

Entrega: Tarea 3

Laura Basalo Tur, Camila Pérez Arévalo, Josep Roman Cardell

1. Implementar el calculo de las matrices de los núcleos de Neumann. La función debería tener dos parámetros: la matriz de adyacencia y el grado de difusión.

```
neumann_kernel = function(A, grado){  
  At = t(A)  
  K = At %*% A  
  Ter = A %*% At  
  K_hat = K %*% solve(diag(100)-grado*K)  
  T_hat = Ter %*% solve(diag(100)-grado*Ter)  
  
  list_result = list(K_hat, T_hat)  
  
  return(list_result)  
}
```

2. Aplicar el algoritmo a la matriz de adyacencia de páginas webs: webPageLinks.txt. Responda, con estos datos, el ejercicio 6.7 del libro.

```
#Creamos una semilla para obtener el mismo resultado  
set.seed(1234)  
  
#Leemos la matriz del fichero txt  
A = as.matrix(read.table("data/webPageLinks.txt"))  
  
#Generamos el grafo a partir de la matriz  
G = graph_from_adjacency_matrix(A)  
  
#Calculamos el grado de difusión  
d_out = max(degree(G, mode = "out"))  
d_in = max(degree(G, mode = "in"))  
grado = runif(1, 0, 1/min(c(d_out, d_in)))  
  
matrix_result = neumann_kernel(A, grado)  
K_hat1 = matrix_result[[1]]  
T_hat1 = matrix_result[[2]]
```

(a) Describe how the HITS algorithm relates to the Neumann Kernel technique.

La técnica del núcleo de Neumann es una generalización del algoritmo de HITS. Por consecuencia, la authority score está relacionada con el score de la matriz K_γ y el hub score con el score de la matriz T_γ . Si se crea un ranking de los nodos atendiendo a su authority score y a su K score, el orden de los nodos será el mismo, por lo que es notoria la relación entre ambos algoritmos.

Cuando el grado de difusión es grande se incrementa la importancia de los vértices más alejados del grafo, por lo que la solución que proporciona el kernel de Neumann es muy similar o igual a la puntuación que calcula el algoritmo de HITS.

```
#Calculamos el valor máximo que puede tener gamma
gamma_max = 1/min(c(d_out, d_in))
```

```
#NK socres.
```

```
Kscore1 = rowSums(K_hat1)
```

```
Tscore1 = rowSums(T_hat1)
```

```
#Comparación algoritmo HITS
```

```
#Authorities
```

```
authority_score = authority_score(G)$vector
```

```
head(authority_score,5)
```

```
##                                http...www.google.com.
##                                2.971358e-17
## http...groups.google.co.il.grphp.ie.ISO.8859.8.I.oe.ISO.8859.8.I.hl.iw.tab.wg
##                                1.980906e-17
##                                http...www.google.com.intl.iw.ads
##                                5.942717e-17
##                                http...www.google.co.il.
##                                1.980906e-17
##                                http...www.google.co.il.intl.iw.about.html
##                                5.615511e-04
```

```
head(Kscore1, 5)
```

```
##                                http...www.google.com.
##                                2.033023
## http...groups.google.co.il.grphp.ie.ISO.8859.8.I.oe.ISO.8859.8.I.hl.iw.tab.wg
##                                0.000000
##                                http...www.google.com.intl.iw.ads
##                                4.134310
##                                http...www.google.co.il.
##                                1.008188
##                                http...www.google.co.il.intl.iw.about.html
##                                15.923374
```

```
#Hub
```

```
hub_score = hub_score(G)$vector
```

```
head(hub_score,5)
```

```
##                                http...www.google.com.
##                                0
## http...groups.google.co.il.grphp.ie.ISO.8859.8.I.oe.ISO.8859.8.I.hl.iw.tab.wg
##                                0
##                                http...www.google.com.intl.iw.ads
##                                0
##                                http...www.google.co.il.
##                                0
##                                http...www.google.co.il.intl.iw.about.html
##                                0
```

```
head(Tscore1, 5)
```

```
##                                http://www.google.com/
##                                0.000000
## http://groups.google.co.il/grphp?ie=ISO-8859-8-I&oe=ISO-8859-8-I&hl=iw&tab=wg
##                                2.033023
##                                http://www.google.com/intl/iw/ads
##                                2.033023
##                                http://www.google.co.il/
##                                4.134310
##                                http://www.google.co.il/intl/iw/about.html
##                                4.134310
```

```
#Correlación entre K y T con los vectores de Authorities y Hub
cor = cor(Kscore1, authority_score)
```

Podemos observar que con una γ aleatoria, en este ejemplo 0.0081217, la correlación entre la K_γ y el Authorities de HITS es muy elevada (0.950861).

