Implementazione di algoritmi per il calcolo efficiente del ranking dato dalla centralità di Katz in grafi molto densi

Gabriel Antonio Videtta (654839)

5 maggio 2025

Sommario

In questa relazione presentiamo la teoria e gli algoritmi proposti da [3] per calcolare efficientemente il ranking dovuto all'indice di centralità di Katz in grafi, eventualmente con pesi non negativi, le cui matrici di adiacenza hanno almeno $\mathcal{O}(n)$ elementi non nulli, sfruttando la nozione di "grafo complementare" e dando significato al calcolo di indici di centralità di Katz con parametri negativi.

1 Prerequisiti teorici

Ricordiamo brevemente che un grafo (eventualmente *con lacci*) è una coppia di insiemi $G = (V, E)$ con $E \subseteq V \times V$ e $V = \{1, \ldots, n\}$ per qualche $n \in \mathbb{N}_{>0}$. Gli elementi dell'insieme V sono detti *nodi*, mentre quelli dell'insieme E sono detti *archi*.

Quando si parla di grafi *senza lacci* (o di grafo *semplice*), si esclude a priori l'esistenza di archi della forma (i, i) con $i \in V$.

Nel corso della relazione useremo n per riferirci al numero di nodi del grafo preso in considerazione, e useremo m per riferirci al numero di archi dello stesso.

L'insieme E induce una relazione $\sim$ su V tale per cui $i \sim j$ se e solo se $(i, j) \in E$. Il grafo G si dice *non orientato* se $\sim$ è simmetrica, e *orientato* altrimenti.

Un sottografo di G è un grafo $G' = (V', E')$ con $V' \subseteq V$ e $E' \subseteq (V' \times V') \cap E$.

Un grafo è rappresentato operativamente tramite la propria *matrice di adiacenza* $A \in \mathbb{R}^{n \times n}$, definita componente per componente come

$$A_{ij} = \begin{cases} 1 & \text{se } i \sim j, \\ 0 & \text{altrimenti.} \end{cases}$$

Si osserva immediatamente che A è simmetrica se e solo se G è non orientato.

Un grafo è detto *sparso* se A ha $\mathcal{O}(n)$ entrate non nulle, mentre è detto *denso* se tutte le entrate di A eccetto per un numero $\mathcal{O}(n)$ di queste sono non nulle. Equivalentemente, un grafo sparso è tale per cui $m = \mathcal{O}(n)$.

Un *cammino orientato* (*directed walk* in inglese) di lunghezza r dal nodo i al nodo j è una sequenza ordinata di $r + 1$ nodi $i_0 = i$, i_1, ..., $i_r = j$ dove $i_k \sim i_{k+1}$ per ogni $k = 0, 1, \ldots, r - 1$.

Un *cammino non orientato* (*undirected walk* in inglese) ignora le direzioni, ovverosia permette di passare da i a j anche se $j \sim i$.

Nel caso di un grafo non orientato, si identificano cammini orientati e non orientati, chiamandoli entrambi semplicemente *cammini*.

Un cammino (orientato o non orientato) è detto *elementare* (*path* in inglese) se ogni nodo è toccato al più una volta.

Un grafo è detto *connesso* se dati due nodi esiste sempre un cammino non orientato che li collega, mentre è detto *fortemente connesso* se dati due nodi esiste sempre un cammino orientato che li collega. Un grafo non orientato è connesso se e solo se è fortemente connesso, dal momento che cammini orientati e non orientati sono identificati. Un grafo è fortemente connesso se e solo se la sua matrice di adiacenza è irriducibile per permutazioni (vd. [1, Theorem 3.2.1]).

Si dice *componente connessa* del nodo i il più grande sottografo connesso di G contenente i. Analogamente si definisce una *componente fortemente connessa*.

Nel corso di questa relazione, scriveremo $\mathbf{1} \in \mathbb{R}^n$ per riferirci al vettore composto da soli uno, $\mathbf{0} \in \mathbb{R}^n$ per riferirci al vettore nullo, $I \in \mathbb{R}^{n\times n}$ per riferirci alla matrice identità e $\mathbf{e}_k \in \mathbb{R}^n$ per riferirci alla k-esima colonna di I. Scriveremo $\mathbf{a}_k$ per riferirci alla k-esima colonna di una matrice A e a_k per riferirci al k-esimo elemento di un vettore $\mathbf{a}$.

