

Proyecto 1 Algoritmo PageRank

(I) ¿Cuál es la página web más importante según el algoritmo si se inicia con el vector $(1/n, \dots, 1/n)$?

```
Vector de importancia (PageRank): [0.05122182 0.05159775 0.05893271
0.03195045 0.04990777 0.03004057 0.03578006 0.03365421 0.05905104
0.01590684 0.04255384 0.03601719 0.04214509 0.01255872 0.05734512
0.03643153 0.03676748 0.03632105 0.03648724 0.04684105 0.03648724
0.03626799 0.03593945 0.04688204 0.03619032 0.00672143]
```

Por lo que la página más importante es: 9

(II) Defina ahora un vector π de la siguiente forma

$\pi_k =$
 $\begin{cases} 1/s & \text{si la página } k \text{ tiene terminación .ru,} \\ 0 & \text{e.o.c.} \end{cases}$

donde s es el número de páginas con terminación .ru. ¿La página de la pregunta 1 sigue siendo la más importante?

```
Vector pi para .ru: [0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
0. 0. 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0. ]
```

Por lo que la página más importante ya no es la 9

(III) Considere ahora la ecuación iterativa $\pi_{m+1} = \pi_m(dP + (1-d)U)$, donde la matriz U queda definida con entradas $u_{i,j} = 1/n$, para toda $i, j \in \{1, \dots, n\}$ y $d \in [0, 1]$. Utilice los valores 0.5, 0.85, 1 para d y reporte los resultados. De los tres valores anteriores para d , ¿cuál considera que es el mejor (según los ordenamientos obtenidos) y por qué?

De acuerdo para cada valor para d

- Vector de importancia con $d=0.5$:

```
[0.04415429 0.04705603 0.04918411 0.03818528 0.04680961
0.03384302 0.03747264 0.03860688 0.04888472 0.02547617 0.03746744
0.03675901 0.0392058 0.02400394 0.04729667 0.03719886 0.0378381
0.0370862 0.03722276 0.04384505 0.03722276 0.03705017 0.03635095
0.04385448 0.03694605 0.02097902]
```

La página más importante con $d=0.5$ es: 3

- Vector de importancia con $d=0.85$:

```
[0.05122182 0.05159775 0.05893271 0.03195045 0.04990777
0.03004057 0.03578006 0.03365421 0.05905104 0.01590684 0.04255384
0.03601719 0.04214509 0.01255872 0.05734512 0.03643153 0.03676748]
```

```
0.03632105 0.03648724 0.04684105 0.03648724 0.03626799 0.03593945
0.04688204 0.03619032 0.00672143]
```

La página más importante con $d=0.85$ es: 9

- Vector de importancia con $d=1$:

```
[5.59805340e-02 5.26195151e-02 6.38710118e-02 2.66558675e-02
4.98894259e-02 2.80673115e-02 3.55618496e-02 2.93129503e-02
6.47526802e-02 1.18540519e-02 4.73189869e-02 3.50427515e-02
4.43391609e-02 6.66393532e-03 6.34545829e-02 3.60575101e-02
3.60578854e-02 3.60576588e-02 3.60577689e-02 4.80769408e-02
3.60577689e-02 3.60574667e-02 3.60575592e-02 4.80770637e-02
3.60577620e-02 1.04385552e-24]
```

La página más importante con $d=1$ es: 9

El mejor valor de d es 0.85 porque mantiene equilibrio entre la aleatoriedad y la estructura del grapho garantizando de cierta forma aleatoriedad que ayuda a evitar problemas como graphos desconectados o ciclos cerrados mientras que en $d=1$ no hay aleatoriedad pues $1-d=0$ lo que significa que puede tener problemas con grafos desconectados o con problemas de acceso y con $d=0.5$ hay menor dispersión en los valores del vector de importancia y da más peso al factor de aleatoriedad.