

2^{as} Jornadas Ingeniería para Matemáticas InxMat2 (Eng2Math2)

CIEM, Castro Urdiales, Cantabria

16 de mayo de 2019

Carlos Marijuán

Universidad de Valladolid

Cómo mejorar el PageRank de una web local

0. Introducción

Es bien sabido que PageRank es el algoritmo que utiliza Google para ordenar los resultados de las búsquedas. El algoritmo está basado en una fórmula que asigna a cada página web un valor, su importancia, que depende básicamente de los vínculos de las demás páginas con ella.

Brinkmeier tiene una visión de la red más amplia: considera todas las trayectorias que conducen a una página determinada, lo que le permite dar una formulación analítica, no lineal, del PageRank.

En cualquiera de estas formulaciones se puede observar que el PageRank de una página depende de la topología de la red, es decir, de la conectividad en el grafo dirigido que es la World Wide Web.

Nosotros nos preguntamos qué modificaciones podemos hacer en una red local para mejorar el PageRank de su página principal. ¿Cómo organizar los vínculos dentro de la red para mejorar su visibilidad en la WWW?

Comenzaremos estudiando estructuras gráficas dirigidas muy rígidas como árboles con raíz: analizaremos qué manipulaciones conviene hacer para mejorar el PageRank de estas redes locales y las extenderemos a estructuras cada vez menos rígidas como digrafos acíclicos o árboles con raíz en los que admitiremos arcos bidireccionales e, incluso, ciclos de cualquier longitud.

Finalmente, aplicaremos las técnicas desarrolladas a cualquier red, por compleja que sea, vista a través de su digrafo condensación.

1. Medidas de centralidad en redes (V, E)

Las técnicas que vamos a desarrollar tienen aplicación en el contexto de grandes redes.

Internet. Red física: $V = \{\text{ordenadores}\}$, $E = \{\text{conexiones físicas entre ordenadores}\}$.

World Wide Web: Red de información: $V = \{\text{páginas web}\}$, $E = \{\text{conexiones por hyperlinks}\}$.

Tim Berners-Lee (MIT, Oxford, 11-3-1989), intercambiar información entre científicos del CERN.

Redes sociales: $V = \{\text{personas}\}$, $E = \{\text{conexiones entre ellas}\}$.

Introducidas por Jacob Moreno (inmigrante rumano en USA) en 1930's como Sociogramas.

Redes neuronales: $V = \{\text{neuronas}\}$, $E = \{\text{conexiones entre ellas}\}$.

Redes bioquímicas: redes metabólicas, interacción entre proteínas, regulación genética,...

Redes tecnológicas: red de conducciones de gas natural en Europa.

Estas redes tienen miles, millones de vértices. Su visualización no ayuda mucho a su análisis.

Entonces se emplean ciertas **métricas** para localizar sus datos estructurales más relevantes.

Algunas de las métricas más útiles son conocidas como **medidas de centralidad**, que tratan de analizar o buscar los vértices *centrales* o *más importantes* de una red.

1. Medidas de centralidad en redes (V, E)

La centralidad del grado: el número de arcos incidentes en un vértice. En una red social, personas con más conexiones pueden estar más informadas, tener más influencia, prestigio, ...

El número de citas que un artículo recibe de otros, su ingrado, da una medida de la influencia del artículo y se utiliza como una métrica para juzgar el impacto de la investigación científica.

La centralidad del autovector, Landau 1895. Un vértice con grado elevado es importante porque está conectado con otros muchos; pero no todos los vecinos tienen la misma importancia.

La centralidad del autovector da a cada vértice una importancia proporcional a la suma de las importancias de sus vecinos.

Si x_i es la importancia del vértice v_i en una red con matriz de adyacencia $A = (a_{ij})$, entonces

$$x_i = \alpha \sum_{j=1}^n a_{ij} x_j \implies \mathbf{x} = \alpha \mathbf{A} \mathbf{x}, \quad \mathbf{A} \mathbf{x} = \frac{1}{\alpha} \mathbf{x}.$$

En una red social o de información como la WWW se considera que la importancia de un vértice es proporcional a la importancia de los vértices que se conectan a él, en este caso

$$x_i = \alpha \sum_{j=1}^n a_{ji} x_j \implies \mathbf{x} = \alpha \mathbf{A}^t \mathbf{x}, \quad \mathbf{A}^t \mathbf{x} = \frac{1}{\alpha} \mathbf{x}.$$

1. Medidas de centralidad en redes (V, E)

La centralidad de Katz, 1953: considera para cada vértice v_i la influencia de todos los vértices v_j conectados a él por una trayectoria;

La contribución de cada vecino v_j está penalizada por un factor de atenuación α , de modo que la contribución a v_i de un vértice v_j conectado a v_i por una trayectoria de longitud k es α^k , entonces

$$x_i = \sum_{k=1}^{\infty} \sum_{j=1}^n \alpha^k a_{ji}^{(k)} \implies \mathbf{x} = \left(\sum_{k=1}^{\infty} \alpha^k (\mathbf{A}^t)^k \right) \mathbf{e} = \left(-\mathbf{I} + \sum_{k=0}^{\infty} (\alpha \mathbf{A}^t)^k \right) \mathbf{e}$$
$$\mathbf{x} = (-\mathbf{I} + (\mathbf{I} - \alpha \mathbf{A}^t)^{-1}) \mathbf{e}$$

Medida α -centralidad, Bonacich 1987: considera que la importancia de un vértice es proporcional a la importancia de los vértices que se conectan a él, pero cada vértice parte con una importancia positiva β :

$$x_i = \beta + \alpha \sum_{j=1}^n a_{ji} x_j \implies \mathbf{x} = \beta \mathbf{e} + \alpha \mathbf{A}^t \mathbf{x} \implies \mathbf{x} = \beta (\mathbf{I} - \alpha \mathbf{A}^t)^{-1} \mathbf{e}$$

Katz y α -centralidad se consideran medidas equivalentes y convergen si $\alpha < \rho(\mathbf{A})$.

El problema de estas medidas es que la importancia de los vértices se transmite uniformemente a los vértices conectados. Si un vértice tiene una elevada α -centralidad, esta se transmite a sus vértices enlazados que pasan a tener una elevada α -centralidad. Para tener una α -centralidad elevada basta tener un “esponsor altamente reputado”. Esta anomalía fue corregida por la medida de centralidad PageRank repartiendo equitativamente la importancia de cada vértice entre sus vecinos.

“If I have seen further it is by standing on the shoulders of giants.” I. Newton

1. Medidas de centralidad en redes (V, E)

Centralidad PageRank, Brin y Page, 1998. La importancia de cada vértice se distribuye equitativamente entre sus vértices vecinos, dividida por su exgrado.

Ahora la importancia x_i del vértice v_i viene dada por

$$x_i = \beta + \alpha \sum_{j=1}^n a_{ji} \frac{x_j}{od(v_j)}$$

Esta es la función PageRank en la que $\beta = \frac{1 - \alpha}{n}$ y $\alpha \in (0, 1)$,

normalmente $\alpha = 0,85$.

“If I have seen further it is by standing on the shoulders of giants.” I. Newton

1. Introducción: un poco de historia

1989: Nace la *Word Wide Wed (WWW)*. Los usuarios de la Red (*Web surfers*) necesitan herramientas eficientes para localizar información en este grafo.

1994 : Nace **Yahoo** que ofrece:

- un directorio de páginas clasificadas por su contenido
- una *máquina de búsqueda (Search Engine)* para hallar *sitios (páginas)* en la Web basados en *palabras claves (Keywords)*.

Yahoo clasifica páginas según su contenido descrito en los *Meta Tags* de las páginas. Presenta lo hallado en una lista ordenada por relevancia:

$$\text{Ranking} = (\text{Relevancia del contenido})$$

Otros buscadores en el mercado de la WWW: Altavista, Lycos, Excite,...

1998: Nace el buscador **Google** (*googol* = 10^{100})

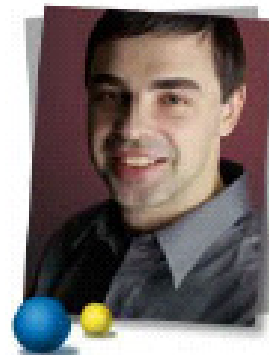
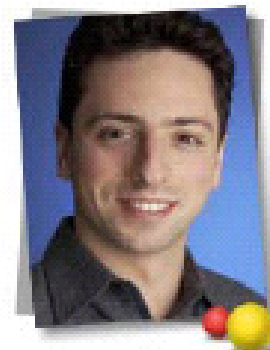
Creado por **Sergei Brin** y **Lawrence Page** en la Univ. de Stanford.

Su fórmula PageRank, junto con análisis contextual de páginas, ofrece búsquedas más precisas y una clasificación de páginas más honesta:

$$\text{Ranking} = (\text{Relevancia del contenido}).(\text{PageRank})$$

15-Nov-2003 : **Google's Florida update** considera también un PageRank local

$$\text{Ranking} = (\text{Relevancia del contenido}).(\text{PR Total}).(\text{PR Local})$$



2. Fórmula de Brin–Page. El surfista aleatorio

Un modelo para la **World Wide Web** es un digrafo (V, A) , donde

un vértice representa cualquier documento depositado en la web (*una página p_i*),

un arco (p_j, p_i) indica que hay un vínculo (link) desde la página p_j a la página p_i .

Brin y Page proponen evaluar cada página p_i con un número real positivo (su PageRank):

$$\mathcal{P}(p_i) = \frac{1 - \alpha}{N} + \alpha \sum_{(p_j, p_i) \in A} \frac{\mathcal{P}(p_j)}{od(p_j)} \quad (1)$$

La motivación, dada por los autores, es que la fórmula (1) **modela el comportamiento de un surfista aleatorio** de la Web que, estando en cierta página p_j , decide seguir por alguno de los vínculos de dicha página con probabilidad α , o decide saltar a cualquier otra página con probabilidad $1 - \alpha$. La probabilidad de elegir el vínculo de la página p_j que conduce a la página p_i depende del número $od(p_j)$ de vínculos en p_j ; de modo que

El PageRank de p_i es la probabilidad de que un usuario alcance la página p_i directamente o siguiendo todos los vínculos adecuados.

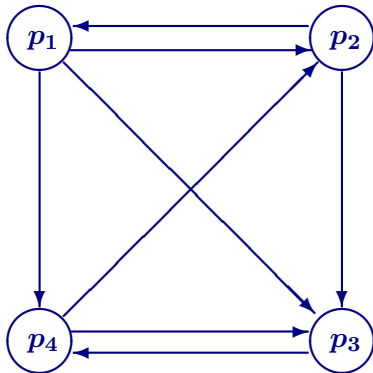
2. Fórmula de Brin–Page. El surfista aleatorio

$$\mathcal{P}(p_i) = \frac{1 - \alpha}{N} + \alpha \sum_{(p_j, p_i) \in A} \frac{\mathcal{P}(p_j)}{od(p_j)} \implies \sum_{i=1}^N \mathcal{P}(p_i) = (1 - \alpha) + \alpha \sum_{i=1}^N \mathcal{P}(p_i)$$

$\sum_{i=1}^N \mathcal{P}(p_i) = 1 \implies$ los PageRanks son una **distribución de probabilidad sobre la WWW**.

$\mathbf{p} = (\mathcal{P}(p_1), \dots, \mathcal{P}(p_N))^t$ se describe por medio de una **matriz de transición** $T = (t_{ij})$ donde t_{ij} simboliza la probabilidad de alcanzar la página p_i desde la página p_j :

$$t_{ij} = \begin{cases} \frac{1-\alpha}{N} + \frac{\alpha}{od(p_j)} & \text{si hay un link desde } p_j \text{ a } p_i, \\ \frac{1-\alpha}{N} & \text{en otro caso} \end{cases} \quad \text{Entonces } \boxed{\mathbf{p} = T\mathbf{p}}$$



$$p_1 = \frac{1-0,84}{4} + 0,84 \frac{p_2}{2}$$

$$p_2 = \frac{1-0,84}{4} + 0,84 \left(\frac{p_1}{3} + \frac{p_4}{2} \right)$$

$$p_3 = \frac{1-0,84}{4} + 0,84 \left(\frac{p_1}{3} + \frac{p_2}{2} + \frac{p_4}{2} \right)$$

$$p_4 = \frac{1-0,84}{4} + 0,84 \left(\frac{p_1}{3} + p_3 \right)$$

$$T = \begin{pmatrix} 0,04 & 0,46 & 0,04 & 0,04 \\ 0,32 & 0,04 & 0,04 & 0,46 \\ 0,32 & 0,46 & 0,04 & 0,46 \\ 0,32 & 0,04 & 0,88 & 0,04 \end{pmatrix}$$

$$\mathbf{p} = (0,13196, 0,21896, 0,31093, 0,33813)^t$$

Para obtener \mathbf{p} basta resolver el sistema de ecuaciones lineales en las incógnitas p_1, p_2, p_3, p_4 o hallar el autovector de la matriz T asociado al autovalor 1.

3. Cálculo del PageRank

Teoría de Perron Frobenius

Perron, 1908. $A > 0$, $\rho(A)$ su radio espectral:

$\rho(A)$ es simple y $\rho(A) > |\lambda_i|$

$\rho(A)$ tiene asociado un autovector positivo.

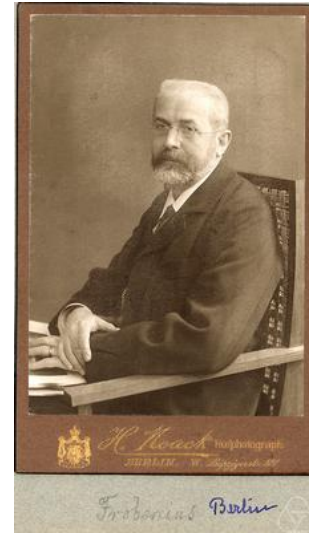
Frobenius, 1912. $A \geq 0$ irreducible:

$\rho(A)$ es simple y $\rho(A) \geq |\lambda_i|$. $P_A(\lambda) = \lambda^m [\lambda^h - \rho^h(A)] [\lambda^h - (\delta_2 \rho(A))^h] \cdots [\lambda^h - (\delta_r \rho(A))^h]$

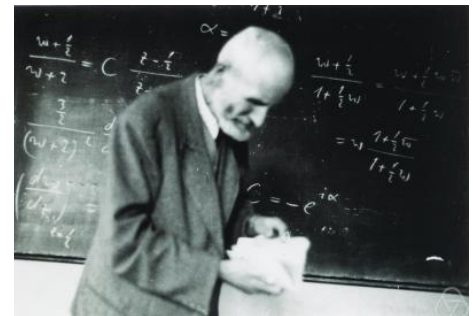
$\rho(A)$ tiene asociado un autovector positivo.

Si $h = 1$, A es **primitiva**, entonces $\rho(A) > |\lambda_i|$.

