

SIW PL7. PageRank.

Hugo Fonseca Díaz
uo258318@uniovi.es

Escuela de Ingeniería Informática. Universidad de Oviedo.

5 de noviembre de 2020

1. Requerimientos

Los requerimientos para el entorno de ejecución de esta práctica son:

- Python 3.8
- Biblioteca *numpy*

2. Detalles de la documentación

Para esta práctica se nos pedía realizar una implementación del algoritmo PageRank ideado por Larry Page y Sergey Brin. Dicho algoritmo está implementado en el script de python *graph.py*. En él se encuentra la clase *Graph*, que modela un grafo. Dicha clase es la encargada de convertir los bordes pasados como argumento en su constructor a una matriz de adyacencia. Dicha matriz sera normalizada y convertida en una matriz estocástica que será la matriz con la que trabajará el algoritmo. Además, en el constructor se definirá un diccionario *PR* que guardará el nombre de los nodos junto a su valor de PageRank. A continuación hablaremos sobre el método *page rank* más en detalle.

El método *page rank* es el siguiente:

```
def page_rank(self, damping = 0.85, limit = 1.0e-8):
    M = self.__stochastic__(self.M) # Stochastic matrix
    N = M.shape[1] # Number of nodes
    R = np.full((1, N), (1 / N)) # Solution eigenvector
    R = R / np.linalg.norm(R, 1)
    M_hat = (1 - damping) / N + damping * M # Matrix for calculating PR
    while True: # Do while the result is over the limit
        prev_R = R
        R = np.dot(R, M_hat) # Recalculate solution eigenvector
        if np.linalg.norm(R - prev_R, 2) <= limit:
            break
    # Dictionary with the PR values
```

```
# key = node, value = PR(node)
pos = 0
for n in self.PR.keys():
    self.PR[n] = R[0, pos] # Insert eigenvectors into PR dictionary
    pos += 1
return self.PR
```

Para entenderlo mejor vamos a hablar sobre las variables utilizadas:

- M : matriz de adyacencia. Normalizada y estocástica.
- N : número de nodos en el grafo.
- R : *eigenvector* solución con los valores del PageRank.
- M_{hat} : Matriz con la que calculamos los valores de R en cada iteración.
- PR : diccionario cuyas claves son los nodos del grafo y cuyos valores son el PageRank asociado a cada uno de esos nodos.

El grueso del algoritmo está inspirado en el código de Python ofrecido por *Wikipedia* en [1].

Referencias

- [1] Entrada de la Wikipedia (Inglés) para el algoritmo *PageRank*, <https://en.wikipedia.org/wiki/PageRank#Python>. Última vez accedido 5 Noviembre 2020.