Información no estructurada

Práctica 4 – Pagerank y sistemas de recomendación

Javier Aróstegui Martín

Contenido

[Parte 1: Pagerank 2](#_Toc71481042)

[Fórmulas utilizadas: 2](#_Toc71481043)

[Resultados: 2](#_Toc71481044)

[R = 0.1 sin personalizar: 2](#_Toc71481045)

[R = 0.85 sin personalizar: 2](#_Toc71481046)

[R = 0.1 con personalización en los nodos A y D: 2](#_Toc71481047)

[R = 0.85 con personalización en los nodos A y D: 2](#_Toc71481048)

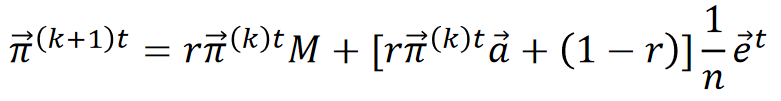
[Parte 2: Recomendación: 3](#_Toc71481049)

[Fórmulas utilizadas: 3](#_Toc71481050)

[Resultados: 4](#_Toc71481051)

# Parte 1: Pagerank

## Fórmulas utilizadas:



He utilizado esta fórmula para no tener que guardar la matriz G, que aunque en este ejemplo no es muy grande, puede llegar a ser muy ineficiente guardarla.



Para el pagerank personalizado he utilizado la misma fórmula, pero utilizando el vector *v* en vez del *e*

## Resultados:

### R = 0.1 sin personalizar:

En esta lista se muestran los resultados finales de pagerank sin personalizar desde el nodo A hasta el nodo E. En este caso se hacen 300 iteraciones, pero no son necesarias tantas para que el resultado converja.



### R = 0.85 sin personalizar:

Cuando aumentamos r le estamos dando más valor a los nodos importantes y por tanto observamos que el nodo B acapara mayor pagerank ya que es el nodo más apuntado, por nodos con un pagerank relativamente alto y no apunta a nadie.



### R = 0.1 con personalización en los nodos A y D:

En este caso al ser r muy pequeña le estamos dando mucha importancia al vector v (vector de personalización) y por tanto los nodos afines al usuario obtienen un pagerank muy elevado.



### R = 0.85 con personalización en los nodos A y D:

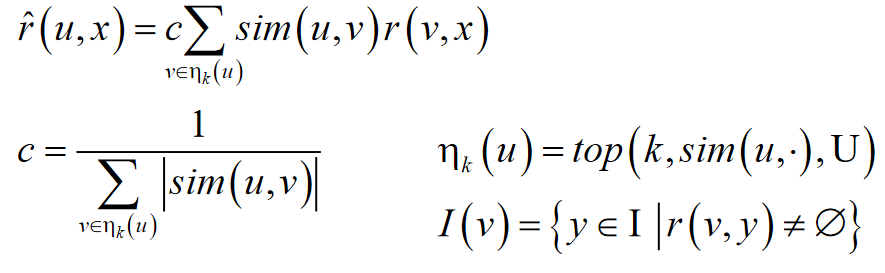
Ahora que hemos aumentado r, ya no se le da tanta importancia al vector de personalización y por tanto el nodo B (al que le corresponde el pagerank más alto) recupera su valor alto. Los nodos A y D han aumentado su valor respecto al pagerank no personalizado.

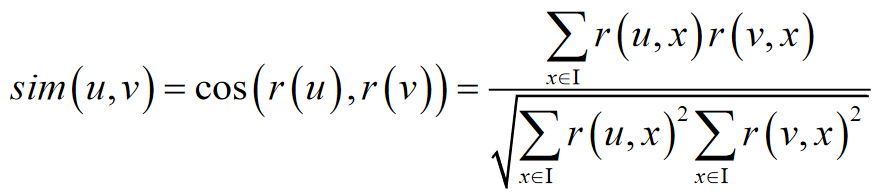


# Parte 2: Recomendación:

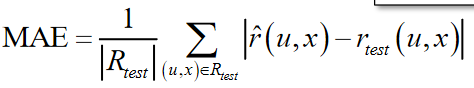
## Fórmulas utilizadas:

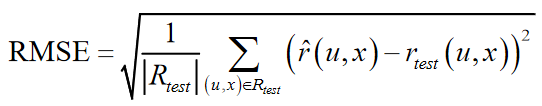
Para la recomendación he utilizado **knn basado en usuarios** con las siguientes fórmulas para el rating esperado y la similitud:

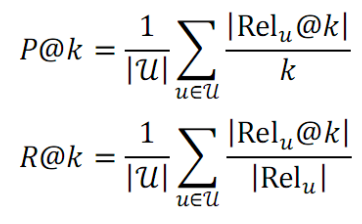




Para la evaluación de recomendación he usado MAE, RMSE, P@k y R@k como se indica en el enunciado usando las siguientes formulas:

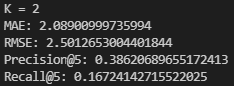




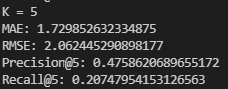


## Resultados:

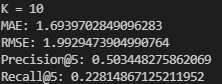
### K = 2:



K = 5:



### K = 10:



Podemos observar como cuando aumentamos K, todas las métricas mejoran. Esto tiene sentido y nos indica que el algoritmo es capaz de aprender lo necesario cuando se le enseña suficiente información.