Teorema. El vector de probabilidad p es la única solución de la ecuación $p = Tp$.



Ferdinand Georg Frobenius 1849-1917



Oskar Perron 1880-1975

3. Cálculo del PageRank

La Web contiene cientos de miles de millones de páginas y la ecuación $p = Tp$ no es computacionalmente resoluble por métodos directos.

Google usa el **método de la potencia** para calcular aproximadamente los PageRanks, empezando con un vector de probabilidad inicial arbitrario $p^{(0)}$:

$$\begin{aligned} p^{(1)} &= Tp^{(0)} \\ p^{(2)} &= Tp^{(1)} = T^2p^{(0)} \\ &\dots\dots\dots \\ p^{(k)} &= Tp^{(k-1)} = T^k p^{(0)} \end{aligned}$$

Método de la potencia. $\lim_{k \rightarrow \infty} p^{(k)} = \lim_{k \rightarrow \infty} T^k p^0 = p.$

Si $p^{(0)}$ es un vector de probabilidad, entonces los $p^{(k)} = T^k p^{(0)}$ son vectores de probabilidad que convergen al vector p solución de la ecuación $p = Tp$.

El método converge si $\rho(T) > |\lambda_i|$ y para esto es suficiente que la matriz T sea primitiva. El módulo del segundo autovalor da la velocidad de convergencia.

Brin y Page: “Son necesarias 100 iteraciones para obtener buenas aproximaciones de los PageRanks de todos los vértices de la web.”

3. Cálculo del PageRank

En nuestro ejemplo, teníamos la solución $p = (0,13196, 0,21896, 0,31093, 0,33813)^t$

Si empezamos con el vector de probabilidad $p^{(0)} = (1, 0, 0, 0)^t$ se tiene

$$p^{(1)} = Tp^{(0)} = \begin{pmatrix} 0,04 & 0,46 & 0,04 & 0,04 \\ 0,32 & 0,04 & 0,04 & 0,46 \\ 0,32 & 0,46 & 0,04 & 0,46 \\ 0,32 & 0,04 & 0,88 & 0,04 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0,4 \\ 0,32 \\ 0,32 \\ 0,32 \end{pmatrix}$$

$$p^{(2)} = Tp^{(1)} = \begin{pmatrix} 0,04 & 0,46 & 0,04 & 0,04 \\ 0,32 & 0,04 & 0,04 & 0,46 \\ 0,32 & 0,46 & 0,04 & 0,46 \\ 0,32 & 0,04 & 0,88 & 0,04 \end{pmatrix} \begin{pmatrix} 0,04 \\ 0,32 \\ 0,32 \\ 0,32 \end{pmatrix} = \begin{pmatrix} 0,1744 \\ 0,1856 \\ 0,32 \\ 0,32 \end{pmatrix}$$

.....

$$p^{(15)} = Tp^{(14)} = \begin{pmatrix} 0,04 & 0,46 & 0,04 & 0,04 \\ 0,32 & 0,04 & 0,04 & 0,46 \\ 0,32 & 0,46 & 0,04 & 0,46 \\ 0,32 & 0,04 & 0,88 & 0,04 \end{pmatrix} \begin{pmatrix} 0,13197 \\ 0,21896 \\ 0,31093 \\ 0,33813 \end{pmatrix} = \begin{pmatrix} 0,13196 \\ 0,21896 \\ 0,31093 \\ 0,33813 \end{pmatrix}$$

3. Cálculo del PageRank

Si empezamos con el vector de probabilidad $p^{(0)} = (0,25, 0,25, 0,25, 0,25)^t$ tendremos

Iteración	$\mathcal{P}(p_1)$	$\mathcal{P}(p_2)$	$\mathcal{P}(p_3)$	$\mathcal{P}(p_4)$
0	0.25	0.25	0.25	0.25
1	0.145	0.215	0.32	0.32
2	0.1303	0.215	0.3053	0.3494
3	0.13029	0.22323	0.31353	0.33293
4	0.13375	0.21631	0.31007	0.33985
5	0.13085	0.22018	0.31104	0.33791
6	0.13247	0.21856	0.31104	0.33791
7	0.13179	0.21901	0.31081	0.33837
8	0.13198	0.21901	0.31100	0.33798
9	0.13198	0.21891	0.31089	0.33820
10	0.13194	0.21900	0.31094	0.33811
11	0.13198	0.21895	0.31093	0.33813
12	0.13196	0.21896	0.31093	0.33813

3. Cálculo del PageRank

Mito: Algunas páginas como las de Yahoo se les asigna un PageRank inicial mayor que al resto y por eso obtienen un ranking superior después de las iteraciones.

No es así: No importa con qué valor arranca el algoritmo, pues éste converge al PageRank real.

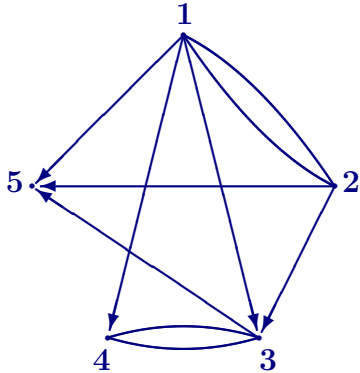
El valor inicial solo influye en el número de iteraciones.

En el mejor de los casos, podemos empezar con un vector de probabilidad próximo a la solución:

Iteración	$\mathcal{P}(p_1)$	$\mathcal{P}(p_2)$	$\mathcal{P}(p_3)$	$\mathcal{P}(p_4)$
0	0.13	0.22	0.31	0.34
1	0.1324	0.2192	0.3116	0.3368
2	0.13206	0.21852	0.31059	0.33881
3	0.13178	0.21928	0.31106	0.33787
4	0.13209	0.21880	0.31090	0.33819
5	0.13189	0.21902	0.31092	0.33814
6	0.13199	0.21895	0.31094	0.33810
7	0.13196	0.21896	0.31092	0.33815
8	0.13196	0.21897	0.31093	0.33812
9	0.13196	0.21896	0.31093	0.33813

4. El PageRank funciona sin sumideros

$$\mathcal{P}(p_i) = \frac{1 - \alpha}{N} + \alpha \sum_{(p_j, p_i) \in A} \frac{\mathcal{P}(p_j)}{\text{od}(p_j)}$$



$$p_1 = \frac{1 - \alpha}{5} + \alpha \frac{p_2}{3}$$

$$p_2 = \frac{1 - \alpha}{5} + \alpha \frac{p_1}{4}$$

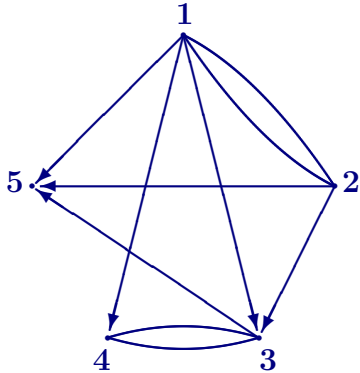
$$p_3 = \frac{1 - \alpha}{5} + \alpha \left(\frac{p_1}{4} + \frac{p_2}{3} + p_4 \right)$$

$$p_4 = \frac{1 - \alpha}{5} + \alpha \left(\frac{p_1}{4} + \frac{p_3}{2} \right)$$

$$p_5 = \frac{1 - \alpha}{5} + \alpha \left(\frac{p_1}{4} + \frac{p_2}{3} + \frac{p_3}{2} \right)$$

4. El PageRank funciona sin sumideros

$$\mathcal{P}(p_i) = \frac{1 - \alpha}{N} + \alpha \sum_{(p_j, p_i) \in A} \frac{\mathcal{P}(p_j)}{\text{od}(p_j)}$$



$$\begin{aligned} p_1 - \alpha \frac{p_2}{3} &= \frac{1 - \alpha}{5} \\ -\alpha \frac{p_1}{4} + p_2 &= \frac{1 - \alpha}{5} \\ -\alpha \frac{p_1}{4} - \alpha \frac{p_2}{3} + p_3 - \alpha p_4 &= \frac{1 - \alpha}{5} \\ -\alpha \frac{p_1}{4} - \alpha \frac{p_3}{2} + p_4 &= \frac{1 - \alpha}{5} \\ -\alpha \frac{p_1}{4} - \alpha \frac{p_2}{3} - \alpha \frac{p_3}{2} + p_5 &= \frac{1 - \alpha}{5} \end{aligned}$$

Sumando las ecuaciones se tiene:

$$(1 - \alpha)p_1 + (1 - \alpha)p_2 + (1 - \alpha)p_3 + (1 - \alpha)p_4 + p_5 = 1 - \alpha$$

$$p_1 + p_2 + p_3 + p_4 + p_5 + \frac{p_5}{1 - \alpha} = 1$$

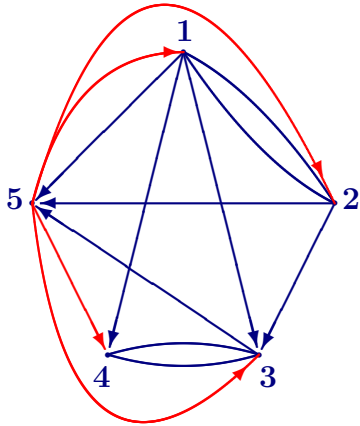
$\implies p_1 + p_2 + p_3 + p_4 + p_5 < 1$. Por tanto, p no es vector de probabilidad.

Para posibles comparaciones con otras situaciones, el vector de PageRanks se 1-normaliza:

$$p^1 = \frac{p}{\sum_{i=1}^N \mathcal{P}(p_i)} = \frac{(\mathcal{P}(p_1), \dots, \mathcal{P}(p_N))}{\sum_{i=1}^N \mathcal{P}(p_i)}$$

4. El PageRank funciona sin sumideros

$$\mathcal{P}(p_i) = \frac{1 - \alpha}{N} + \alpha \sum_{(p_j, p_i) \in A} \frac{\mathcal{P}(p_j)}{od(p_j)}$$



$$\begin{array}{rcl} p_1 & -\alpha \frac{p_2}{3} & -\alpha \frac{p_5}{4} = \frac{1-\alpha}{5} \\ -\alpha \frac{p_1}{4} & +p_2 & -\alpha \frac{p_5}{4} = \frac{1-\alpha}{5} \\ -\alpha \frac{p_1}{4} & -\alpha \frac{p_2}{3} & +p_3 & -\alpha p_4 & -\alpha \frac{p_5}{4} = \frac{1-\alpha}{5} \\ -\alpha \frac{p_1}{4} & & -\alpha \frac{p_3}{2} & +p_4 & -\alpha \frac{p_5}{4} = \frac{1-\alpha}{5} \\ -\alpha \frac{p_1}{4} & -\alpha \frac{p_2}{3} & -\alpha \frac{p_3}{2} & & +p_5 = \frac{1-\alpha}{5} \end{array}$$

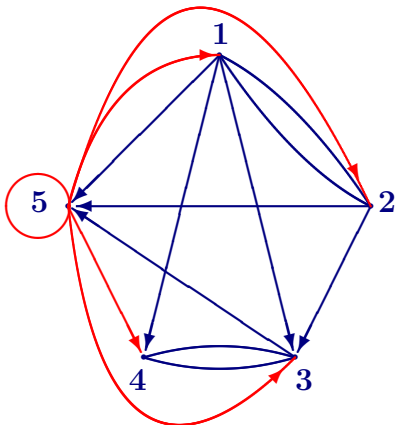
El vector p es un vector de probabilidad: $\sum_{i=1}^N \mathcal{P}(p_i) = 1$.

Los valores p_1 y p_2 son bastante mayores que en el caso anterior.

En consecuencia, en este caso p_3 y p_4 disminuyen, mientras que p_5 aumenta.

De modo que con esta operación aparentemente equitativa se ha modificado sustancialmente la posición relativa de los PageRanks de las páginas.

4. El PageRank funciona sin sumideros



$$\mathcal{P}(p_i) = \frac{1 - \alpha}{N} + \alpha \sum_{(p_j, p_i) \in A} \frac{\mathcal{P}(p_j)}{od(p_j)}$$

$$\begin{array}{rcccc} p_1 & -\alpha \frac{p_2}{3} & & & -\alpha \frac{p_5}{5} = \frac{1-\alpha}{5} \\ -\alpha \frac{p_1}{4} & +p_2 & & & -\alpha \frac{p_5}{5} = \frac{1-\alpha}{5} \\ -\alpha \frac{p_1}{4} & -\alpha \frac{p_2}{3} & +p_3 & -\alpha p_4 & -\alpha \frac{p_5}{5} = \frac{1-\alpha}{5} \\ -\alpha \frac{p_1}{4} & & -\alpha \frac{p_3}{2} & +p_4 & -\alpha \frac{p_5}{5} = \frac{1-\alpha}{5} \\ -\alpha \frac{p_1}{4} & -\alpha \frac{p_2}{3} & -\alpha \frac{p_3}{2} & & +p_5 & -\alpha \frac{p_5}{5} = \frac{1-\alpha}{5} \end{array}$$

Todas las ecuaciones incluyen la misma cantidad $\alpha \frac{p_5}{5}$. En consecuencia, todos los p_i aumentan.

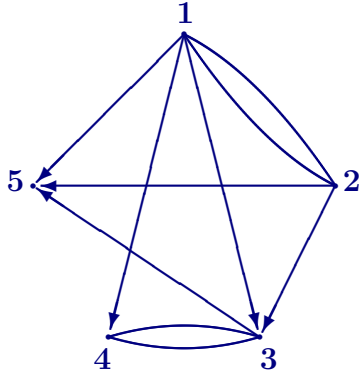
En una web sin sumideros: $\sum_{i=1}^N \mathcal{P}(p_i) = 1 - \alpha + \alpha \sum_{i=1}^N \mathcal{P}(p_i)$

$\sum_{i=1}^N \mathcal{P}(p_i) = 1 \implies$ los PageRanks son una distribución de probabilidad.

La conexión de los sumideros con todos los vértices, incluidos ellos mismos, equivale a la 1-normalización del vector de PageRank con sumideros.

4. El PageRank funciona sin sumideros

$p = (\mathcal{P}(p_1), \dots, \mathcal{P}(p_N))^t$ se describe por medio de una **matriz de transición** $T = (t_{ij})$ donde t_{ij} simboliza la probabilidad de alcanzar la página p_i desde la página p_j :



matriz de adyacencia $A = (a_{ij})$,

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

matriz de conectividad ponderada

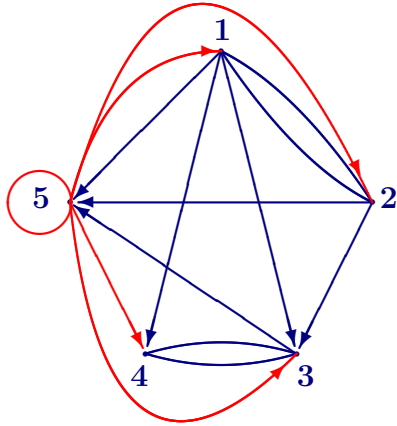
$$C = \left(\frac{a_{ji}}{od(p_i)} \right) = \begin{pmatrix} 0 & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{4} & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{3} & 0 & 1 & 0 \\ \frac{1}{4} & \frac{1}{3} & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{pmatrix}$$

matriz de transición $T = \frac{1-\alpha}{N}J + \alpha C$, $t_{ij} = \begin{cases} \frac{1-\alpha}{N} + \frac{\alpha}{od(p_j)} & \text{si } p_j \longrightarrow p_i, \\ \frac{1-\alpha}{N} & \text{en otro caso} \end{cases}$

$$T = \begin{pmatrix} \frac{1-\alpha}{5} & \frac{1-\alpha}{5} + \frac{\alpha}{3} & \frac{1-\alpha}{5} & \frac{1-\alpha}{5} & \frac{1-\alpha}{5} \\ \frac{1-\alpha}{5} + \frac{\alpha}{4} & \frac{1-\alpha}{5} & \frac{1-\alpha}{5} & \frac{1-\alpha}{5} & \frac{1-\alpha}{5} \\ \frac{1-\alpha}{5} + \frac{\alpha}{4} & \frac{1-\alpha}{5} + \frac{\alpha}{3} & \frac{1-\alpha}{5} & \frac{1-\alpha}{5} + \alpha & \frac{1-\alpha}{5} \\ \frac{1-\alpha}{5} + \frac{\alpha}{4} & \frac{1-\alpha}{5} & \frac{1-\alpha}{5} + \frac{\alpha}{2} & \frac{1-\alpha}{5} & \frac{1-\alpha}{5} \\ \frac{1-\alpha}{5} + \frac{\alpha}{4} & \frac{1-\alpha}{5} + \frac{\alpha}{3} & \frac{1-\alpha}{5} + \frac{\alpha}{2} & \frac{1-\alpha}{5} & \frac{1-\alpha}{5} \end{pmatrix}$$

4. El PageRank funciona sin sumideros

$p = (\mathcal{P}(p_1), \dots, \mathcal{P}(p_N))^t$ se describe por medio de una **matriz de transición** $T = (t_{ij})$ donde t_{ij} simboliza la probabilidad de alcanzar la página p_i desde la página p_j :



matriz de adyacencia $A = (a_{ij})$,

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} \end{pmatrix},$$

matriz de conectividad ponderada

$$C = \left(\frac{a_{ji}}{od(p_i)} \right) = \begin{pmatrix} 0 & \frac{1}{3} & 0 & 0 & \mathbf{\frac{1}{5}} \\ \frac{1}{4} & 0 & 0 & 0 & \mathbf{\frac{1}{5}} \\ \frac{1}{4} & \frac{1}{3} & 0 & 1 & \mathbf{\frac{1}{5}} \\ \frac{1}{4} & 0 & \frac{1}{2} & 0 & \mathbf{\frac{1}{5}} \\ \frac{1}{4} & \frac{1}{3} & \frac{1}{2} & 0 & \mathbf{\frac{1}{5}} \end{pmatrix}$$

matriz de transición

$$T = \frac{1-\alpha}{N}J + \alpha C, \quad t_{ij} = \begin{cases} \frac{1-\alpha}{N} + \frac{\alpha}{od(p_j)} & \text{si } p_j \longrightarrow p_i, \\ \frac{1-\alpha}{N} & \text{en otro caso} \end{cases}$$

$$T = \begin{pmatrix} \frac{1-\alpha}{5} & \frac{1-\alpha}{5} + \frac{\alpha}{3} & \frac{1-\alpha}{5} & \frac{1-\alpha}{5} & \frac{1-\alpha}{5} + \frac{\alpha}{5} \\ \frac{1-\alpha}{5} + \frac{\alpha}{4} & \frac{1-\alpha}{5} & \frac{1-\alpha}{5} & \frac{1-\alpha}{5} & \frac{1-\alpha}{5} + \frac{\alpha}{5} \\ \frac{1-\alpha}{5} + \frac{\alpha}{4} & \frac{1-\alpha}{5} + \frac{\alpha}{3} & \frac{1-\alpha}{5} & \frac{1-\alpha}{5} + \alpha & \frac{1-\alpha}{5} + \frac{\alpha}{5} \\ \frac{1-\alpha}{5} + \frac{\alpha}{4} & \frac{1-\alpha}{5} & \frac{1-\alpha}{5} + \frac{\alpha}{2} & \frac{1-\alpha}{5} & \frac{1-\alpha}{5} + \frac{\alpha}{5} \\ \frac{1-\alpha}{5} + \frac{\alpha}{4} & \frac{1-\alpha}{5} + \frac{\alpha}{3} & \frac{1-\alpha}{5} + \frac{\alpha}{2} & \frac{1-\alpha}{5} & \frac{1-\alpha}{5} + \frac{\alpha}{5} \end{pmatrix}, \quad \text{entonces } \boxed{Tp^1 = p^1}$$

El autovector p^1 de T es el 1-normalizado del vector p del digrafo original con sumidero en el vértice 5.

5. Formulación de Brinkmeier

Brinkmeier considera que el ranking de una página p_i depende de la contribución de cada página p_j en cada trayectoria $w : p_j \xrightarrow{*} p_i$ que conduce a p_i .

Para cada trayectoria $w = p_j p_2 \dots p_i$ se considera la contribución de w a p_i graduada por el **branching** de w :

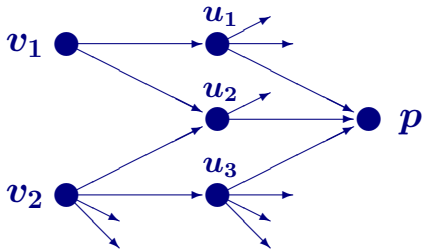
$$B(w) = \frac{1}{od(p_j)od(p_2) \cdots od(p_{i-1})}$$

Entonces, para cada página $p_i \in V$, se tiene

$$\mathcal{P}(p_i) = \frac{1 - \alpha}{N} \sum_{p_j \in V} \sum_{w : p_j \xrightarrow{*} p_i} \alpha^{l(w)} B(w) \quad (2)$$

$w : p_j \xrightarrow{*} p_i$ denota una trayectoria w desde la página p_j a la página p_i ,

$l(w)$ es la longitud de la trayectoria w .



$$\mathcal{P}(p) = \frac{1 - \alpha}{N} \left(1 + \alpha \left(\frac{1}{3} + \frac{1}{2} + \frac{1}{4} \right) + \alpha^2 \left(\frac{1}{2 * 3} + \frac{1}{2 * 2} + \frac{1}{4 * 2} + \frac{1}{4 * 4} \right) \right)$$

6. PageRank de árboles con raíz

$$\mathcal{P}(p_i) = \frac{1 - \alpha}{N} + \alpha \sum_{(p_j, p_i) \in A} \frac{\mathcal{P}(p_j)}{\text{od}(p_j)} = \frac{1 - \alpha}{N} + \alpha \sum_{(p_j, p_i) \in A} \mathcal{P}(p_j)$$

$$p_1 = \frac{1-\alpha}{14} + \alpha(p_2 + p_3)$$

$$p_2 = \frac{1-\alpha}{14} + \alpha(p_4 + p_5 + p_6)$$

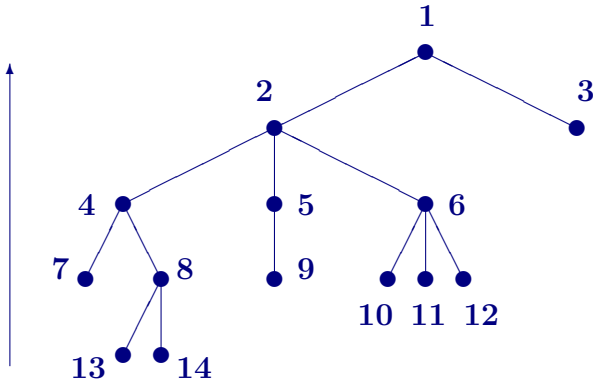
$$p_3 = \frac{1-\alpha}{14} = p_7 = p_9 = p_{10} = \dots = p_{14}$$

$$p_4 = \frac{1-\alpha}{14} + \alpha(p_3 + p_8)$$

$$p_5 = \frac{1-\alpha}{14} + \alpha(p_3)$$

$$p_6 = \frac{1-\alpha}{14} + \alpha(3p_3)$$

$$p_8 = \frac{1-\alpha}{14} + \alpha(2p_3)$$



$$\mathcal{P}(p_i) = \frac{1 - \alpha}{N} \sum_{p_j \in V} \sum_{w: p_j \xrightarrow{*} p_i} \alpha^{l(w)} B(w) = \frac{1 - \alpha}{N} \sum_{p_j \in V, p_j \xrightarrow{*} p_i} \alpha^{l(w)}$$

$$p_1 = \frac{1-\alpha}{14} (\alpha^0 + 2\alpha + 3\alpha^2 + 6\alpha^3 + 2\alpha^4)$$

$$p_2 = \frac{1-\alpha}{14} (1 + 3\alpha + 6\alpha^2 + 2\alpha^3)$$

$$p_3 = \frac{1-\alpha}{14} = p_7 = p_9 = p_{10} = \dots = p_{14}$$

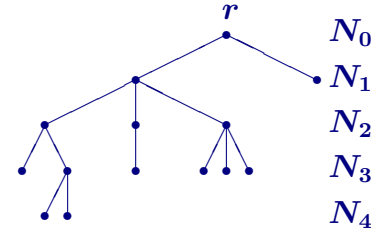
$$p_4 = \frac{1-\alpha}{14} (1 + 2\alpha + 2\alpha^2)$$

$$p_5 = \frac{1-\alpha}{14} (1 + \alpha)$$

$$p_6 = \frac{1-\alpha}{14} (1 + 3\alpha)$$

$$p_8 = \frac{0,15}{14} (1 + 2\alpha)$$

6. PageRank de árboles con raíz



\mathcal{T}^r es un árbol dirigido con raíz r .

La raíz r es la página objetivo para mejorar su PageRank.

altura de un vértice en \mathcal{T}^r es la longitud del camino desde el vértice a la raíz.

nivel N_k es el conjunto de vértices de altura k ; la raíz está en el nivel N_0 .

altura de \mathcal{T}^r es la longitud del camino más largo desde una hoja a la raíz.

Nota. Estamos interesados en estudiar el comportamiento del PageRank en ciertos subgrafos cerrados de la web. En particular, vemos un árbol con raíz como un sitio web local cerrado. Esto significa que el valor de N en la fórmula (2) es el número de vértices en el árbol.

Teorema. Si \mathcal{T}^r tiene N vértices y altura h , el PageRank de su raíz r es

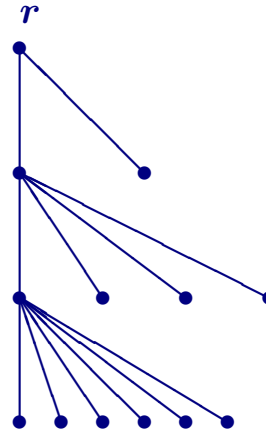
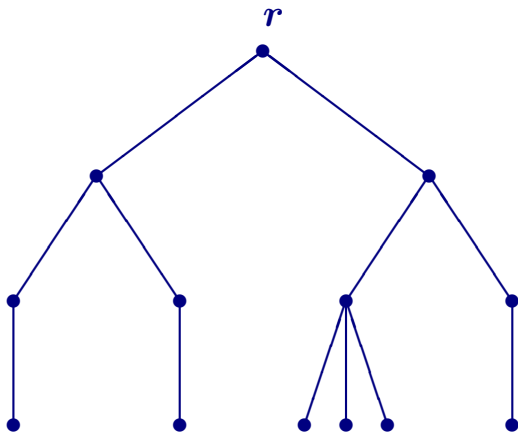
$$\mathcal{P}(r) = \frac{1 - \alpha}{N} \sum_{p_j \in V, p_j \xrightarrow{*} r} \alpha^{l(w)} = \frac{1 - \alpha}{N} \sum_{k=0}^h \alpha^k n_k \quad (3)$$

$n_k := |N_k|$ es el número de vértices del nivel N_k .

$\mathcal{P}(r)$ depende de la altura h , del número total de vértices N y de su distribución por niveles.

6. PageRank de árboles con raíz

Podemos reordenar los vínculos entre niveles consecutivos, sin que el PageRank de la raíz varíe.

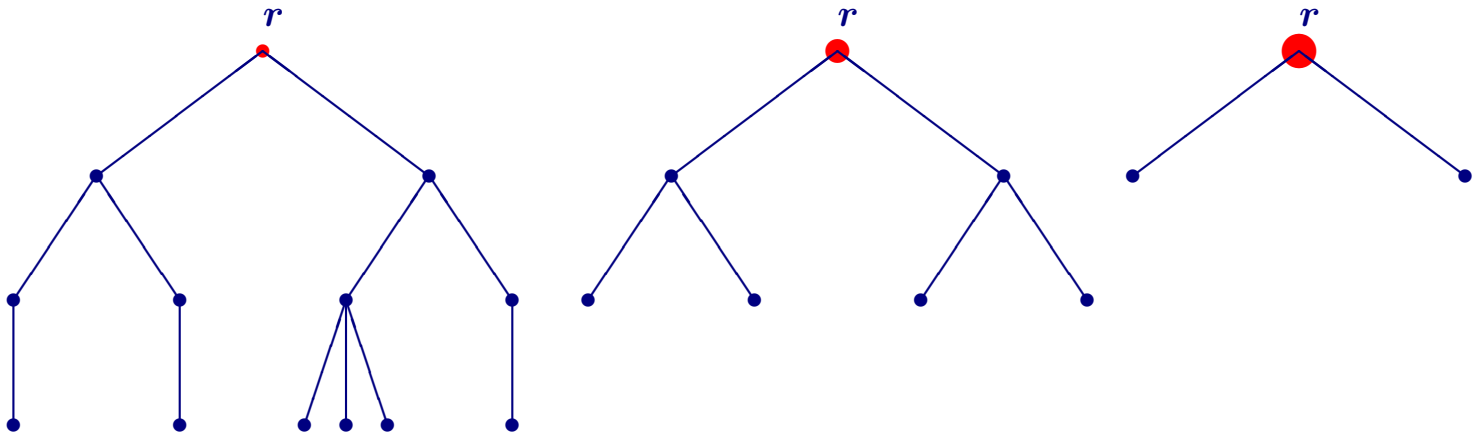


$\mathcal{T}^r = 1n_1 \dots n_h$ denota un árbol con raíz r de altura h con n_k vértices en el nivel N_k , $k = 1, \dots, h$.

7. Supresión de vértices. Árboles cola

Modificaciones en la estructura de un árbol para mejorar el PageRank de su raíz.

Teorema. Si en un árbol $\mathcal{T}^r = 1n_1 \dots n_h$ suprimimos el nivel N_h , entonces aumenta el PageRank de su raíz r .



Para un número de vértices fijo, la configuración óptima es una estrella con raíz.

7. Supresión de vértices. Árboles cola

Modificaciones en la estructura de un árbol para mejorar el PageRank de su raíz.

Teorema. Si en un árbol $\mathcal{T}^r = 1n_1 \dots n_h$ suprimimos el nivel N_h , entonces aumenta el PageRank de su raíz r .

Prueba. Al pasar del árbol $\mathcal{T}^r = 1n_1 \dots n_{h-1}n_h$ al árbol $\mathcal{T}'^r = 1n_1 \dots n_{h-1}$, se tiene

$$\mathcal{P}(r) = \frac{1 - \alpha}{N} \sum_{k=0}^h \alpha^k n_k, \quad \mathcal{P}'(r) = \frac{1 - \alpha}{N - n_h} \sum_{k=0}^{h-1} \alpha^k n_k$$

$$\begin{aligned} \mathcal{P}'(r) - \mathcal{P}(r) &= \frac{n_h(1 - \alpha)}{(N - n_h)N} (1 + n_1\alpha + \dots + n_{h-1}\alpha^{h-1} - (N - n_h)\alpha^h) \\ &= \frac{n_h(1 - \alpha)}{(N - n_h)N} (1 + n_1\alpha + \dots + n_{h-1}\alpha^{h-1} - (1 + n_1 + \dots + n_{h-1})\alpha^h) \\ &= \frac{n_h(1 - \alpha)}{(N - n_h)N} ((1 - \alpha^h) + n_1(\alpha - \alpha^h) + \dots + n_{h-1}(\alpha^{h-1} - \alpha^h)) > 0 \end{aligned}$$

ya que $0 < \alpha < 1$ and $h \geq 1$.

7. Supresión de vértices. Árboles cola

Modificaciones en la estructura de un árbol para mejorar el PageRank de su raíz.

Teorema. Si en un árbol $\mathcal{T}^r = 1n_1 \dots n_h$ suprimimos el nivel N_h , entonces aumenta el PageRank de su raíz r .

Consecuencia. Para mejorar el PageRank de la raíz de un árbol uno puede suprimir, de abajo a arriba, tantos niveles como lo permita el contexto.

Recíprocamente: si añadimos un nuevo nivel de vértices al árbol, entonces el PageRank de su raíz decrece.

Si estuviéramos obligados a conservar una determinada altura,

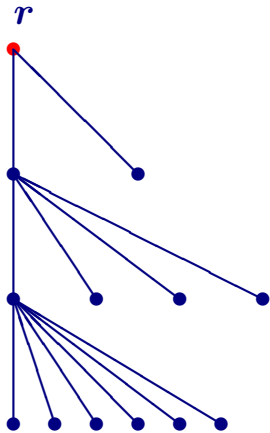
¿qué parte del árbol podemos suprimir mejorando su PageRank?

La situación extrema es suprimir todos los vértices de cada nivel, menos uno para conservar la altura, teniendo la estructura como cota.

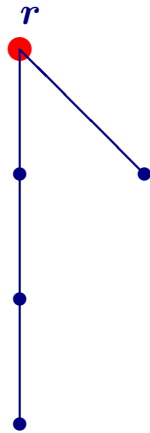
7. Supresión de vértices. Árboles cola

Definición. El **árbol cola** de un árbol $\mathcal{T}^r = 1n_1 \dots n_h$ es el árbol

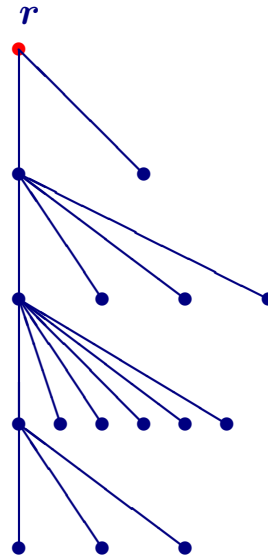
$$\mathcal{T}_q^r = 1n_1 \dots n_{\lfloor \frac{h-1}{2} \rfloor} \underbrace{1 \dots 1}_{\lfloor \frac{h}{2} \rfloor + 1}$$



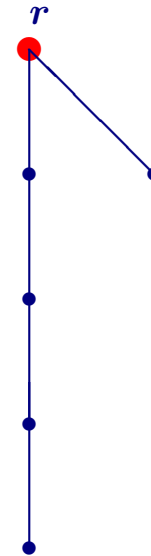
$$\mathcal{T}^r = 1246$$



$$\mathcal{T}_q^r = 1211$$



$$\mathcal{T}^r = 12463$$



$$\mathcal{T}_q^r = 12111$$

7. Supresión de vértices. Árboles cola

Teorema. El PageRank de la raíz de un árbol es menor que el PageRank de la raíz de su árbol cola.

El Teorema no puede mejorarse, en el sentido de que suprimir más vértices (conservando la altura) en un árbol cola puede o no mejorar el PageRank de la raíz.

Para valores pequeños de h , el árbol cola es la poda óptima para mejorar el PageRank.

Ejemplo. Si $h = 4$ el árbol cola es $\mathcal{T}_q = 1n_1111$ con PageRank $\mathcal{P}(r)$, y al suprimir un vértice de N_1 nos queda el árbol $\mathcal{T}' = 1(n_1 - 1)111$ con PageRank $\mathcal{P}'(r)$. Entonces

$$\mathcal{P}(r) = \frac{1 - \alpha}{(n_1 + 4)}(1 + n_1\alpha + \alpha^2 + \alpha^3 + \alpha^4), \quad \mathcal{P}'(r) = \frac{1 - \alpha}{(n_1 + 3)}(1 + (n_1 - 1)\alpha + \alpha^2 + \alpha^3 + \alpha^4)$$

$$\mathcal{P}'(r) - \mathcal{P}(r) = \frac{1 - \alpha}{(n_1 + 3)(n_1 + 4)}(1 - 4\alpha + \alpha^2 + \alpha^3 + \alpha^4) < 0, \quad \alpha \in (0, 27568, 1)$$

Si $h > 4$, una mejora del PageRank depende de α y de los cardinales $n_1, \dots, n_{\lfloor \frac{h-1}{2} \rfloor}$.

8. Problema: optimización de la estructura de vínculos

Objetivo principal: maximizar la función

$$\mathcal{P}(\alpha, h, n_1, \dots, n_h) = \frac{1 - \alpha}{1 + n_1 + \dots + n_h} \sum_{k=0}^h \alpha^k n_k$$

para $\alpha \in (0, 1)$ y $h, n_1, \dots, n_h \in \mathbb{N}$.

Si el número de vértices N está acotado, el máximo existe.

La complejidad es equivalente a la de hallar las particiones enteras positivas de N (con las condiciones estructurales impuestas por el contexto).

Aproximación heurística:

Regla 1: reducir la altura tanto como lo permita el contexto.

Regla 2: reordenar vínculos entre niveles consecutivos no modifica el PageRank.

Regla 3: Una vez fijada la altura óptima¹, suprimir vértices (tantos como sea posible) en la mitad inferior del árbol, aproximándose al árbol cola subyacente.

Regla 4: acercar los vértices no suprimibles, tanto como sea posible, al nivel N_1 .

¹La optimalidad depende del contexto: la altura significa el mínimo número de niveles necesarios para reflejar el sitio web; por ejemplo, una corporación o un hipertexto.

9. PageRank de digrafos acíclicos

Los digrafos acíclicos son estructuras de orden antisimétricas, no necesariamente transitivas, intermedias entre diagramas de Hasse y órdenes parciales.

Las estructuras más amplias en las que se puede asegurar la existencia de maximales y minimales y costituyen la clase más amplia de digrafos cuyos vértices se pueden estratificar en niveles.

Un vértice v es **maximal** si $exgr(v) = 0$ (**minimal** si $ingr(v) = 0$).

Lema. En un digrafo acíclico hay, al menos, un vértice maximal (minimal).

En un digrafo acíclico (V, A) se pueden definir los siguientes **niveles**:

$$N_0 = \{v \in V : v \text{ es minimal de } (V, A)\},$$

$$N_1 = \{v \in V - N_0 : v \text{ es minimal de } (V - N_0, A/(V - N_0))\}$$

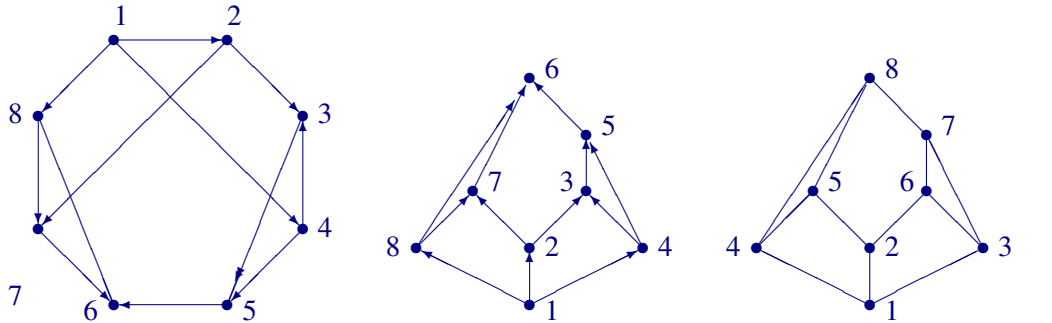
$$N_2 = \{v \in V - (N_0 \cup N_1) : v \text{ es minimal de } (V - (N_0 \cup N_1), A/(V - (N_0 \cup N_1)))\} \dots$$

y, en general, para todo p , con $0 < p \leq n$:

$$N_p = \{v \in V - \bigcup_{r=0}^{p-1} N_r : v \text{ es minimal de } (V - \bigcup_{r=0}^{p-1} N_r, A/(V - \bigcup_{r=0}^{p-1} N_r))\}.$$

Teorema. Un digrafo (V, A) es acíclico $\iff V$ se puede clasificar en niveles.

9. PageRank de digrafos acíclicos



Se caracterizan, además, por admitir etiquetados **naturalmente ordenados**:

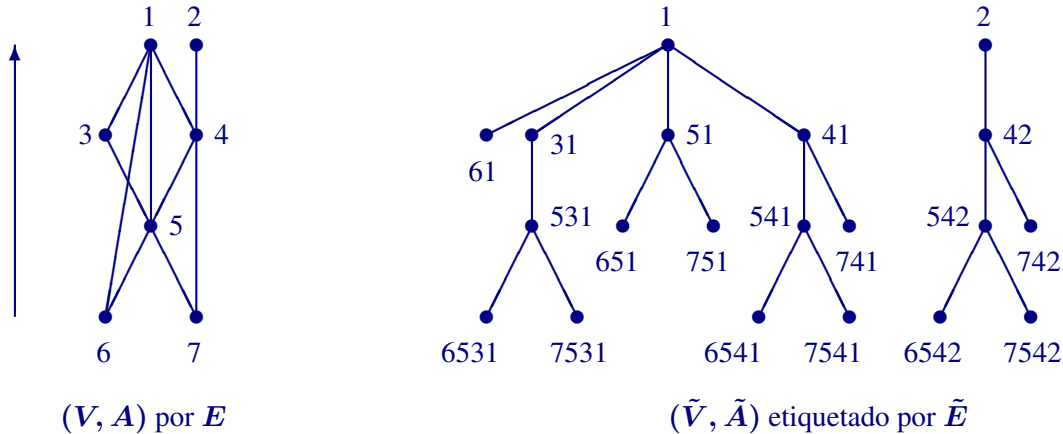
Teorema. Un digrafo (V, A) de orden n es acíclico $\iff V$ se puede etiquetar con $\{1, 2, \dots, n\}$ de modo que si $(i, j) \in A \implies i < j$.

Demostración. \implies) Basta asignar etiquetas respetando el orden de los niveles:

$$\text{Si } n_r = \text{Card}(N_r), \text{ entonces } m \in N_p \iff \sum_{r=0}^{p-1} n_r < m \leq \sum_{r=0}^p n_r$$

\iff) Si (V, A) tiene un ciclo $k_1 k_2 \cdots k_p k_1$ y $k_i = \max\{k_1, k_2, \dots, k_p\}$, entonces $(k_i, k_{i+1}) \in A$ con $k_i > k_{i+1}$.

9. PageRank de digrafos acíclicos



(V, A) un digrafo acíclico etiquetado por $E = \{1, \dots, N\}$; $M = \{\text{maximales}\}$

(\tilde{V}, \tilde{A}) es el **bosque de caminos** de (V, A) etiquetado por $\tilde{E} = \{\text{caminos con fin maximal}\}$,

$\tilde{V} = \{\text{caminos con fin maximal de } (V, A)\} \cup M$

$\tilde{A} = \{(p, q) \in \tilde{V} \times \tilde{V} : q = p - \text{el primer elemento, } p \notin M\}$.

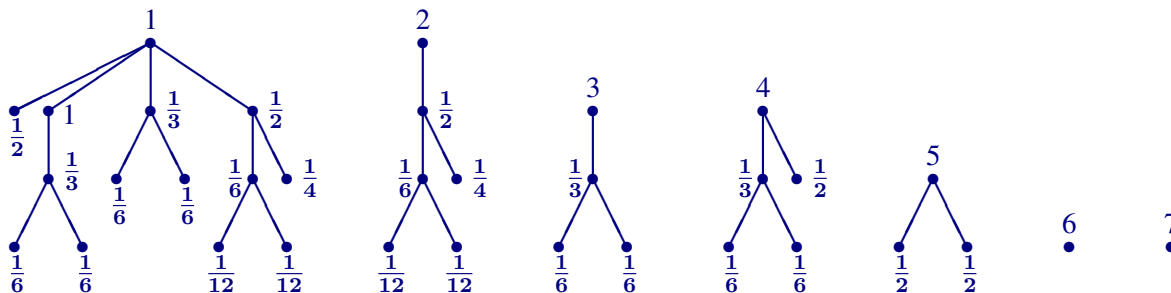
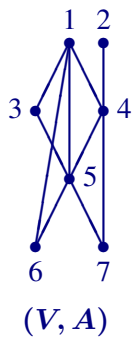
(\tilde{V}, \tilde{A}) es un bosque con $Card(M)$ árboles con raíz.

Cada vértice $v \in V$ da lugar a un nuevo vértice en \tilde{V} por cada uno de los caminos con fin maximal en (V, A) que empiezan en v .

9. PageRank de digrafos acíclicos

Teorema. Sea (V, A) un digrafo acíclico, (\tilde{V}, \tilde{A}) su bosque de caminos, $i \in V$.

Entonces $\mathcal{P}_i = \frac{1 - \alpha}{N} \sum_{k=0}^{h_i} b_{i_k} \alpha^k$, donde b_{i_k} es la suma de los branching de los vértices de nivel N_{i_k} en el subárbol con raíz i y altura h_i del bosque de caminos (\tilde{V}, \tilde{A}) .

