DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Ranking web

11 de octubre de 2014

Métodos Numéricos
Trabajo Práctico Nro. 2

Integrante	LU	Correo electrónico
Martin Carreiro	45/10	martin301290@gmail.com
Kevin Kujawski	459/10	kevinkuja@gmail.com
Juan Manuel Ortíz de Zárate	403/10	jmanuoz@gmail.com

Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires
Ciudad Universitaria - (Pabellón I/Planta Baja)
Intendente Güiraldes 2160 - C1428EGA
Ciudad Autónoma de Buenos Aires - Rep. Argentina
Tel/Fax: (54 11) 4576-3359
http://www.fcen.uba.ar

Índice

1. Resumen

Los sitios web a medida que fueron creciendo en cantidad en la época de los 90's, se complicó el acceso a ellos y ,a menos que alguien te comentara o a través de publicidades, era muy dificil acceder a la información deseada. Es por eso que se produjo el auge de los buscadores, que a partir de palabras claves, podrían devolverte sitios que puedan llegar a responder tu pregunta o decirte algo al respecto. Un primer problema de entrada, es que, como todo en la vida, la calidad de dicho contenido puede no ser el deseado y existan mejores. Durante este trabajo repasaremos 3 algoritmos conocidos de ranqueo de páginas web, veremos los resultados y los compararemos.
Una vez que sepamos cómo funcionan y cómo ordenan y ubican los resultados, intentaremos responder a la pregunta: cuáles son los pasos a seguir para poder mejorar tu sitio y que salga con mejor puntaje que la competencia.

2. Introducción teórica

2.1. Matriz Dispersa

Se define como matriz dispersa a aquella a la que la mayoría de sus elementos son cero. Ejemplo:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & a_{04} \\ 0 & a_{11} & a_{12} & 0 & 0 \\ 0 & 0 & 0 & a_{23} & 0 \\ 0 & 0 & 0 & a_{33} & 0 \\ a_{40} & 0 & 0 & 0 & 0 \end{bmatrix}$$

2.2. DOK vs CRS vs CSC

La matriz dispersa al tener la propiedad de tener muy pocos valores no−cero es conveniente solo guardar estos y asumir el resto como cero. Existen varias estructuras como Dictionary of Keys (dok), Compressed Sparse Row (CSR) o Compressed Sparse Column (CSC) pensadas para optimizar el espacio y las operaciones con estas estructuras de datos. En el desarrollo de este TP, utilizamos DOK por facilidad en el uso del mismo. Tanto CSR o CSC se basan en la estructura Yale y se diferencian en como guardan los mismos valores, uno priorizando las columnas y otro las filas respectivamente.
La estructura Yale consiste en a partir de la matriz original obtener tres vectores que contengan

- A = los elementos no−cero de arriba-abajo,izquierda-derecha
- IA = los indices para cada fila i del primer elemento no-cero de dicha fila
- JA = los indices de columna para cada valor de A

Si bien en caso de que haya en una fila con muchos números no-ceros es más beneficioso la utilización de esta estructura, la facilidad con DOK permite hacer pruebas más rápido.Y nos pareció poco práctico ponernos a implementar todas las lógicas requeridas para la eliminación o agregación de nuevos datos en estas estructuras ya que no hacían a la escencia del TP y complejizaban el código y el debagueo durante las pruebas y el desarrollo. Consideramos que la optimización otorgada por DOK es suficiente para el tipo de análisis que deseamos hacer sobre los algoritmos de rankeo solicitados.

3. Desarrollo

3.1. Page Rank

El algoritmo de PageRank lo dividimos en dos etapas, primero la inicialización en donde se crea la matriz estocástica y luego la corrida en donde se itera y calcula el pagerank hasta que la diferencia de norma entre los vectores sea menor que la tolerancia establecida.

Inicialización:

```
Genero un vector inicial.
Para cada nodo:
        - Si tiene salidas, inserto en cada celda de la columna correspondiente
           de los nodos de salida
        - Si no, guardo el nodo en el vector de desconectados.
```

Calculo del PageRank:

```
Hasta que converja:
        Multiplico la matriz por el vector actual.
        Aplico el algoritmo para tener en cuentas los nodos desconectados.
        Aplico el algoritmo para tener en cuenta el navegante aleatorio.
        Guardo el vector actual.
```

3.2. HITS

Este también lo dividimos en la etapa en la etapa de iniciaclización y de calculo de sus vectores. En la primera creamos la matriz estocástica e inicializamos los vectores y en la segunda calculamos los mismos hasta que iteremos k veces o la diferencia obtenida sea menor que la tolerancia.

