltiplicazione per la costante λ che lo fa semplicemente scalare (cambiare di dimensione) ma non ne cambia la direzione.

Vediamo un esempio. Consideriamo ancora il seguente grafo indiretto con 5 nodi e

5 archi che viene rappresentato dalla seguente matrice 5×5:

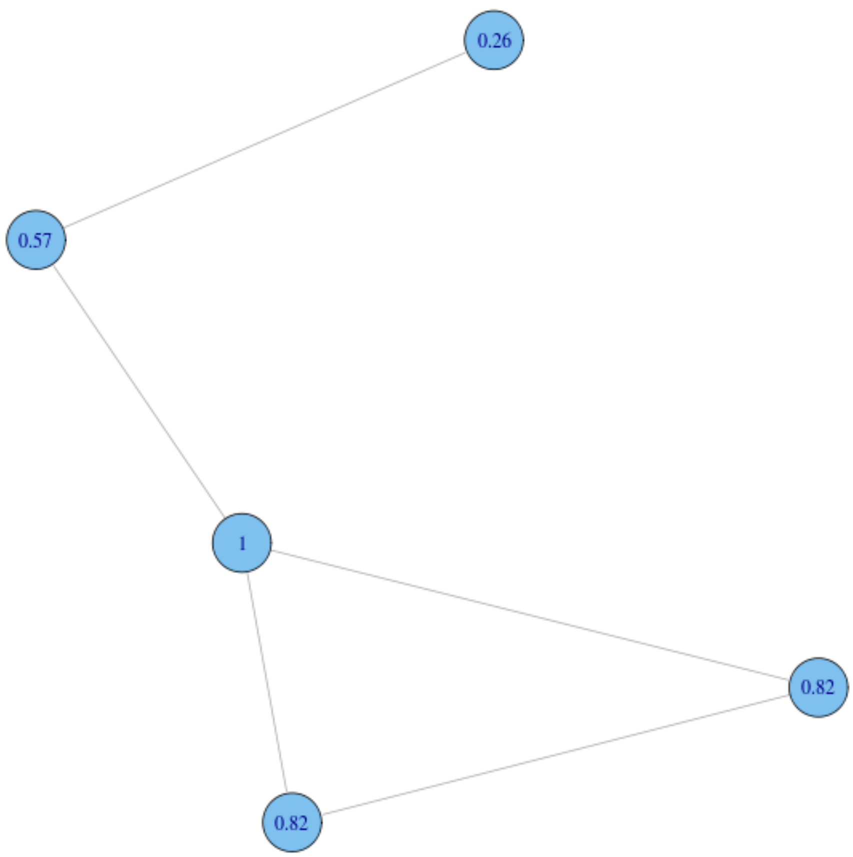


$$\begin{pmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

L'insieme di equazioni per la centralità risulta:

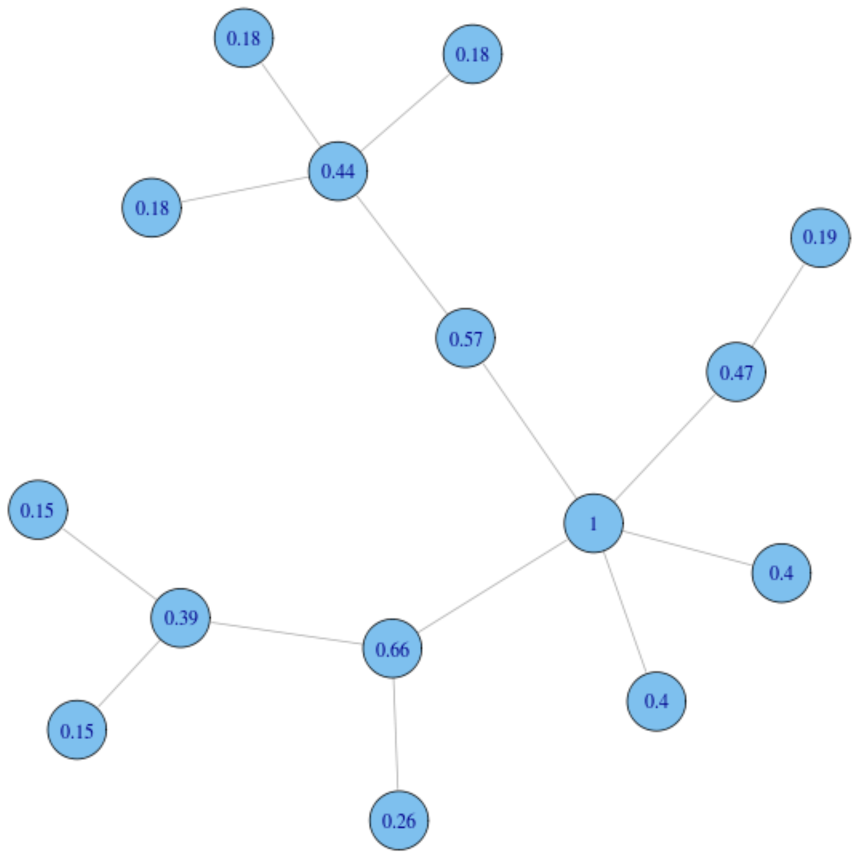
$$
\begin{aligned}
\lambda x_1 &= x_2 + x_3 + x_4 \\
\lambda x_2 &= x_1 + x_3 \\
\lambda x_3 &= x_1 + x_2 \\
\lambda x_4 &= x_1 + x_5 \\
\lambda x_5 &= x_4
\end{aligned}
$$