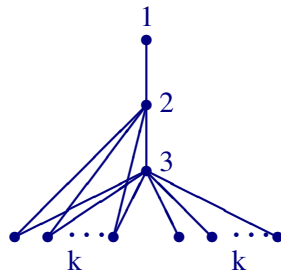


PageRank:
$$p = \frac{1 - \alpha}{7} \left(1 + \frac{7}{3}\alpha + \frac{13}{12}\alpha^2 + \frac{1}{2}\alpha^3, 1 + \frac{1}{2}\alpha + \frac{5}{12}\alpha^2 + \frac{1}{6}\alpha^3, \right. \\ \left. 1 + \frac{1}{3}\alpha + \frac{1}{3}\alpha^2, 1 + \frac{5}{6}\alpha + \frac{1}{3}\alpha^2, 1 + \alpha, 1, 1 \right)$$

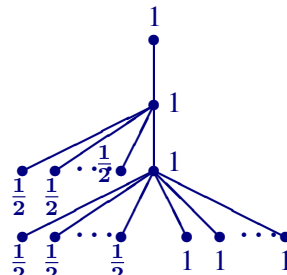
9. PageRank de digrafos acíclicos

Teorema. 1. La supresión de las hojas del último nivel de un digrafo acíclico aumenta el PageRank de sus correspondientes vértices maximales.

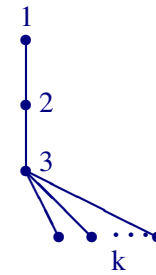
2. Sin embargo, la supresión de vértices con exgrado mayor que 1 del último nivel de un digrafo acíclico puede aumentar o disminuir el PageRank de su raíz.



digrafo acíclico (V, A)



bosque de caminos (\tilde{V}, \tilde{A})



$(V, A) - \{v_i \in N_0 : od(v_i) > 1\}$

$$\mathcal{P}_1 = \frac{1 - \alpha}{2k + 3} \left(1 + \alpha + \left(\frac{k}{2} + 1 \right) \alpha^2 + \frac{3k}{2} \alpha^3 \right), \quad \mathcal{P}_2 = \frac{1 - \alpha}{2k + 3} \left(1 + \left(\frac{k}{2} + 1 \right) \alpha + \frac{3k}{2} \alpha^2 \right),$$

$$\mathcal{P}_3 = \frac{1 - \alpha}{2k + 3} \left(1 + \frac{3k}{2} \alpha \right), \quad \text{para } i = 4, \dots, 2k + 3, \quad \mathcal{P}_i = \frac{1 - \alpha}{2k + 3}.$$

$$\sum_{i=1}^{2k+3} \mathcal{P}_i = \frac{1 - \alpha}{2k + 3} \left(2k + 3 + (2k + 2)\alpha + (2k + 1)\alpha^2 + \frac{3k}{2}\alpha^3 \right),$$

para el digrafo sin los k vértices del nivel N_0 con exgrado 2 se tiene

$$\mathcal{P}'_1 = \frac{1 - \alpha}{k + 3}(1 + \alpha + \alpha^2 + k\alpha^3), \quad \mathcal{P}'_2 = \frac{1 - \alpha}{k + 3}(1 + \alpha + k\alpha^2),$$

$$\mathcal{P}'_3 = \frac{1 - \alpha}{k + 3}(1 + k\alpha), \quad \text{para } i = 4, \dots, k + 3, \quad \mathcal{P}'_i = \frac{1 - \alpha}{k + 3}.$$

$$\sum_{i=1}^{k+3} \mathcal{P}'_i = \frac{1 - \alpha}{k + 3}(k + 3 + (k + 2)\alpha + (k + 1)\alpha^2 + k\alpha^3)$$

En términos relativos, se tiene

$$\frac{\mathcal{P}'_1}{\sum_{i=1}^{k+3} \mathcal{P}'_i} \geq \frac{\mathcal{P}_1}{\sum_{i=1}^{2k+3} \mathcal{P}_i} \iff$$

$$\frac{(1 + \alpha + \alpha^2 + k\alpha^3)}{(k + 3 + (k + 2)\alpha + (k + 1)\alpha^2 + k\alpha^3)} \geq \frac{(1 + \alpha + (\frac{k}{2} + 1)\alpha^2 + \frac{3k}{2}\alpha^3)}{(2k + 3 + (2k + 2)\alpha + (2k + 1)\alpha^2 + \frac{3k}{2}\alpha^3)}$$

$$\iff (k - 3)\alpha^2 - 4\alpha - 2 \leq 0$$

Por tanto,

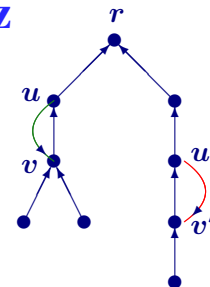
$$\left\{ \begin{array}{l} \frac{\mathcal{P}'_1}{\sum_{i=1}^{k+3} \mathcal{P}'_i} \geq \frac{\mathcal{P}_1}{\sum_{i=1}^{2k+3} \mathcal{P}_i} \quad \text{para } k \leq 9, \alpha \in (0, 1), \text{ y} \\ \frac{\mathcal{P}'_1}{\sum_{i=1}^{k+3} \mathcal{P}'_i} \leq \frac{\mathcal{P}_1}{\sum_{i=1}^{2k+3} \mathcal{P}_i} \quad \text{para } k > 9, \alpha \in \left(\frac{2 + \sqrt{2k - 2}}{k - 3}, 1\right) \end{array} \right.$$

10. PageRank de árboles bidireccionales con raíz

\mathcal{B}^r denota un **árbol bidireccional con raíz r** :

$\mathcal{B}^r = (V, A)$ es un digrafo con A partido en dos $A = A_1 \cup A_2$ tal que:

- (V, A_1) es un árbol parcial con raíz r (el árbol subyacente de \mathcal{B}^r)
- si $uv \in A_2 \implies vu \in A_1$.



Árbol bidireccional

Cada arco bidireccional uvu proporciona una **infinitad de trayectorias** desde u a la raíz r : $u \dots r$, $uvu \dots r$, $uvuvu \dots r$, ...

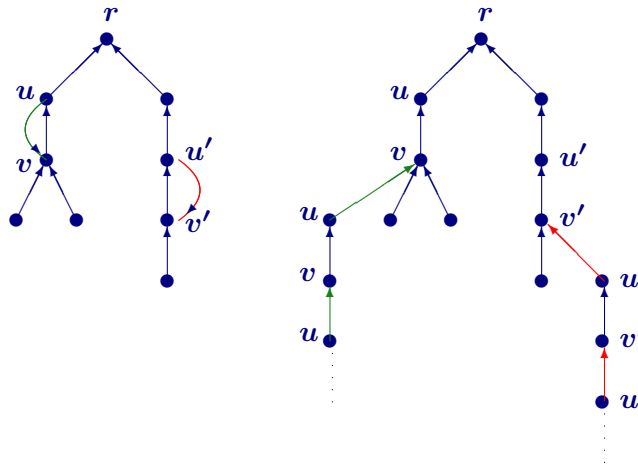
Para calcular el PageRank de r con la formula de Brinkmeier,

$$\mathcal{P}(r) = \frac{1 - \alpha}{N} \sum_{p_j \in V} \sum_{w: p_j \xrightarrow{*} r} \alpha^{l(w)} B(w)$$

podemos ver **cada arco bidireccional uvu** como **un camino infinito** colgando del vértice u .

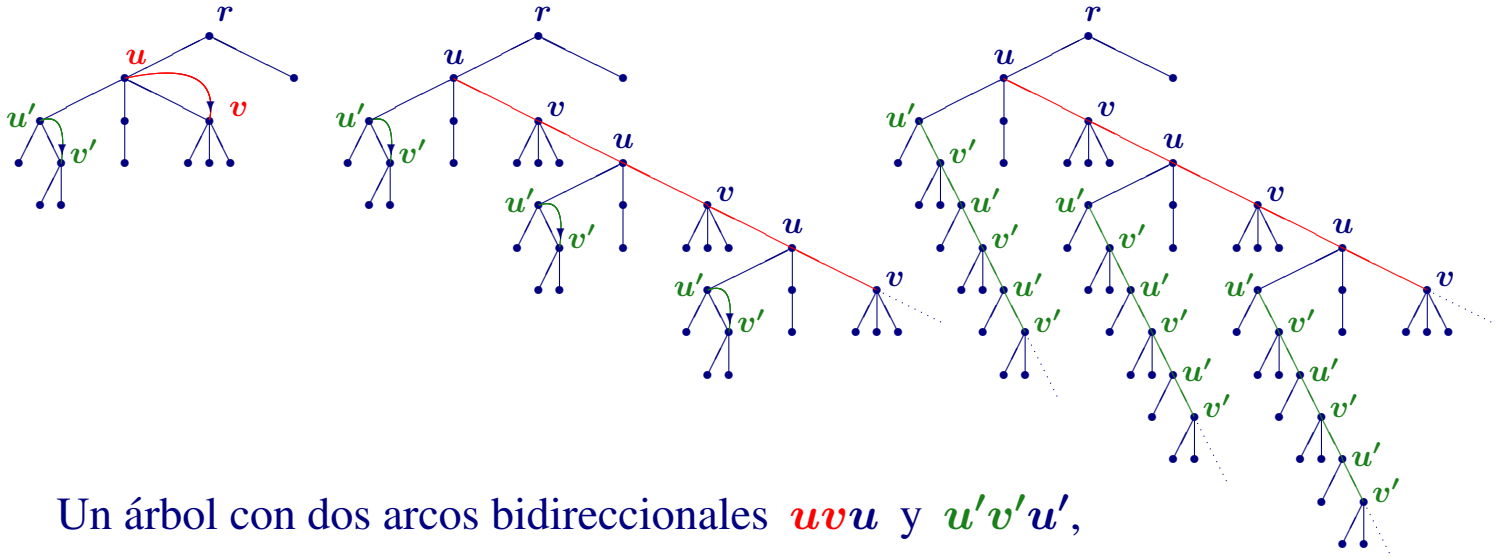
Extendiendo esta idea a todos los arcos bidireccionales, podemos ver \mathcal{B}^r como **un árbol con raíz infinito**.

10. PageRank de árboles bidireccionales con raíz



Árbol bidireccional y su árbol infinito asociado

10. PageRank de árboles bidireccionales con raíz



Un árbol con dos arcos bidireccionales uvu y $u'v'u'$,
el árbol bidireccional con una **rama infinita** asociada al arco bidireccional uvu
y el árbol unidireccional infinito asociado.

10. PageRank de árboles bidireccionales con raíz

$$\text{Brin-Page } \mathcal{P}(p_i) = \frac{1 - \alpha}{N} + \alpha \sum_{(p_j, p_i) \in A} \frac{\mathcal{P}(p_j)}{\text{od}(p_j)}$$

$$p_1 = \frac{1 - \alpha}{14} + \alpha \left(\frac{p_2}{2} + p_3 \right)$$

$$p_2 = \frac{1 - \alpha}{14} + \alpha(p_4 + p_5 + p_6)$$

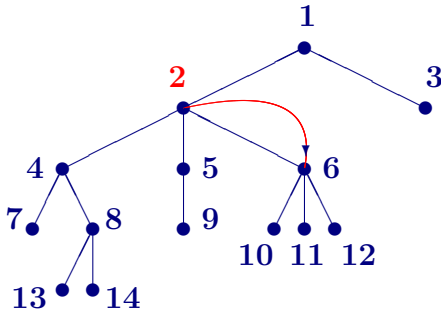
$$p_3 = \frac{1 - \alpha}{14} = p_7 = p_9 = p_{10} = p_{11} = p_{12} = p_{13} = p_{14}$$

$$p_4 = \frac{1 - \alpha}{14} + \alpha(p_3 + p_8)$$

$$p_5 = \frac{1 - \alpha}{14} + \alpha(p_3)$$

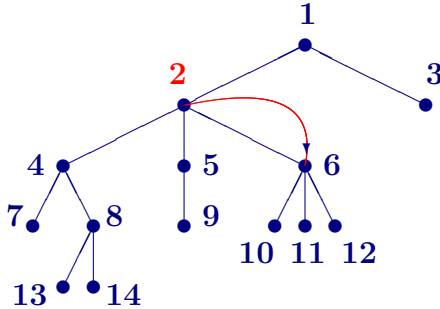
$$p_6 = \frac{1 - \alpha}{14} + \alpha \left(\frac{p_2}{2} + 3p_3 \right)$$

$$p_8 = \frac{1 - \alpha}{14} + \alpha(2p_3)$$



10. PageRank de árboles bidireccionales con raíz

$$\text{Brinkmeier } \mathcal{P}(p_i) = \frac{1-\alpha}{N} \sum_{p_j \in V} \sum_{w: p_j \xrightarrow{*} p_i} \alpha^{l(w)} B(w), \quad B(2) = \frac{1}{2}$$



$$p_1 \longrightarrow 1 + (21, 2621, 262621, \dots) + 31 + 3(421, 42621, \dots) + 6(7421, 742621, \dots) + 2(138421, 13842621, \dots)$$

$$p_2 \longrightarrow (2, 262, \dots) + 3(42, 4262, \dots) + 6(742, 74262, \dots) + 2(13842, 1384262, \dots)$$

$$p_6 \longrightarrow (6, 626, \dots) + 3(106, 10626, \dots) + (26, 2626, \dots) + 2(426, 42626, \dots) + 3(7426, 742626, \dots) + 2(138426, 13842626, \dots)$$

$$p_1 = \frac{1-\alpha}{14} (1 + (\frac{\alpha}{2} + \frac{\alpha^3}{4} + \dots) + \alpha + 3(\frac{\alpha^2}{2} + \frac{\alpha^4}{4} + \dots) + 6(\frac{\alpha^3}{2} + \frac{\alpha^5}{4} + \dots) + 2(\frac{\alpha^4}{2} + \frac{\alpha^6}{4} + \dots))$$

$$p_2 = \frac{1-\alpha}{14} ((1 + \frac{\alpha^2}{2} + \frac{\alpha^4}{4} + \dots) + 3(\alpha + \frac{\alpha^3}{2} + \dots) + 6(\alpha^2 + \frac{\alpha^4}{2} + \dots) + 2(\alpha^3 + \frac{\alpha^5}{2} + \dots))$$

$$p_3 = \frac{1-\alpha}{14} = p_7 = p_9 = p_{10} = \dots = p_{14}$$

$$p_4 = \frac{1-\alpha}{14} (1 + 2\alpha + 2\alpha^2), \quad p_5 = \frac{1-\alpha}{14} (1 + \alpha), \quad p_8 = \frac{0,15}{14} (1 + 2\alpha)$$

$$p_6 = \frac{1-\alpha}{14} ((1 + \frac{\alpha^2}{2} + \dots) + 3(\alpha + \frac{\alpha^3}{2} + \dots) + (\frac{\alpha}{2} + \frac{\alpha^3}{2} + \dots) + 2(\frac{\alpha^2}{2} + \frac{\alpha^4}{4} + \dots) + 3(\frac{\alpha^3}{2} + \frac{\alpha^5}{4} + \dots) + 2(\frac{\alpha^4}{2} + \frac{\alpha^6}{4} + \dots))$$

10. PageRank de árboles bidireccionales con raíz

$$\mathcal{P}(p_i) = \frac{1 - \alpha}{N} \sum_{p_j \in V} \sum_{w: p_j \xrightarrow{*} p_i} \alpha^{l(w)} B(w)$$

Para cada vértice p_j conectado por trayectorias $w : p_j \xrightarrow{*} p_i$ con el vértice p_i , en realidad existe un único camino $c : p_j \xrightarrow{*} p_i$.