Inicialización:

```
Creo el Dok vacio
Para cada arista:
        defino en el dok el nodo desde y hasta

Inicio los vectores de hubs y autoridades con todos sus valores en 1 y
   normalizados
```

Cálculo de vectores Hubs y Autoridades

```
Itero de 1 a K
        Para el vector de hubs multiplico el dok transpuesto por el vector de
           autoridades y normalizo
        Para el vector de autoridades multiplicoel dok transpuesto por el vector
           de hubs y normalizo
        Si la diferencia entre el nuevo vector de hubs o autoridades con su valor
            previo es menor a la tolerancia termino la iteracion
```

3.3. Indeg

Indeg utiliza plenamente la información de los links de las páginas que los apuntan y arma un promedio en la cantidad de estos sobre el total de links existentes en la red

```
Inicializo el vector de resultados con ceros.
Para cada conjunto de referencias de una pagina:
        Para cada referencia:
                Al vector resultados le sumo 1 / cantidad total de los links
```

4. Experimentación Y Resultados

4.1. Casos de prueba

A continuación se listarán los casos utilizados y después se compararán los resultados.

- MOVIES: Este caso incluye 5797 páginas
- ABORTION: Este caso incluye 2293 páginas
- GENETIC: Este caso incluye 3468 páginas
- STANFORD: Este caso incluye 281903 páginas
- GOOGLE: Este caso incluye 916428 páginas

4.2. Comparación de Normas

En esta sección vamos a mostrar como evoluciona la norma Manhattan (también conocida como distancia L1) entre dos vectores a medida que se suceden las iteraciones. La norma Manhattan es la distancia entre dos vectores, o en otras palabras, la suma de la diferencia coordenada a coordenada en modulo:

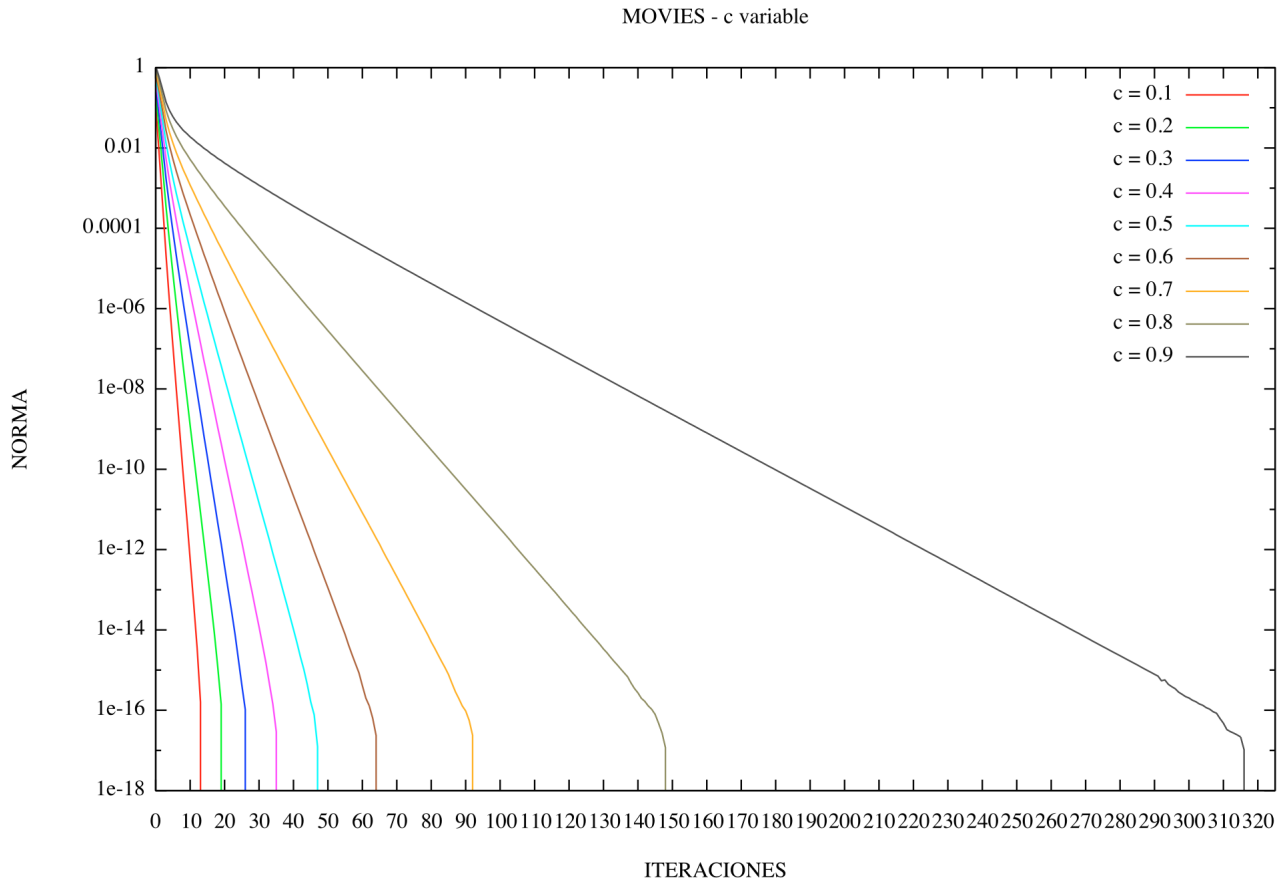
$$d_1(\mathbf{p}, \mathbf{q}) = \|\mathbf{p} - \mathbf{q}\|_1 = \sum_{i=1}^{n} |p_i - q_i|$$