Notate che questo **non è un sistema lineare** perché l'autovalore λ non è noto, è una variabile, ed esso viene moltiplicato per gli elementi del vettore incognita x. La soluzione risulta la seguente in cui i nodi sono etichettati con la corrispondente centralità (i valori sono normalizzati nell'intervallo [0,1] e arrotondati alla seconda cifra decimale):



Il valore (arrotondato) dell'autovettore $\lambda = 2.21$. Si può verificare che con la soluzione trovata vale $\lambda x = x A$.

Vediamo un esempio più complicato e direttamente la sua soluzione nel grafo:

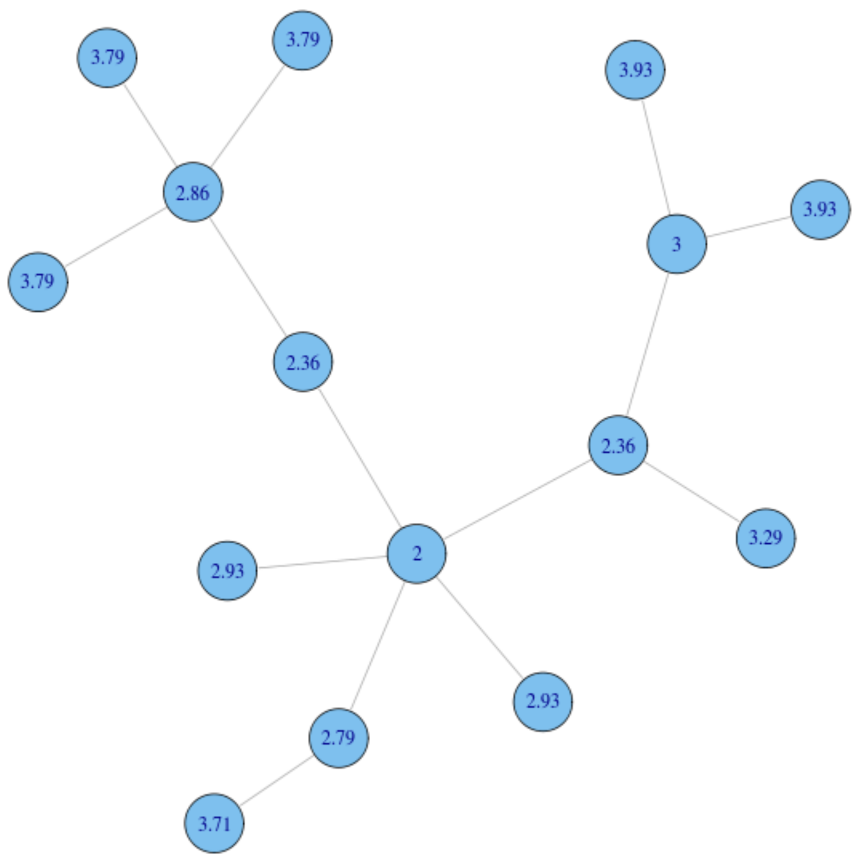


Ma come risolvere l'equazione matriciale $\lambda x = x A$? La soluzione si può trovare con un metodo iterativo chiamato **Metodo delle Potenze**. Il metodo è molto semplice e consiste nel partire da un vettore x_0 arbitrario e moltiplicare $x_0 A$ ottenendo un nuovo vettore x_1, quindi dividere x_1 per la sua massima componente e ripetere quanto fatto con il nuovo vettore x_1, cioè moltiplicare $x_1 A$ ottenendo un nuovo vettore x_2 da dividere per la sua massima componente, e così via. Accade che per k sufficientemente grande x_k diventa l'autovettore x soluzione della nostra equazione e la sua massima componente diventa l'autovalore λ.

Centralità basate sui cammini minimi

Vi sono altri due metodi importanti per determinare la centralità di un nodo che si basano sul concetto di cammino minimo.

Definiamo **distanza** tra due nodi come la lunghezza di un cammino minimo tra i due nodi (oppure il numero di nodi del grafo se non esiste alcun cammino tra i due nodi). La **centralità per vicinanza** (*closeness centrality*) misura la distanza media di un nodo da tutti gli altri nodi del grafo. I nodi poco distanti da altri nodi in questo senso hanno il vantaggio di ricevere velocemente le informazioni che circolano sulla rete e allo stesso tempo di far arrivare velocemente le proprie informazioni agli altri attori della rete. Vediamo un esempio con soluzione:



La **centralità per intermediazione** (*betweenness centrality*) misura il numero di cammini minimi nel grafo che contengono quel nodo. Un nodo che giace su molti cammini minimi tra altri nodi è centrale nel senso che esercita un controllo sullo scambio di informazione tra i nodi, cioè è un buon intermediatore (*broker*) tra gli altri nodi. Vediamo un esempio con soluzione: