

Algoritmo PageRank (ranking de página) de Google

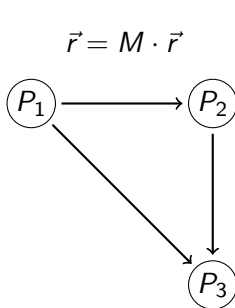
La importancia de estar bien conectado.

Dado un conjunto de páginas web P_1, P_2, \dots, P_N con enlaces entre ellas se pretende definir un ranking de página de forma que:

“El ranking de la página rp_i es la media de los rankings de las páginas que llegan a ella”

- El conjunto de distribuciones Δ . Distribuciones propias.
- Matriz de enlaces E y matriz de enlaces normalizada M .
- Vector de ranking $\vec{r} = (rp_1, rp_2, \dots, rp_N)$ (en los calculos se toma como una columna. Además $\vec{r} \in \Delta$).

Ejemplo



Matriz de enlaces $\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}$

Normalizada $\begin{pmatrix} 0 & 0 & 0 \\ 1/2 & 0 & 0 \\ 1/2 & 1 & 0 \end{pmatrix}$

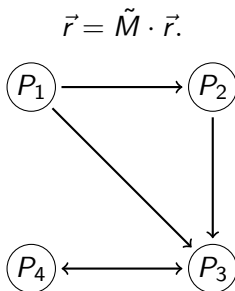
Matriz extendida $\begin{pmatrix} 0 & 0 & 1/3 \\ 0.5 & 0 & 1/3 \\ 0.5 & 1 & 1/3 \end{pmatrix}$

Solución

$$\vec{r} = \vec{0}.$$

- Matriz de enlaces extendida (normalizada) \tilde{M} .
- Matriz estocástica.

Segundo ejemplo



Solución

$$\vec{r} = (0, 0, 0.5, 0.5).$$

La matriz de Google

- Matriz de unos, $\mathbb{1}$.
- Matriz de Google $\mathbb{G}_\alpha = \alpha \tilde{M} + \frac{1-\alpha}{N} \mathbb{1}$.

Ecuación de Google

$$\vec{r} = \mathbb{G}_\alpha \cdot \vec{r}, \quad \vec{r} \in \Delta$$

- La ecuación de Google tiene una única solución, y es siempre una distribución propia.
- Sergey Brin y Larry Page, creadores de PageRank, sugieren $\alpha = 0.85$.