Este camino $c : p_j \xrightarrow{*} p_i$ puede ser incidente con varios arcos bidireccionales.

Cada arco bidireccional uxu , con $od(u) = 2$, genera infinitas trayectorias

$$u, \quad uxu, \quad uxuxu, \dots$$

con branchings $B(u) = 1, B(uxu) = \frac{1}{2}, B(uxuxu) = \frac{1}{2^2}, \dots$

$$\sum_{w: u \xrightarrow{*} u} \alpha^{l(w)} B(w) = 1 + \frac{\alpha^2}{2} + \frac{\alpha^4}{2^2} + \dots = \frac{1}{1 - \alpha^2/2}$$

Si en el camino $c : p_j \xrightarrow{*} p_i$ hay q vértices distintos en los que inciden q arcos bidireccionales, la aportación de todas las trayectorias es $\frac{1}{(1 - \alpha^2/2)^q}$.

Si el arco bidireccional es rxr , $od(r) = 1$, $B(rxr \dots xr) = 1$ y la aportación es $\frac{1}{1 - \alpha^2}$.

10. PageRank de árboles bidireccionales con raíz

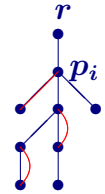
En la fórmula $\mathcal{P}(p_i) = \frac{1 - \alpha}{N} \sum_{p_j \in V} \sum_{w: p_j \xrightarrow{*} p_i} \alpha^{l(w)} B(w)$

para cada vértice $p_j \in V$ basta considerar el camino único $c : p_j \xrightarrow{*} p_i$ cuya aportación depende de los siguientes datos: $l(c)$, su longitud,

$n :=$ número de vértices bidireccionales (con $od(u) = 2$) no finales en c y

$q :=$ número de arcos bidireccionales incidentes en c . Se tiene

1) \mathcal{B}^r con $od(r) = 0 \implies \mathcal{P}(p_i) = \frac{1 - \alpha}{N} \sum_{p_j \xrightarrow{*} p_i} \frac{\alpha^{l(c)}}{2^n (1 - \alpha^2/2)^q}, p_i \in V$



2) \mathcal{B}^r con $rur, od(r) = 1 \implies \mathcal{P}(p_i) = \frac{1 - \alpha}{N} \sum_{p_j \xrightarrow{*} p_i} \frac{\alpha^{l(c)}}{2^n (1 - \alpha^2/2)^q}, p_i \notin \{r, u\}$



$$\mathcal{P}(p_i) = \frac{1 - \alpha}{N} \sum_{p_j \xrightarrow{*} p_i} \frac{\alpha^{l(c)}}{2^n (1 - \alpha^2/2)^{q-1} (1 - \alpha^2)}, p_i \in \{r, u\}$$

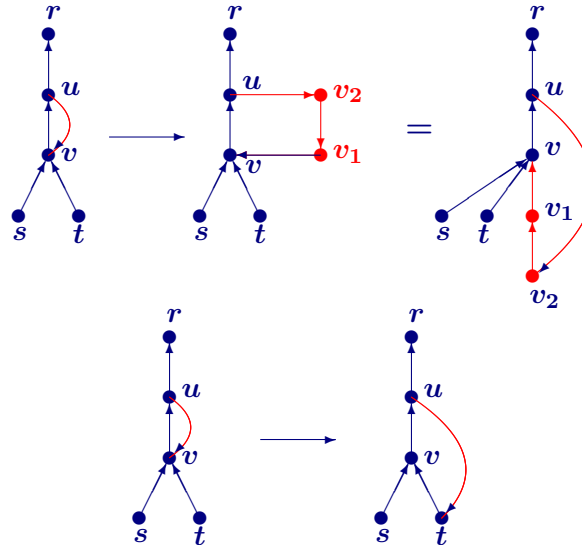
En particular

1) si $od(r) = 0, n = q \implies \mathcal{P}(r) = \frac{1 - \alpha}{N} \sum_{p_j \xrightarrow{*} r} \frac{\alpha^{l(c)}}{(2 - \alpha^2)^q}$.

2) si $od(r) = 1, n = q - 1 \implies \mathcal{P}(r) = \frac{1 - \alpha}{N} \sum_{p_j \xrightarrow{*} r} \frac{\alpha^{l(c)}}{(2 - \alpha^2)^{q-1} (1 - \alpha^2)}$.

11. PageRank de árboles cíclicos con raíz

Ahora admitimos ciclos de longitud mayor cerrando caminos de longitud arbitraria.



\mathcal{C}^r denota un árbol cíclico con raíz r :

$\mathcal{C}^r = (V, A)$ es un digrafo con A partido en dos $A = A_1 \cup A_2$ tal que:

- (V, A_1) es un árbol parcial con raíz r (el árbol subyacente de \mathcal{C}^r) y
- si $uv \in A_2$, hay un camino $v \cdots u$, y u es el origen del ciclo $uv \cdots u$.

11. PageRank de árboles cíclicos con raíz

Cada ciclo $ux \dots u$, de longitud l con $od(u) = 2$, genera una infinidad de trayectorias $u, ux \dots u, ux \dots ux \dots u, \dots$ con branchings $B(u) = 1, B(ux \dots u) = \frac{1}{2}, B(ux \dots ux \dots u) = \frac{1}{2^2}, \dots$

$$\sum_{w: u \xrightarrow{*} u} \alpha^{l(w)} B(w) = 1 + \frac{\alpha^l}{2} + \frac{\alpha^{2l}}{2^2} + \dots = \frac{1}{1 - \alpha^l/2}$$

Si en el camino $c : v \xrightarrow{*} a$ hay q vértices distintos en los que inciden q ciclos de longitudes l_1, l_2, \dots, l_q , la aportación de todas las trayectorias es

$$\frac{1}{1 - \alpha^{l_1}/2} \frac{1}{1 - \alpha^{l_2}/2} \dots \frac{1}{1 - \alpha^{l_q}/2}$$

Si el ciclo es $rx \dots r$, con $od(r) = 1$, $B(rx \dots rx \dots r \dots) = 1$ y la aportación es $\frac{1}{1 - \alpha^l}$.

11. PageRank de árboles cíclicos con raíz

En la fórmula $\mathcal{P}(p_i) = \frac{1 - \alpha}{N} \sum_{p_j \in V} \sum_{w: p_j \xrightarrow{*} p_i} \alpha^{l(w)} B(w)$

para cada vértice $p_j \in V$ basta considerar el camino único $c : p_j \xrightarrow{*} p_i$ cuya aportación depende de los siguientes datos: $l(c)$, su longitud,

$n :=$ número de vértices bidireccionales (con $od(u) = 2$) no finales en c y

$q :=$ número de ciclos incidentes en c de longitudes l_1, l_2, \dots, l_q . Se tiene

1) si $od(r) = 0 \implies \mathcal{P}(p_i) = \frac{1 - \alpha}{N} \sum_{p_j \xrightarrow{*} p_i} \frac{\alpha^{l(c)}}{2^n (1 - \alpha^{l_1}/2) \dots (1 - \alpha^{l_q}/2)}, p_i \in V$

2) si $rp_1 \dots p_{l_{q-1}} r, od(r) = 1 \implies \mathcal{P}(p_i) = \frac{1 - \alpha}{N} \sum_{p_j \xrightarrow{*} p_i} \frac{\alpha^{l(c)}}{2^n (1 - \alpha^{l_1}/2) \dots (1 - \alpha^{l_q}/2)},$
 $p_i \notin \{r, p_1, \dots, p_{l_{q-1}}\}$

$\mathcal{P}(p_i) = \frac{1 - \alpha}{N} \sum_{p_j \xrightarrow{*} p_i} \frac{\alpha^{l(c)}}{2^n (1 - \alpha^{l_1}/2) \dots (1 - \alpha^{l_{q-1}}/2) (1 - \alpha^{l_q})}, p_i \in \{r, p_1, \dots, p_{l_{q-1}}\}$

En particular

1) si $od(r) = 0, n = q \implies \mathcal{P}(r) = \frac{1 - \alpha}{N} \sum_{p_j \xrightarrow{*} r} \frac{\alpha^{l(c)}}{(2 - \alpha^{l_1}) \dots (2 - \alpha^{l_q})}$

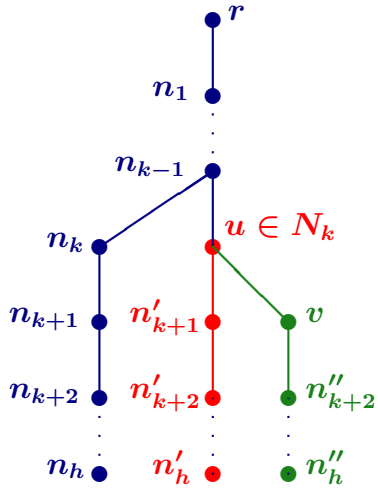
2) si $od(r) = 1, n = q - 1 \implies \mathcal{P}(r) = \frac{1 - \alpha}{N} \sum_{p_j \xrightarrow{*} r} \frac{\alpha^{l(c)}}{(2 - \alpha^{l_1}) \dots (2 - \alpha^{l_{q-1}}) (1 - \alpha^{l_q})}$

12. Manipulaciones en la topología

Teorema. Si en un árbol \mathcal{T}^r transformamos un arco vu , con $u \neq r$, en un arco bidireccional uvu , entonces $\mathcal{P}(u)$ y $\mathcal{P}(v)$ aumentan; $\mathcal{P}(r)$ disminuye.

$\mathcal{P}_x(\mathcal{T}^y)$ denota el PageRank del vértice x en el árbol \mathcal{T}^y , con raíz y .

$n_p(\mathcal{T}^y)$ denota el número de vértices en el nivel N_p del árbol \mathcal{T}^y . Se tiene

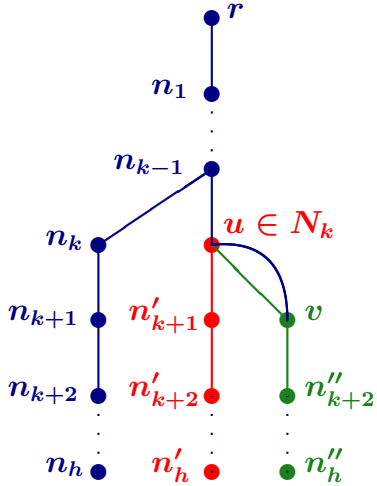


12. Manipulaciones en la topología

Teorema. Si en un árbol \mathcal{T}^r transformamos un arco vu , con $u \neq r$, en un arco bidireccional uvu , entonces $\mathcal{P}(u)$ y $\mathcal{P}(v)$ aumentan; $\mathcal{P}(r)$ disminuye.

$\mathcal{P}_x(\mathcal{T}^y)$ denota el PageRank del vértice x en el árbol \mathcal{T}^y , con raíz y .

$n_p(\mathcal{T}^y)$ denota el número de vértices en el nivel N_p del árbol \mathcal{T}^y . Se tiene



$$\mathcal{P}_r(\mathcal{T}^r) = \frac{1-\alpha}{N} \sum_{p=0}^h n_p(\mathcal{T}^r) \alpha^p$$

$$= \frac{1-\alpha}{N} \left(\sum_{p=0}^k n_p(\mathcal{T}^r - \mathcal{T}^u) \alpha^p + \sum_{p=k}^h n_p(\mathcal{T}^u) \alpha^p \right)$$

$$\mathcal{P}_r(\mathcal{B}^r) = \frac{1-\alpha}{N} \left(\sum_{p=0}^k n_p(\mathcal{T}^r - \mathcal{T}^u) \alpha^p + \frac{1}{2-\alpha^2} \sum_{p=k}^h n_p(\mathcal{T}^u) \alpha^p \right) < \mathcal{P}_r(\mathcal{T}^r)$$

$$\mathcal{P}_u(\mathcal{B}^u) = \frac{1-\alpha}{N(1-\alpha^2/2)} \sum_{p=k}^h n_p(\mathcal{T}^u) \alpha^{p-k} = \frac{1}{1-\alpha^2/2} \mathcal{P}_u(\mathcal{T}^r) > \mathcal{P}_u(\mathcal{T}^r)$$

$$\mathcal{P}_v(\mathcal{B}^v) = \frac{1-\alpha}{N(1-\alpha^2/2)} \left(\frac{\alpha}{2} \sum_{p=k}^h n_p(\mathcal{T}^u - \mathcal{T}^v) \alpha^{p-k} + \sum_{p=k+1}^h n_p(\mathcal{T}^v) \alpha^{p-(k+1)} \right) > \mathcal{P}_v(\mathcal{T}^v)$$

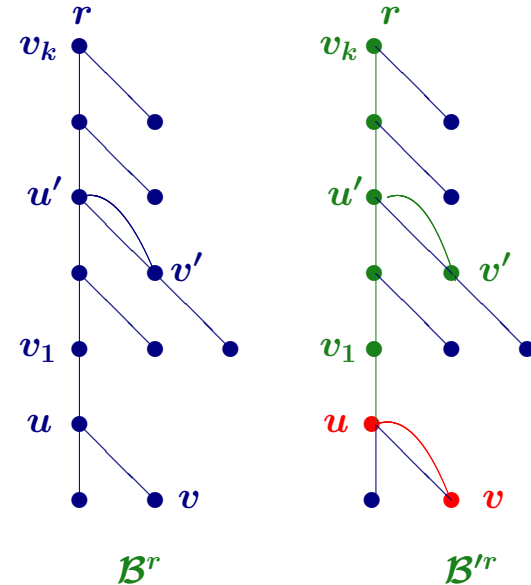
$$\mathcal{P}_u(\mathcal{B}^u) = \frac{1}{1-\alpha^2/2} (\mathcal{P}_u(\mathcal{T}^u - \mathcal{T}^v) + \alpha \mathcal{P}_v(\mathcal{T}^v))$$

$$\mathcal{P}_v(\mathcal{B}^v) = \frac{1}{1-\alpha^2/2} \left(\frac{\alpha}{2} \mathcal{P}_u(\mathcal{T}^u - \mathcal{T}^v) + \mathcal{P}_v(\mathcal{T}^v) \right)$$