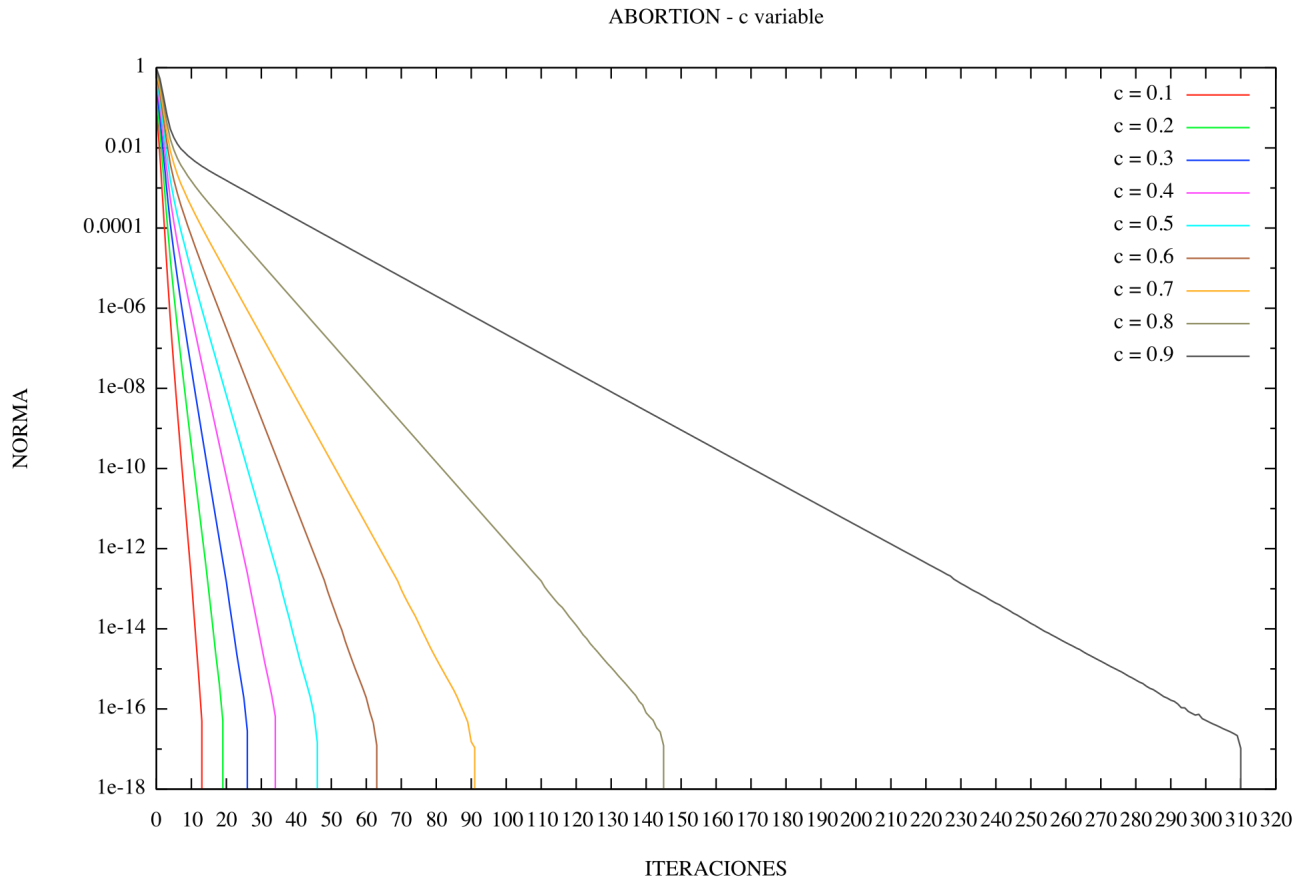
4.2.1. PageRank

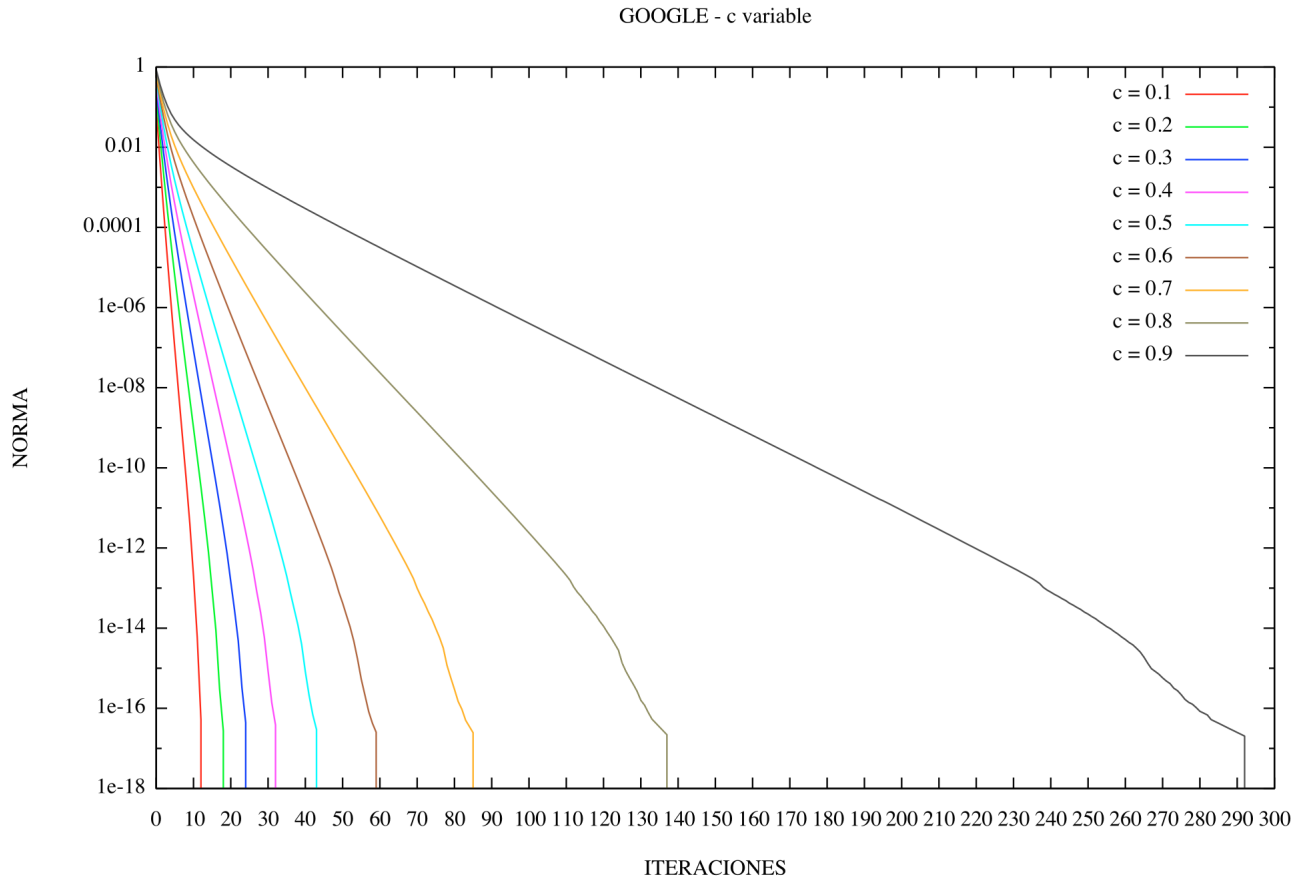
Para evaluar el comportamiento de la norma manhattan variando la probabilidad del navegante aleatorio, el cual de ahora en más lo denotaremos como el parámetro **c**