1.1 Il vettore di Katz e buona definizione

Ricordiamo un altro classico risultato della teoria dei grafi, facilmente dimostrabile per induzione.

Lemma 1. *Sia A la matrice di adiacenza di un grafo G e sia $r \in \mathbb{N}$. Allora l'entrata (i, j)-esima di A^r rappresenta il numero di cammini di lunghezza r da i a j.*

Osservazione 1. *Dal Lemma 1 segue facilmente che l'i-esima coordinata del vettore $A^r\mathbf{1}$ rappresenta il numero di cammini di lunghezza r che partono da i e che terminano in un qualsiasi nodo.*

Definizione 1. *Se $\alpha \in (-1/\rho(A), 1/\rho(A))$*[1]*,* il vettore di Katz in α di A *è definito come il vettore*

$$\mathbf{x} := (I + \alpha A + \alpha^2 A^2 + \cdots)\mathbf{1} = (I - \alpha A)^{-1}\mathbf{1}.$$

Equivalentemente, $\mathbf{x}$ è l'unica soluzione del sistema lineare $(I - \alpha A)\mathbf{x} = \mathbf{1}$.

Osservazione 2. *Se $\alpha > 0$, il vettore di Katz induce un naturale ranking dei nodi di G, dove l'importanza di un nodo i è determinata dai cammini che hanno sorgente i: a un cammino di lunghezza r è associato un valore α^r, e sommando tutti questi valori si ottiene la coordinata di i nel vettore di Katz.*

Nella teoria si vorrebbe che α^r decresca all'aumentare di r, ovverosia che i cammini più lunghi abbiano sempre meno importanza. Questo non è immediatamente ovvio, dal momento che $1/\rho(A)$ potrebbe essere maggiore di 1. La Proposizione 1 *mostra che $\rho(A) > 1$ nella maggior parte dei casi considerati, garantendoci di star dando la giusta interpretazione alla costruzione del vettore di Katz.*

[1] Il limite superiore $1/\rho(A)$ è necessario affinché la serie converga a $(I - \alpha A)^{-1}$.

Definizione 2. *Se G è un grafo non orientato, si definisce* grado $\deg(i)$ *del nodo i il numero di archi insistenti su i. Equivalentemente* $\deg(i) = \mathbf{1}^T A \mathbf{e}_i$.

Lemma 2 (Handshaking lemma)**.** *Se G è un grafo semplice non orientato, allora*

$$\sum_{i \in V} \deg(i) = 2m.$$

Dimostrazione. Si tratta di una semplice verifica combinatoriale:

$$\begin{aligned} \sum_{i \in V} \deg(i) &= \sum_{i \in V} \mathbf{1}^T A \mathbf{e}_i \\ &= \mathbf{1}^T A \mathbf{1} \\ &= 2m, \end{aligned}$$

dove si è usato che non esistono lacci in G e che gli elementi non nulli di A sono tanti quanto il doppio degli archi in G. $\square$

Definizione 3. *Se G è un grafo non orientato, si definisce* il grado medio d_G *di G come la media dei gradi dei nodi, ovverosia*

$$d_G = \frac{\sum_{i \in G} \deg(i)}{n}.$$

Proposizione 1. *Se un grafo semplice non orientato G ha almeno un nodo con grado almeno 2 e $A \in \mathbb{R}^{n \times n}$ è la sua matrice di adiacenza, allora $\rho(A) > 1$.*

Dimostrazione. Sia C una componente connessa di G a cui appartiene almeno un nodo di grado almeno 2, e indichiamo con $V(C)$ l'insieme dei nodi di C, con n_C il numero di nodi $|V(C)|$ e con m_C il numero di archi di C. Sia $\mathbf{e}_C$ il vettore indicatore dei membri di C, ovvero tale che

$$(e_C)_i = \begin{cases} 1 & \text{se } i \in V(C), \\ 0 & \text{altrimenti.} \end{cases}$$