A continuación realizamos la comparación utilizando el γ con valor máximo (0.0714286):

```
#Gamma Máxima (0.07142857)
matrix_result = neumann_kernel(A, gamma_max)
K_hat2 = matrix_result[[1]]
T_hat2 = matrix_result[[2]]

Kscore2 = rowSums(K_hat2)
Tscore2 = rowSums(T_hat2)

head(authority_score,5)
```

```
##                                http...www.google.com.
##                                2.971358e-17
## http...groups.google.co.il.grphp.ie.ISO.8859.8.I.oe.ISO.8859.8.I.hl.iw.tab.wg
##                                1.980906e-17
##                                http...www.google.com.intl.iw.ads
##                                5.942717e-17
##                                http...www.google.co.il.
##                                1.980906e-17
##                                http...www.google.co.il.intl.iw.about.html
##                                5.615511e-04
```

```
head(Kscore2, 5)
```

```
##                                http...www.google.com.
##                                2.333333
## http...groups.google.co.il.grphp.ie.ISO.8859.8.I.oe.ISO.8859.8.I.hl.iw.tab.wg
##                                0.000000
##                                http...www.google.com.intl.iw.ads
##                                5.600000
##                                http...www.google.co.il.
##                                1.076923
##                                http...www.google.co.il.intl.iw.about.html
##                                -34.683034
```

```
head(hub_score,5)
```

```
##                                http...www.google.com.
##                                0
## http...groups.google.co.il.grphp.ie.IS0.8859.8.I.oe.IS0.8859.8.I.hl.iw.tab.wg
##                                0
##                                http...www.google.com.intl.iw.ads
##                                0
##                                http...www.google.co.il.
##                                0
##                                http...www.google.co.il.intl.iw.about.html
##                                0
```

```
head(Tscore2, 5)
```

```
##                                http://www.google.com/
##                                0.000000
## http://groups.google.co.il/grphp?ie=ISO-8859-8-I&oe=ISO-8859-8-I&hl=iw&tab=wg
##                                2.333333
##                                http://www.google.com/intl/iw/ads
##                                2.333333
##                                http://www.google.co.il/
##                                5.600000
##                                http://www.google.co.il/intl/iw/about.html
##                                5.600000
```

```
#Correlación entre K y T con los vectores de Authorities y Hub
cor2 = cor(Kscore2, authority_score)
```

En este caso, observamos que no hay correlación (-0.0110786). Teóricamente, para valores más altos de γ , NK tendría que converger a HITS, pero en este caso, el resultado no es el esperado, puesto que la correlación indica que los valores de NK y Auth se alejan más entre ellos.

Procedemos a usar otra formula para el calculo de las puntuaciones de Neuman Kernel. Esto solucionaría el error anterior en caso de que hubiera algún fallo en el cálculo con la función que hemos creado en el apartado anterior.

Funciones:

$$\hat{K}_{\gamma} = K \sum_{n=0}^{\infty} \gamma^n K^n$$

$$\hat{T}_{\gamma} = T \sum_{n=0}^{\infty} \gamma^n T^n$$

```
library(expm)
```

```
At = t(A)
K = At %*% A
Ter = A %*% At
sumatorio = 0
for (n in 0:7) {
```

```

    sumatorio = sumatorio + gamma_max^n * K^%n
}
K_hat = K*sumatorio
sumatorio = 0
for (n in 0:7) {
    sumatorio = sumatorio + gamma_max^n * Ter^%n
}
T_hat = Ter*sumatorio

Kscore3 = rowSums(K_hat)
Tscore3 = rowSums(T_hat)
cor(Kscore3, authority_score)

```

```
## [1] 1
```

```
cor(Tscore3, hub_score)
```

```
## [1] 1
```

Podemos observar que estos resultados son los esperados de acuerdo con la teoría. En este caso al ser un grafo relativamente pequeño el sumatorio de $n = (0, \infty)$ no es necesario, pues las puntuaciones de NK y HITS convergen con un sumatorio de $n = (0, 7)$

$$\hat{K}_\gamma = K \sum_{n=0}^7 \gamma^n K^n$$

$$\hat{T}_\gamma = T \sum_{n=0}^7 \gamma^n T^n$$

(b) How does the decay factor γ affect the output of the Neumann Kernel algorithm?

El factor de difusión (γ) da menos importancia a la influencia de los nodos más lejanos. Es decir, tiene mayor peso la conexión entre dos vértices directamente conectados que los que tienen n aristas entre ellos. La teoría dice que las matrices de KN convergen si $\gamma \geq \frac{1}{\min(\Delta^-, \Delta^+)}$, donde Δ^- es el máximo grado de entrada y Δ^+ es el máximo grado de salida del grafo. Cuando se elige un valor elevado de γ la puntuación de K_γ converge a la de autoridades de HITS.