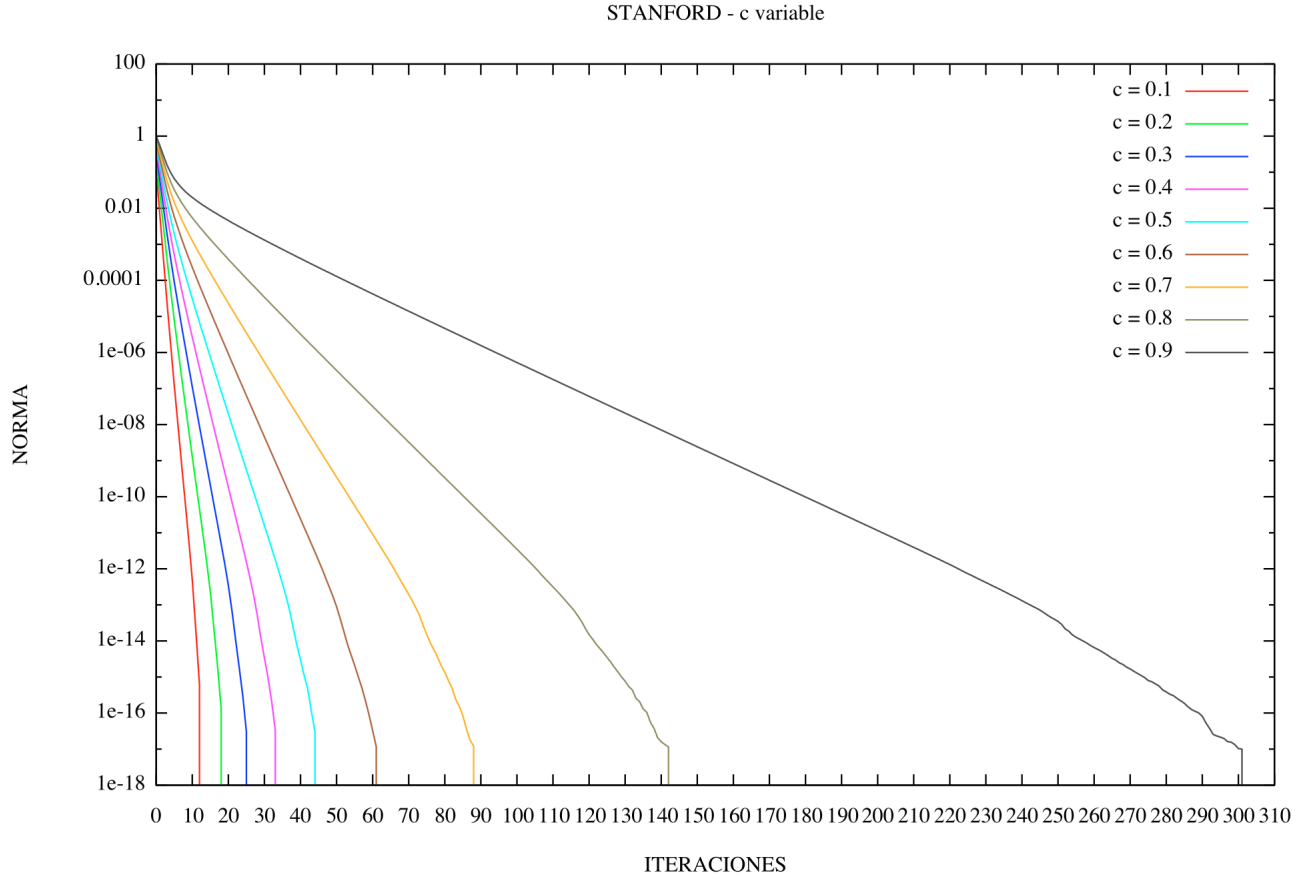
Los casos de prueba se corrieron sin un limite entre normas pero si con un limite de 1000 iteraciones, ya que, según lo que investigamos, con un c $\approx$ 0.15 la matriz suele converger en un máximo de 50 iteraciones, que por lo que se puede observar suele ser proporcional esta relación a medida que aumenta el c.

A continuación se muestran los resultados para cuatro tests de como evoluciona la norma a lo largo de las iteraciones y como varía la misma con distintos c, y que luego discutiremos más adelante. Cabe aclarar que expresamos los valores de la norma en escala logarítmica para una mejor visualización y para que se obtenga un mejor entendimiento de como disminuye de a varias magnitudes en cada iteración.


MOVIES - c variable
NORMA
ITERACIONES
c = 0.1
c = 0.2
c = 0.3
c = 0.4
c = 0.5
c = 0.6
c = 0.7
c = 0.8
c = 0.9
1
0.01
0.0001
1e-06
1e-08
1e-10
1e-12
1e-14
1e-16
1e-18
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280 290 300 310 320


ABORTION - c variable
NORMA
ITERACIONES
c = 0.1
c = 0.2
c = 0.3
c = 0.4
c = 0.5
c = 0.6
c = 0.7
c = 0.8
c = 0.9
1
0.01
0.0001
1e-06
1e-08
1e-10
1e-12
1e-14
1e-16
1e-18
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280 290 300 310 320


GOOGLE - c variable
NORMA
ITERACIONES
1
0.01
0.0001
1e-06
1e-08
1e-10
1e-12
1e-14
1e-16
1e-18
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280 290 300
c = 0.1
c = 0.2
c = 0.3
c = 0.4
c = 0.5
c = 0.6
c = 0.7
c = 0.8
c = 0.9


STANFORD - c variable
NORMA
ITERACIONES
100
1
0.01
0.0001
1e-06
1e-08
1e-10
1e-12
1e-14
1e-16
1e-18
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280 290 300 310
c = 0.1
c = 0.2
c = 0.3
c = 0.4
c = 0.5
c = 0.6
c = 0.7
c = 0.8
c = 0.9

4.2.2. HITS

Estas corridas se hicieron para k= 100 ya que con esto alcanzaba para analizar los compartamientos deseados. La tolerancia en todos los casos fue de 0 ya que nos pareció mas interesante ver que tanto converge mas allá de que para nosotros una divergencia de 0.00001 ya es despreciable.