Allora $\mathbf{e}_C^T \mathbf{e} = n_C$, il numero di nodi nella componente C, mentre vale che

$$\mathbf{e}_C^T A \mathbf{e}_C = \sum_{i \in C} \deg(C) = 2 \cdot m_C$$

dove m_C è il numero degli archi in C e l'ultima uguaglianza è data dal Lemma 2. Dal momento che C è connesso, allora C deve contenere almeno $n_C - 1$ archi, ovverosia $m_C \geq n_C - 1$. Se d_C è il grado medio di C, allora:

$$d_C = \frac{\mathbf{e}_C^T A \mathbf{e}_C}{\mathbf{e}_C^T \mathbf{e}_C} \geq \frac{2(n_C - 1)}{n_C} \geq \frac{4}{3},$$

dove nell'ultima disuguaglianza si è sfruttato che $n_C \geq 3$, dal momento che esiste almeno un nodo in C di grado 2.

Poiché A è una matrice simmetrica reale, $\rho(A)$ maggiora sicuramente d_C, che è il quoziente di Rayleigh per la matrice A e il vettore e_C (vd. [2, Theorem 4.4.2]), e pertanto $\rho(A) \geq \frac{4}{3} > 1$. $\square$

2 Grafo complementare e vettore di Katz

In questa sezione della relazione, presentiamo il concetto di grafo complementare, eventualmente senza lacci, e tramite i Teoremi 1 e 2, preceduti dai preziosi Lemmi 3 e 4, troviamo una corrispondenza che ci permette di calcolare il vettore di Katz sul grafo complementare, ottenendo lo stesso ranking che avremmo ottenuto calcolandolo sul grafo originale.

Definizione 4. *Su un grafo (con eventualmente lacci) $G = (V, E)$ si definisce il* **grafo complementare** *$G^c = (V, E^c)$ come il grafo tale per cui E^c è il complementare di E in $V \times V$.*

Su un grafo senza lacci $G = (V, E)$ si definisce il **grafo complementare senza lacci** *come il grafo complementare di G a cui si tolgono i lacci.*

Osservazione 3. *La matrice di adiacenza di un grafo complementare si ottiene come una modifica di rango 1 della matrice di adiacenza di partenza (cambiata di segno), ovverosia, se A^C è la matrice di adiacenza del complementare e A è la matrice di adiacenza originale vale*

$$A^C = \mathbf{e}\mathbf{e}^T - A.$$

La matrice di adiacenza di un grafo complementare senza lacci si ottiene invece come una modifica di rango 1 su $A + I$ (cambiata di segno):

$$A^C = \mathbf{e}\mathbf{e}^T - A - I$$

Lemma 3. *Sia $A \in \mathbb{C}^{n\times n}$ una matrice invertibile e siano $\mathbf{u}$ e $\mathbf{v}$ vettori in $\mathbb{C}^n$. Allora $\det(A + \mathbf{u}\mathbf{v}^T) = \det(A)(1 + \mathbf{v}^T A^{-1}\mathbf{u})$.*

Dimostrazione. Dimostriamo innanzitutto il lemma nel caso in cui $A = I$.

Poiché $\mathbf{u}\mathbf{v}^T$ è una matrice di rango 1, $\mathbf{u}\mathbf{v}^T$ ha come autovalore 0 con almeno molteplicità $n - 1$. L'unico altro autovalore, che è eventualmente 0, è allora $\mathrm{tr}(\mathbf{u}\mathbf{v}^T)$, ovverosia $\mathbf{v}^T\mathbf{u}$. Pertanto il polinomio caratteristico di $\mathbf{u}\mathbf{v}^T$ risulta essere

$$p_{\mathbf{u}\mathbf{v}^T}(t) = \det(\mathbf{u}\mathbf{v}^T - tI) = (-1)^n t^{n-1}(t - \mathbf{v}^T\mathbf{u}).$$

Dunque, $\det(I + \mathbf{u}\mathbf{v}^T) = \det(\mathbf{u}\mathbf{v}^T - (-1)\cdot I) = 1 + \mathbf{v}^T\mathbf{u}$, dimostrando il caso in cui $A = I$.