12. Manipulaciones en la topología

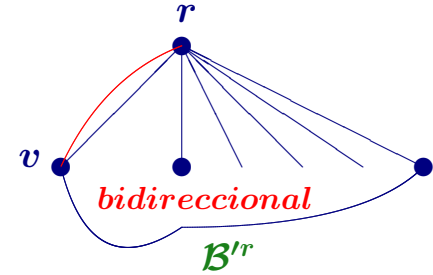
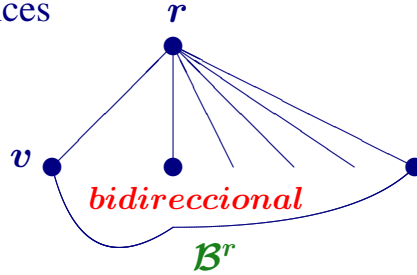
Teorema. Si en un árbol bidireccional \mathcal{B}^r transformamos un arco vu , $u \neq r$, en un arco bidireccional uvu , entonces

- 1) $\mathcal{P}_u(\mathcal{B}'^u) = \frac{1}{1 - \alpha^2/2} \mathcal{P}_u(\mathcal{B}^u) > \mathcal{P}_u(\mathcal{B}^u)$
- 2) $\mathcal{P}_v(\mathcal{B}'^u) > \mathcal{P}_v(\mathcal{B}^u)$
- 3) Si $u'v'u'$ es un arco bidireccional incidente en el camino $uv_1 \dots v_k = r$, entonces $\mathcal{P}_{u'}(\mathcal{B}'^r) < \mathcal{P}_{u'}(\mathcal{B}^r)$ y $\mathcal{P}_{v'}(\mathcal{B}'^r) < \mathcal{P}_{v'}(\mathcal{B}^r)$
- 4) $\mathcal{P}_{v_i}(\mathcal{B}'^r) < \mathcal{P}_{v_i}(\mathcal{B}^r) \quad \forall v_i$ del camino $v_1 \dots v_k$
- 5) En particular $\mathcal{P}_r(\mathcal{B}'^r) < \mathcal{P}_r(\mathcal{B}^r)$
- 6) Los vértices exteriores al camino $uv_1 \dots v_k = r$ y a sus arcos bidireccionales incidentes **conservan el PageRank.**



12. Manipulaciones en la topología

Teorema. Si en un árbol bidireccional \mathcal{B}^r , con $od(r) = 0$, transformamos un arco vr en un arco bidireccional rvr , entonces



$$1) \mathcal{P}_r(\mathcal{B}'^r) = \frac{\mathcal{P}_r(\mathcal{B}^r)}{1 - \alpha^2} \approx 3,6\mathcal{P}_r(\mathcal{B}^r) = \mathcal{P}_r(\mathcal{B}^r) + 2,6\mathcal{P}_r(\mathcal{B}^r)$$

$$\mathcal{P}_r(\mathcal{B}^r) = \alpha\mathcal{P}_v(\mathcal{B}^r) + \mathcal{P}_r(\mathcal{B}^r - \mathcal{B}^v)$$

$$2) \mathcal{P}_v(\mathcal{B}'^r) = \frac{\mathcal{P}_v(\mathcal{B}^r)}{1 - \alpha^2} + \alpha\mathcal{P}_r(\mathcal{B}'^r - \mathcal{B}'^v) = \frac{\mathcal{P}_v(\mathcal{B}^r)}{1 - \alpha^2} + \frac{\alpha(\mathcal{P}_r(\mathcal{B}^r) - \alpha\mathcal{P}_v(\mathcal{B}^r))}{1 - \alpha^2}$$

$$= \mathcal{P}_v(\mathcal{B}^r) + \frac{\alpha\mathcal{P}_r(\mathcal{B}^r)}{1 - \alpha^2} \approx \mathcal{P}_v(\mathcal{B}^r) + 3,06\mathcal{P}_r(\mathcal{B}^r)$$

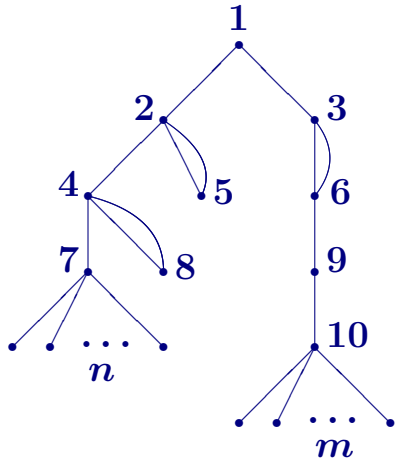
$$3) \mathcal{P}_r(\mathcal{B}'^r) \geq \mathcal{P}_v(\mathcal{B}'^r) \iff \frac{\mathcal{P}_r(\mathcal{B}^r)}{1 - \alpha^2} \geq \mathcal{P}_v(\mathcal{B}^r) + \frac{\alpha\mathcal{P}_r(\mathcal{B}^r)}{1 - \alpha^2} \iff \mathcal{P}_r(\mathcal{B}^r) \geq (1 + \alpha)\mathcal{P}_v(\mathcal{B}^r)$$

4) Los demás vértices conservan el PageRank.

Teorema. Análogos resultados para árboles cíclicos con modulación $\frac{1}{1 - \alpha^l}$.

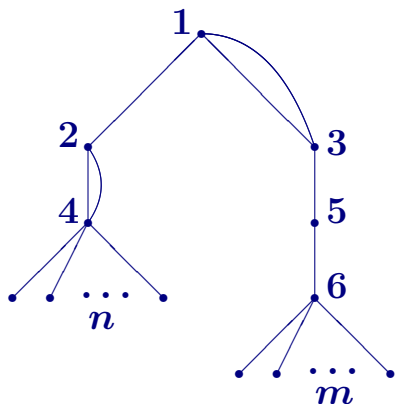
12. Manipulaciones en la topología

Teorema. La supresión de los vértices del último nivel en un árbol cíclico puede aumentar o disminuir el PageRank de la raíz.



Para $n \leq 75$ y para cualquier $m \geq 1$, al suprimir progresivamente los m vértices del último nivel $\mathcal{P}(1)$ aumenta.

Para $n \geq 76$ y para cualquier $m \geq 1$, al suprimir progresivamente los m vértices del último nivel $\mathcal{P}(1)$ disminuye.



Para $n \leq 31$ y para cualquier $m \geq 1$, al suprimir progresivamente los m vértices del último nivel $\mathcal{P}(1)$ aumenta.

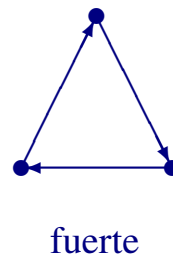
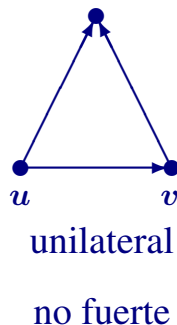
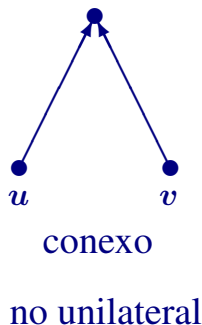
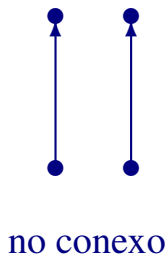
Para $n \geq 32$ y para cualquier $m \geq 1$, al suprimir progresivamente los m vértices del último nivel $\mathcal{P}(1)$ disminuye.

13. Topologías más complejas: estructura conectiva de un digrafo

conexión (débil): si $\forall u, v \in V$ hay semitrayectoria conectando u y v

conexión unilateral: si $\forall u, v \in V$ hay $u \xrightarrow{\text{camino}} v$ o $v \xrightarrow{\text{camino}} u$

conexión fuerte: si $\forall u, v \in V$ hay $u \xrightarrow{\text{camino}} v$ y $v \xrightarrow{\text{camino}} u$



En un digrafo (V, A) Fuertemente Conexo se tiene:

si $(u, v) \in A \implies uv \xrightarrow{\text{camino}} u$ es un ciclo que lo contiene.

La estructura conectiva de un digrafo se describe en términos de reducibilidad e irreducibilidad de sus matrices de adyacencia.

13. Topologías más complejas: estructura conectiva de un digrafo

A reducible si $\exists P$ de permutación / $P^{-1}AP = \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}$ con A_{ii} cuadradas.

$A_{21} = 0$: completamente reducible \longleftrightarrow digrafo no conexo.

No reducible = irreducible \longleftrightarrow digrafo fuertemente conexo.

Forma normal de A reducible: $P^{-1}AP$

$$= \begin{pmatrix} A_{11} & & & & & & & \\ 0 & \ddots & 0 & & & & & 0 \\ & & A_{pp} & & & & & \\ A_{p+1,1} & \cdots & \cdots & A_{p+1,p+1} & & & & \\ \cdots & \cdots & \cdots & \cdots & \ddots & & & 0 \\ A_{p+q,1} & \cdots & \cdots & \cdots & \cdots & \cdots & A_{p+q,p+q} & \end{pmatrix}$$

A_{ii} irreducible. $Spec(A) = \bigcup_{i=1}^{p+q} Spec(A_{ii})$.

13. Topologías más complejas: estructura conectiva de un digrafo

A reducible si $\exists P$ de permutación / $P^{-1}AP = \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}$ con A_{ii} cuadradas.

$A_{21} = 0$: completamente reducible \longleftrightarrow digrafo no conexo.

No reducible = irreducible \longleftrightarrow digrafo fuertemente conexo.

Forma normal de A reducible: $P^{-1}AP$

$$= \begin{pmatrix} A_{11} & & & & & & \\ 0 & \ddots & 0 & & & & 0 \\ & & A_{pp} & & & & \\ A_{p+1,1} & \dots & \dots & A_{p+1,p+1} & & & \\ \dots & \dots & \dots & \dots & \dots & \ddots & 0 \\ A_{p+q,1} & \dots & \dots & \dots & \dots & \dots & A_{p+q,p+q} \end{pmatrix}$$

A_{ii} irreducible. $Spec(A) = \bigcup_{i=1}^{p+q} Spec(A_{ii})$.

La conexión fuerte en un digrafo $G = (V, E)$ es una relación de equivalencia: $V = V_1 \cup \dots \cup V_r$.

$G_i = (V_i, E \cap V_i \times V_i)$ comp. con. fuertes de G

13. Topologías más complejas: estructura conectiva de un digrafo

A reducible si $\exists P$ de permutación / $P^{-1}AP = \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}$ con A_{ii} cuadradas.

$A_{21} = 0$: completamente reducible \longleftrightarrow digrafo no conexo.

No reducible = irreducible \longleftrightarrow digrafo fuertemente conexo.

Forma normal de A reducible: $P^{-1}AP$

$$= \begin{pmatrix} A_{11} & & & & & & & & \\ & 0 & \dots & 0 & & & & & \\ & & & A_{pp} & & & & & \\ A_{p+1,1} & \dots & \dots & A_{p+1,p+1} & & & & & \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & & \\ A_{p+q,1} & \dots & \dots & \dots & \dots & \dots & A_{p+q,p+q} & & \end{pmatrix}$$

A_{ii} irreducible. $Spec(A) = \bigcup_{i=1}^{p+q} Spec(A_{ii})$.

La conexión fuerte en un digrafo $G = (V, E)$ es una relación de equivalencia: $V = V_1 \cup \dots \cup V_r$.

$G_i = (V_i, E \cap V_i \times V_i)$ comp. con. fuertes de G

Digrafo condensación:

$G^* = (V^*, E^*)$ con $V^* = \{G_1, \dots, G_r\}$

$(G_i, G_j) \in E^* \Leftrightarrow i \neq j, \exists v_i \in V_i, v_j \in V_j, (v_i, v_j) \in E$.

13. Topologías más complejas: estructura conectiva de un digrafo

A reducible si $\exists P$ de permutación $/ P^{-1}AP = \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}$ con A_{ii} cuadradas.

$A_{21} = 0$: completamente reducible \longleftrightarrow digrafo no conexo.

No reducible = irreducible \longleftrightarrow digrafo fuertemente conexo.

Forma normal de A reducible: $P^{-1}AP$

$$= \begin{pmatrix} A_{11} & & & & & & & & \\ & \ddots & & & & & & & \\ & & 0 & & & & & & \\ & & & & A_{pp} & & & & \\ & A_{p+1,1} & \dots & \dots & & A_{p+1,p+1} & & & \\ & \dots & \dots & \dots & & & \ddots & & \\ & A_{p+q,1} & \dots & \dots & \dots & \dots & \dots & A_{p+q,p+q} & \end{pmatrix}$$

A_{ii} irreducible. $\text{Spec}(A) = \bigcup_{i=1}^{p+q} \text{Spec}(A_{ii})$.

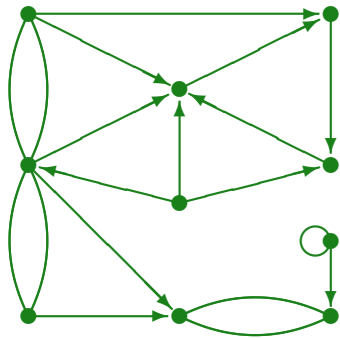
La conexión fuerte en un digrafo $G = (V, E)$ es una relación de equivalencia: $V = V_1 \cup \dots \cup V_r$.

$G_i = (V_i, E \cap V_i \times V_i)$ comp. con. fuertes de G

Digrafo condensación:

$G^* = (V^*, E^*)$ con $V^* = \{G_1, \dots, G_r\}$

$(G_i, G_j) \in E^* \Leftrightarrow i \neq j, \exists v_i \in V_i, v_j \in V_j, (v_i, v_j) \in E$.



13. Topologías más complejas: estructura conectiva de un digrafo

A reducible si $\exists P$ de permutación / $P^{-1}AP = \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}$ con A_{ii} cuadradas.

$A_{21} = 0$: completamente reducible \longleftrightarrow digrafo no conexo.

No reducible = irreducible \longleftrightarrow digrafo fuertemente conexo.

Forma normal de A reducible: $P^{-1}AP$

$$= \begin{pmatrix} A_{11} & & & & & & & \\ & \ddots & & & & & & \\ 0 & \dots & 0 & & & & & \\ & & & A_{pp} & & & & \\ A_{p+1,1} & \dots & \dots & A_{p+1,p+1} & & & & \\ \dots & \dots & \dots & \dots & \dots & \ddots & & 0 \\ A_{p+q,1} & \dots & \dots & \dots & \dots & \dots & A_{p+q,p+q} & \end{pmatrix}$$

A_{ii} irreducible. $Spec(A) = \bigcup_{i=1}^{p+q} Spec(A_{ii})$.

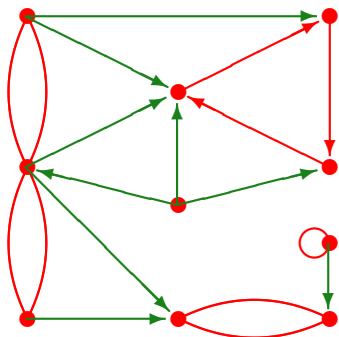
La conexión fuerte en un digrafo $G = (V, E)$ es una relación de equivalencia: $V = V_1 \cup \dots \cup V_r$.