A continuación se muestran los resultados para cuatro instancias distintas, 3 medianas y una grande, de como evoluciona la norma a lo largo de las iteraciones en ambos vectores :

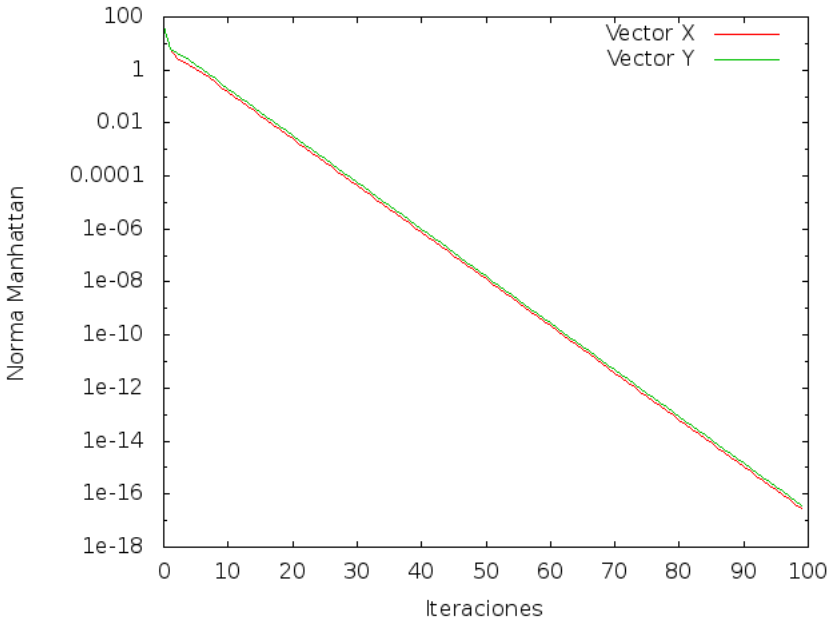


Figura 1: Abortion expanded

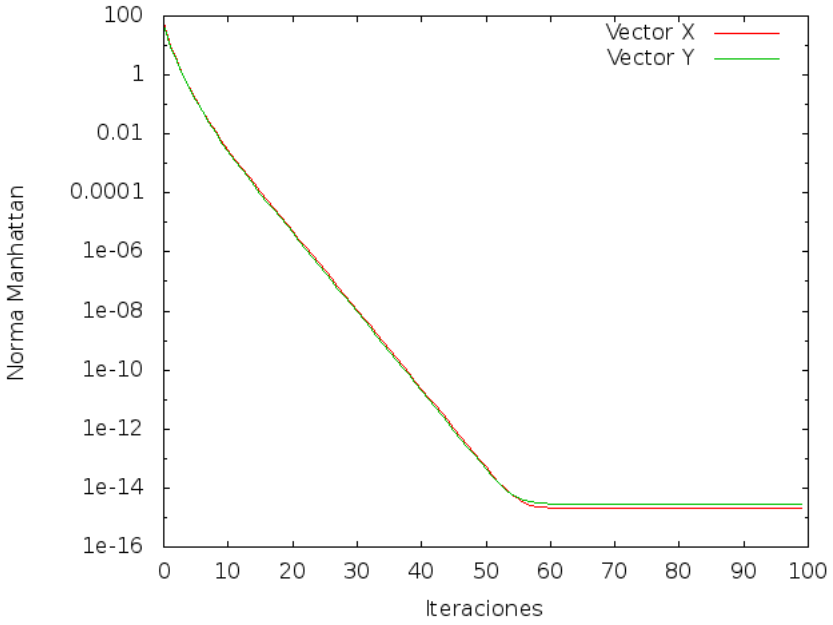


Figura 2: Genetic expanded

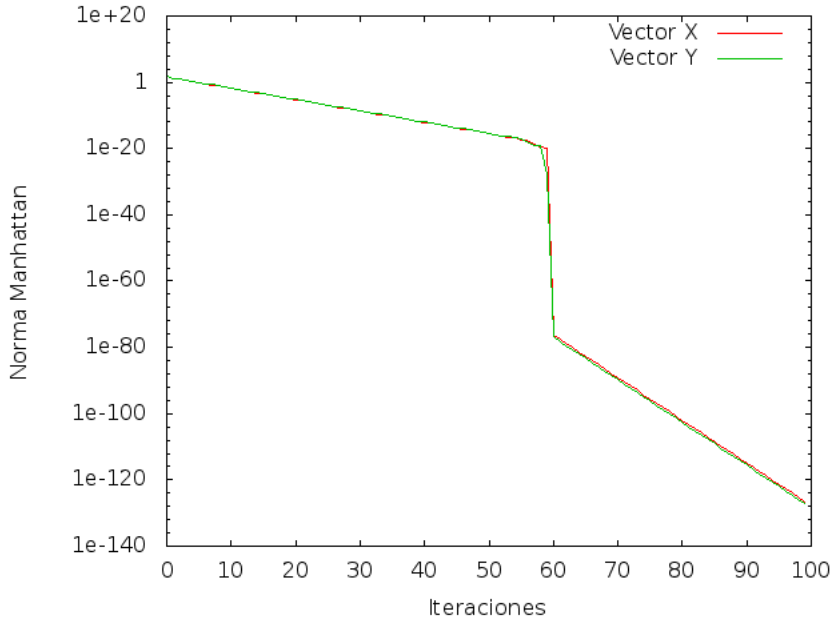


Figura 3: Movies expanded

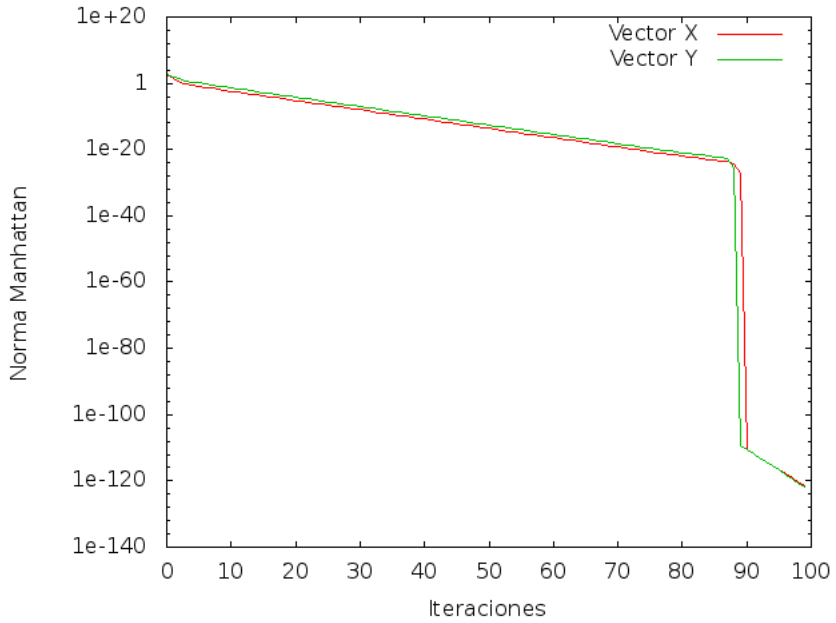


Figura 4: Standford

4.3. Comparación de Tiempos

El siguiente gráfico muestra la evolución del tiempo de computo en función del tamaño de la red para cada algoritmo. La red utilizada en todos los casos es una red estrella en la que todos los nodos (o sitios) apuntan al primero de ellos. Utilizamos este tipo de grafo ya que en c++ es el más rápido y simple de crear teniendo en cuenta además que la forma del grafo no tiene un impacto de eficiencia en los algoritmos sino su tamaño en nodos y aristas es el que cambia el tiempo de ejecución. Por esto no nos pareció pertinente probar distintos tipos de grafos (arboles, completos, bipartito, etc) sin más bien el tamaño de los mismos.