Nel caso generale, considerato che $A + \mathbf{u}\mathbf{v}^T = A(I + A^{-1}\mathbf{u}\mathbf{v}^T)$, applicando l'identità di Binet si ottiene

$$\det(A + \mathbf{u}\mathbf{v}^T) = \det(A)\det(I + A^{-1}\mathbf{u}\mathbf{v}^T) = \det(A)(1 + \mathbf{v}^T A^{-1}\mathbf{u}),$$

completando la dimostrazione. □

Lemma 4 (Sherman-Morrison)**.** *Se $A \in \mathbb{C}^{n\times n}$ è invertibile e $\mathbf{u}$ e $\mathbf{v}$ sono vettori in $\mathbb{C}^n$, allora $A + \mathbf{u}\mathbf{v}^T$ è invertibile se e solo se $1 + \mathbf{v}^T A^{-1}\mathbf{u}$ è diverso da 0, e in tal caso vale che*

$$(A + \mathbf{u}\mathbf{v}^T)^{-1} = A^{-1} - \frac{A^{-1}\mathbf{u}\mathbf{v}^T A^{-1}}{1 + \mathbf{v}^T A^{-1}\mathbf{u}}.$$

Dimostrazione. La prima parte del lemma è un corollario del Lemma 3. La seconda parte è una semplice verifica diretta:

$$
\begin{aligned}
(A+\mathbf{u}\mathbf{v}^T)\left(A^{-1}-\frac{A^{-1}\mathbf{u}\mathbf{v}^T A^{-1}}{1+\mathbf{v}^T A^{-1}\mathbf{u}}\right) &= I+\mathbf{u}\mathbf{v}^T A^{-1}-\frac{\mathbf{u}\mathbf{v}^T A^{-1}+\mathbf{u}\mathbf{v}^T A^{-1}\mathbf{u}\mathbf{v}^T A^{-1}}{1+\mathbf{v}^T A^{-1}\mathbf{u}} \\
&= I+\mathbf{u}\mathbf{v}^T A^{-1}-\frac{(1+\mathbf{v}^T A^{-1}\mathbf{u})\mathbf{u}\mathbf{v}^T A^{-1}}{1+\mathbf{v}^T A^{-1}\mathbf{u}} \\
&= I+\mathbf{u}\mathbf{v}^T A^{-1}-\mathbf{u}\mathbf{v}^T A^{-1} \\
&= I.
\end{aligned}
$$

□

Teorema 1. *Sia G un grafo. Se $\alpha \in (0, 1/\rho(A))$, allora il vettore di Katz di parametro $-\alpha$ calcolato sul complementare (con lacci) G^C induce lo stesso ranking del vettore di Katz di parametro α calcolato su G.*

Dimostrazione. Innanzitutto, osserviamo che

$$I+\alpha A^C = (I-\alpha A)+\alpha\mathbf{e}\mathbf{e}^T.$$

Pertanto, per il Lemma 3, $I+\alpha A^C$ è invertibile se e solo se

$$1+\alpha\mathbf{e}^T(I-\alpha A)^{-1}\mathbf{e} \neq 0.$$

Dal momento che $\alpha > 0$, il termine $\alpha\mathbf{e}^T(I-\alpha A)^{-1}\mathbf{e}$ è certamente positivo, e dunque $I+\alpha A^C$ è invertibile.

Sia A la matrice di adiacenza di G. Allora la matrice di adiacenza A^C di G^C è tale per cui $A^C = \mathbf{e}\mathbf{e}^T - A$. Si osserva che, per il Lemma 4, vale che

$$
\begin{aligned}
(I-\alpha A)^{-1} &= (I-\alpha(\mathbf{e}\mathbf{e}^T - A^C))^{-1} \\
&= (I+\alpha A^C-\alpha\mathbf{e}\mathbf{e}^T)^{-1} \\
&= (I+\alpha A^C)^{-1}-\frac{(I+\alpha A^C)^{-1}(-\alpha\mathbf{e}\mathbf{e}^T)(I+\alpha A^C)^{-1}}{1+\mathbf{e}^T(I+\alpha A^C)^{-1}(-\alpha\mathbf{e})} \\
&= (I+\alpha A^C)^{-1}+\alpha\frac{(I+\alpha A^C)^{-1}\mathbf{e}\mathbf{e}^T(I+\alpha A^C)^{-1}}{1-\alpha\mathbf{e}^T(I+\alpha A^C)^{-1}\mathbf{e}}. \qquad (1)
\end{aligned}
$$