$G_i = (V_i, E \cap V_i \times V_i)$ comp. con. fuertes de G

Digrafo condensación:

$G^* = (V^*, E^*)$ con $V^* = \{G_1, \dots, G_r\}$

$(G_i, G_j) \in E^* \Leftrightarrow i \neq j, \exists v_i \in V_i, v_j \in V_j, (v_i, v_j) \in E$.



13. Topologías más complejas: estructura conectiva de un digrafo

A reducible si $\exists P$ de permutación / $P^{-1}AP = \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}$ con A_{ii} cuadradas.

$A_{21} = 0$: **completamente reducible** \longleftrightarrow **digrafo no conexo**.

No reducible = **irreducible** \longleftrightarrow **digrafo fuertemente conexo**.

Forma normal de A reducible: $P^{-1}AP$

$$= \begin{pmatrix} A_{11} & & & & & \\ 0 & \ddots & 0 & & & \\ & & A_{pp} & & & \\ A_{p+1,1} & \dots & \dots & A_{p+1,p+1} & & \\ \dots & \dots & \dots & \dots & \ddots & 0 \\ A_{p+q,1} & \dots & \dots & \dots & \dots & A_{p+q,p+q} \end{pmatrix}$$

A_{ii} irreducible. $Spec(A) = \bigcup_{i=1}^{p+q} Spec(A_{ii})$.

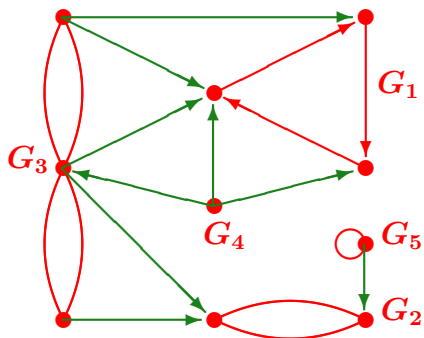
La conexión fuerte en un digrafo $G = (V, E)$ es una relación de equivalencia: $V = V_1 \cup \dots \cup V_r$.

$G_i = (V_i, E \cap V_i \times V_i)$ comp. con. fuertes de G

Digrafo condensación:

$G^* = (V^*, E^*)$ con $V^* = \{G_1, \dots, G_r\}$

$(G_i, G_j) \in E^* \Leftrightarrow i \neq j, \exists v_i \in V_i, v_j \in V_j, (v_i, v_j) \in E$.



13. Topologías más complejas: estructura conectiva de un digrafo

A reducible si $\exists P$ de permutación / $P^{-1}AP = \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}$ con A_{ii} cuadradas.

$A_{21} = 0$: completamente reducible \longleftrightarrow digrafo no conexo.

No reducible = irreducible \longleftrightarrow digrafo fuertemente conexo.

Forma normal de A reducible: $P^{-1}AP$

$$= \begin{pmatrix} A_{11} & & & & & & \\ 0 & \ddots & 0 & & & & \\ & & A_{pp} & & & & \\ A_{p+1,1} & \cdots & \cdots & A_{p+1,p+1} & & & \\ \cdots & \cdots & \cdots & \cdots & \ddots & & 0 \\ A_{p+q,1} & \cdots & \cdots & \cdots & \cdots & A_{p+q,p+q} \end{pmatrix}$$

A_{ii} irreducible. $Spec(A) = \bigcup_{i=1}^{p+q} Spec(A_{ii})$.

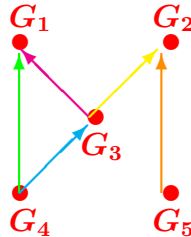
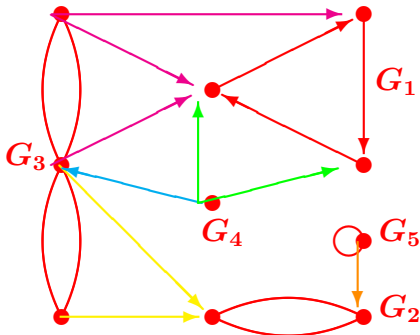
La conexión fuerte en un digrafo $G = (V, E)$ es una relación de equivalencia: $V = V_1 \cup \dots \cup V_r$.

$G_i = (V_i, E \cap V_i \times V_i)$ comp. con. fuertes de G

Digrafo condensación:

$G^* = (V^*, E^*)$ con $V^* = \{G_1, \dots, G_r\}$

$(G_i, G_j) \in E^* \Leftrightarrow i \neq j, \exists v_i \in V_i, v_j \in V_j, (v_i, v_j) \in E$.



13. Topologías más complejas: estructura conectiva de un digrafo

A reducible si $\exists P$ de permutación / $P^{-1}AP = \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}$ con A_{ii} cuadradas.

$A_{21} = 0$: completamente reducible \longleftrightarrow **digrafo no conexo.**

No reducible = **irreducible** \longleftrightarrow **digrafo fuertemente conexo.**

Forma normal de A reducible: $P^{-1}AP$

$$= \begin{pmatrix} A_{11} & & & & & \\ 0 & \ddots & 0 & & & 0 \\ & & A_{pp} & & & \\ A_{p+1,1} & \cdots & \cdots & A_{p+1,p+1} & & \\ \cdots & \cdots & \cdots & \cdots & \ddots & 0 \\ A_{p+q,1} & \cdots & \cdots & \cdots & \cdots & A_{p+q,p+q} \end{pmatrix}$$

A_{ii} irreducible. $\text{Spec}(A) = \bigcup_{i=1}^{p+q} \text{Spec}(A_{ii})$.

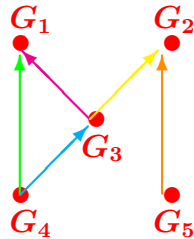
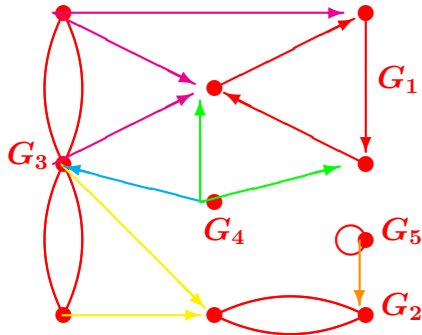
La conexión fuerte en un digrafo $G = (V, E)$ es una relación de equivalencia: $V = V_1 \cup \dots \cup V_r$.

$G_i = (V_i, E \cap V_i \times V_i)$ comp. con. fuertes de G

Digrafo condensación:

$$G^* = (V^*, E^*) \quad \text{con} \quad V^* = \{G_1, \dots, G_r\}$$

$$(G_i, G_j) \in E^* \Leftrightarrow i \neq j, \exists v_i \in V_i, v_j \in V_j, (v_i, v_j) \in E.$$



G^* es acíclico:

fuera de las componentes
conexas fuertes no hay ciclos

13. Topologías más complejas: estructura conectiva de un digrafo

A reducible si $\exists P$ de permutación $P^{-1}AP = \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}$ con A_{ii} cuadradas.

$A_{21} = 0$: completamente reducible \iff digrafo no conexo.

No reducible = irreducible \iff digrafo fuertemente conexo.

Forma normal de A reducible: $P^{-1}AP$

$$= \begin{pmatrix} A_{11} & & & & & \\ 0 & \ddots & 0 & & & \\ & & A_{pp} & & & \\ A_{p+1,1} & \cdots & \cdots & A_{p+1,p+1} & & \\ \cdots & \cdots & \cdots & \cdots & \ddots & 0 \\ A_{p+q,1} & \cdots & \cdots & \cdots & \cdots & A_{p+q,p+q} \end{pmatrix}$$

A_{ii} irreducible. $Spec(A) = \bigcup_{i=1}^{p+q} Spec(A_{ii})$.

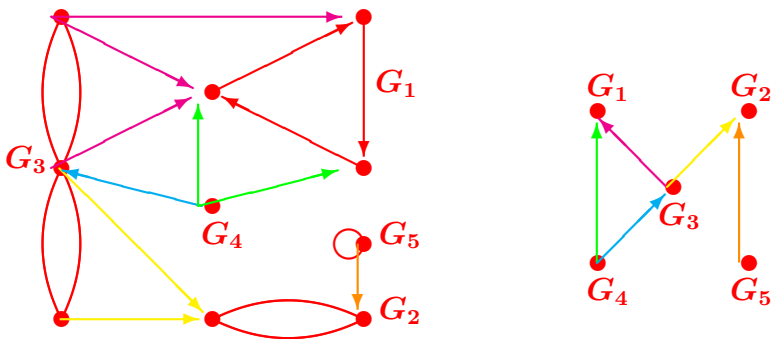
La conexión fuerte en un digrafo $G = (V, E)$ es una relación de equivalencia: $V = V_1 \cup \dots \cup V_r$.

$G_i = (V_i, E \cap V_i \times V_i)$ comp. con. fuertes de G

Digrafo condensación:

$G^* = (V^*, E^*)$ con $V^* = \{G_1, \dots, G_r\}$

$(G_i, G_j) \in E^* \iff i \neq j, \exists v_i \in V_i, v_j \in V_j, (v_i, v_j) \in E$.



Cualquier digrafo es el digrafo acíclico de sus componentes conexas fuertes

14. Topologías más complejas: cálculo paralelizado del PageRank

Observación. En redes suficientemente complejas las CCFs no son simples ciclos, sino uniones de ciclos con vértices y arcos comunes. En este caso, el uso de la fórmula de Brinkmeier puede ser confuso.

Brinkmeier. $D = (V, A)$ un multidigrafo, $v \in V$ y C su CCF.

Entonces la sucesión $r^{(i)}(v)$:

$$r^{(0)}(v) = \frac{1 - \alpha}{N}, \quad r^{(i+1)}(v) = \frac{1 - \alpha}{N} + \alpha \sum_{\substack{u \in C \\ u \rightarrow v}} \frac{r^{(i)}(u)}{od(u)} + \alpha \sum_{\substack{u \notin C \\ u \rightarrow v}} \frac{PageRank(u)}{od(u)}$$

converge a PageRank(v).

Cálculo en paralelo.

- 1) Hallar las CCFs del digrafo D y el digrafo condensación D^* .
- 2) Clasificar el digrafo acíclico D^* en niveles y etiquetarlo naturalm. ordenado.
- 3) Calcular los PageRanks de todos los vértices en todas las CCFs de cada nivel.

Como las CCFs de cada nivel son independientes, esta etapa puede paralelizarse para cada nivel.

14. Topologías más complejas: cálculo paralelizado del PageRank

Datos de la WWW en 2001

número de vértices	118 142 115
número de arcos	1 019 903 190
número de CCFs	41 126 852
CCF mayor	53 891 939
2ª CCF mayor	9 428
3ª CCF mayor	5 925
CCF de tamaño 1	39 843 421
CCF de tamaño 2	323 994
CCF de tamaño 3	154 786
ratio de vértices en la CCF mayor	45.6 %
ratio de vértices en CCFs de tamaño ≤ 3	34.1 %

Cada CCF, salvo la gigante, puede almacenarse en la memoria principal de un ordenador estándar.

El 34 % de la web se puede calcular directamente, sin iteración, resolviendo un sistema lineal de a lo sumo 3 incógnitas.

El cálculo de la CCF gigante se puede distribuir entre varios clientes reduciendo los recursos requeridos a cada cliente pero aumentando el número de comunicaciones y/o de iteraciones.

Referencias

- [1] R. Albert, H. Jeong and A.-L. Barabasi, Diameter of the World Wide Web, *Nature* **401**, 130-131, Sep 1999.
- [2] A. Arasu, J. Novak, A. Tomkins and J. Tomlin, PageRank computation and the structure of the Web: Experiments and algorithms, *The Eleventh International World Wide Web Conference, Posters.*, 2002.
- [3] A. Arratia and C. Marijuán, Cómo mejorar el PageRank de un árbol, *In: E. M. Moro et al (eds.): V Jornadas de Matemática Discreta y Algoritmica* , Pub. Univ. de Valladolid (ISBN: 84-8448-380-0), 53-60, 2006.
- [4] A. Arratia and C. Marijuán, On graph combinatorics to improve eigenvector-based measures of centrality in directed networks, *Linear Algebra Appl.* **504**, 325-353, 2016.
- [5] M. Bianchini, M. Gori and F. Scarselli, Inside PageRank, *ACM Transactions on Internet Technologies* **5** (1), 2005.
- [6] P. Bonacich, Power and centrality: a family of measures, *American Journal of Sociology* **92**, 1170-1182, 1987.
- [7] S. Brin and L. Page, The anatomy of a large scale hypertextual web search engine, *Computer Networks and ISDN Systems*, **33**, 107-117, 1998.
- [8] S. Brin, R. Motwami, L. Page and T. Winograd, The PageRank citation ranking: Bringing order to the web, *Technical Report*, Comp. Sci. Dept., Stanford University, 1998.

- [9] M. Brinkmeier, Distributed calculation of PageRank using strongly connected components, *Proceedings of the I2CS'05 in Paris (LNCS)*, 2005.
- [10] M. Brinkmeier, PageRank revisited, *ACM Transactions on Internet Technologies*, **6** (3), 2006.
- [11] T. Cormen, C. Leiserson, R. Rivest and C. Stein, Introduction to Algorithms, *The MIT Press*, 2nd. Edition, 2001.
- [12] E. Landau, Zur relativen Wertbemessung der Turnierresultate, *Deutsches Wochhenschach* **11**, 366-369, 1895.
- [13] P. Craven, Google's "Florida" Update, In: *WebWorks* (<http://www.webworkshop.net/florida-update.html>), Dec. 3, 2003.
- [14] Google Acquires Kaltix Corp, In *Google Press* (<http://www.google.com/intl/ro/press/pressrel/kaltix.html>), Sept. 30, 2003.
- [15] F. Harary, Graph Theory, *Addison-Wesley*, 1972.
- [16] L. Katz, A new status index derived from sociometric analysis, *Psychometrika*, **18**, 39-43, 1953.
- [17] S. D. Kamvar, T. H. Haveliwala, C. D. Manning, and G. H. Golub, Exploiting the block structure of the Web for computing PageRank, *Stanford University Technical Report*, 2003.

- [18] D. E. Knuth, *The Art of Computer Programming*, Addison-Wesley, volumes 1, 2, and 3, 3rd edition, 1998.
- [19] A. N. Langville and C. D. Meyer, Deeper Inside PageRank, *Internet Mathematics* **1** (3), 335–380, 2005.
- [20] A. N. Langville and C. D. Meyer, *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton University Press, June 2006.
- [21] C. Marijuán, Birational equivalence of reduced graphs, *J. Pure Appl. Algebra* **124**, 173-199, 1998.
- [22] S. Olsen, Searching for the personal touch, In: *CNET News.com* (<http://news.com.com/2100-1024-5061873.html>), August 11, 2003.