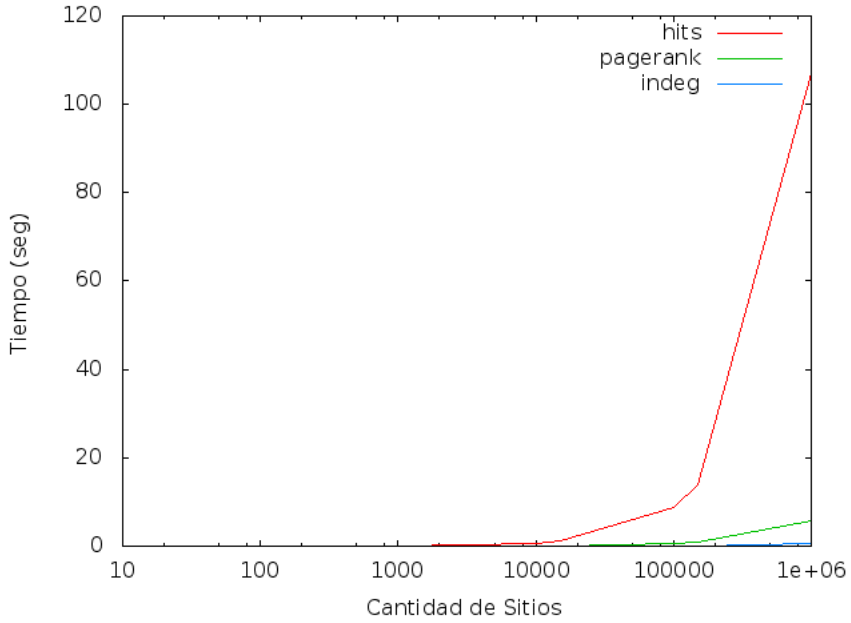


Figura 5: Tiempo de ejecución en función del tamaño de la red

5. Discusión

5.1. PageRank

Claramente podemos notar que a medida que el C crece, el algoritmo toma más iteraciones en achicar la norma. Esto se debe a que el grado de aleatoriedad elimina el peso de la unión entre los sitios e indica una uniformidad en el comportamiento, entonces la matriz si bien estocástica ahora se encuentra distribuida esa suma = 1 por columna en varias filas. Esto produce mayor cantidad de iteraciones en el método de la potencia ya que la mayor uniformidad de la matriz provoca que ninguna 'zona' de la matriz absorba más que las demás. [1]
También es bastante notorio que a pesar de que los distintos casos de prueba sean muy diferentes entre si y hasta cientos de veces más grandes, la evolución de la norma converge de formas casi idénticas y lo mismo sucede para las iteraciones requeridas hasta llegar a la norma variando el parámetro c.