Sia γ definito come

$$\gamma = \mathbf{e}^T(I+\alpha A^C)^{-1}\mathbf{e}.$$

Allora, grazie all'eq. (1), il vettore di Katz di A si riscrive come

$$(I-\alpha A)^{-1}\mathbf{e} = \frac{1}{1-\alpha\gamma}(I+\alpha A^C)^{-1}\mathbf{e}. \qquad (2)$$

Dall'eq. (2) si ricava inoltre che

$$(I+\alpha A^C)(I-\alpha A)^{-1}\mathbf{e} = \frac{1}{1-\alpha\gamma}\mathbf{e}. \qquad (3)$$

Poiché il termine $(I+\alpha A^C)(I-\alpha A)^{-1}\mathbf{e}$ dell'eq. (3) è non negativo, allora anche il termine a destra è non negativo, da cui $1-\alpha\gamma > 0$, e dunque $\alpha\gamma < 1$.

Pertanto $(I+\alpha A^C)^{-1}\mathbf{e}$, che è proporzionale a $(I-\alpha A)^{-1}\mathbf{e}$ per un fattore positivo per l'eq. (2), è un vettore non negativo ed induce correttamente un ranking dei nodi equivalente a quello generato dal vettore di Katz di A. □

Teorema 2. *Sia G un grafo. Se $\alpha \in (0, 1/\rho(A))$, allora il vettore di Katz di parametro $-\frac{\alpha}{1+\alpha}$ calcolato sul complementare senza lacci G^C induce lo stesso ranking del vettore di Katz di parametro α calcolato su G.*

Dimostrazione. La dimostrazione segue pressoché gli stessi passaggi della dimostrazione del Teorema 1 definendo $A^C = \mathbf{e}\mathbf{e}^T - (A + I)$ al posto di $A^C = \mathbf{e}\mathbf{e}^T - A$. $\square$

3 Sperimentazione numerica

Le simulazioni sono state eseguite in ambiente GNU Octave 9.4.0 su un PC laptop equipaggiato con 16 GB di RAM e processore Intel i7-13620H con un clock rate di 2.40 GHz. Tutti i codici sono inclusi inoltre in [4].

Prima di implementare operativamente il calcolo del vettore di Katz mediante il grafo complementare, si è innanzitutto generato due grafi densi con `generate_matrices.m`:

```
n = 4000;
density = 0.97;

% si assicura la densita' di A e B
Am = double(rand(n) < density);
Bm = double(rand(n) < density);

% forza la simmetria ed elimina i loop su A
Am = triu(Am, 1);
Am = Am + Am';

% forza la simmetria, ma non elimina i loop su B
Dm = diag(diag(Bm));
Bm = triu(Bm, 1);
Bm = Bm + Bm' + Dm;

A = struct();
B = struct();

A.matrix = Am;
B.matrix = Am;

A.rho = svds(Am, 1);
B.rho = svds(Bm, 1);

save(''A.mat'', ''A'')
save(''B.mat'', ''B'')
```

Figura 1: `generate_matrices.m`

Riferimenti bibliografici

[1] Richard A Brualdi, Herbert John Ryser et al. *Combinatorial matrix theory*. Vol. 39. Springer, 1991.

[2] Shmuel Friedland. *Matrices: algebra, analysis and applications*. World Scientific, 2015.

[3] Vanni Noferini e Ryan Wood. «Efficient computation of Katz centrality for very dense networks via negative parameter Katz». In: *Journal of Complex Networks* 12.5 (set. 2024), cnae036. ISSN: 2051-1329. DOI: 10.1093/comnet/cnae036. eprint: https://academic.oup.com/comnet/article-pdf/12/5/cnae036/59073581/cnae036.pdf. URL: https://doi.org/10.1093/comnet/cnae036.

[4] G. A. Videtta. *Implementazione di due algoritmi per il calcolo della centralità di Katz in grafi molto densi*. https://github.com/hearot/katz-centrality-dense. GitHub repository. 2025.