5.2. HITS

En todos los casos podemos observar que tanto el vector de hubs como el de autridades convergen de forma muy similar, sólo en la instancia grande hay una pequeña diferencia pero es bastante despreciable. Por otro lado podemos ver que los casos en los que mas drásitca es la convergencia (abortion y genetic) los valores inciales de la norma manhattan son muy altos (alrrededor de 100), provocando asi que se equiparen con las que comienzan en valores mas bajos pero convergen mas lentamente (movies y standford). En estos dos últimos casos además podemos notar grandes saltos de convergencia pasando en pocas iteracion de $1e^{20}$ a menos de $1e^{80}$, entiendiendo, aca sí, que la diferencia es totalmente despreciable y el valor obtenido ya ha convergido. De todas formas consideramos que puede ser un punto de interés para analizar mas en profundidad ya que más allá de decir que entendemos de eso, no sabríamos explicar porque se produce ese salto.

5.3. Ejemplos de comportamiento esperado

A continuación veremos en redes pequeñas como se comporta cada algoritmo para ver si su comportamiento es el esperado.

5.3.1. PageRank

Para mostrar un ejemplo del comportamiento del PageRank generamos una red de 11 nodos y lo corrimos con un $c = 0.85$.

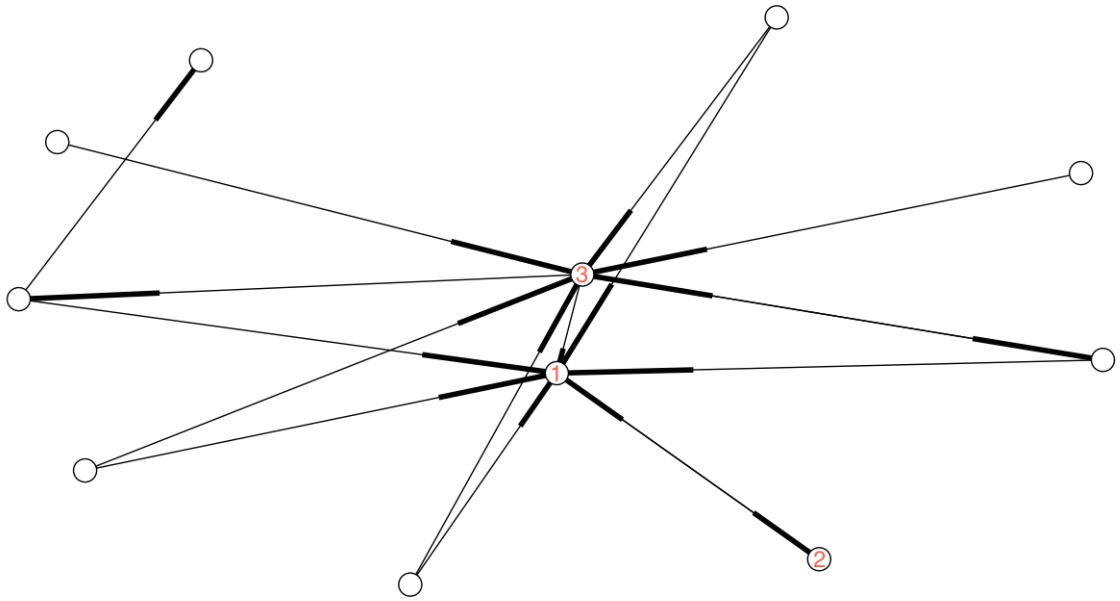


Figura 6: Red de 11 nodos, c=0.85

Lo particular de esta red es que uno *ingenuamente* podría pensar que los dos sitios centrales 1 y 3 van a ser los que mas PageRank obtengan, pero esa suposición se basaría en que el algoritmo solo tiene en cuenta los grados de entrada de cada nodo. El resultado real que se obtiene de esta red es que el orden de PageRank se da por el 1, 2 y 3 (los demás nodos no son importantes para ilustrar el comportamiento). El nodo 2 le gana al 3 ya que la diferencia sustancial es que el 1 que tiene un alto valor lo apunta únicamente al 2, es decir, le da todo el peso que él tiene, mientras que el nodo 3 a pesar de tener muchos sitios que lo apuntan estos son sitios de muy bajo valor de los cuales solo tiene nodos de salida.

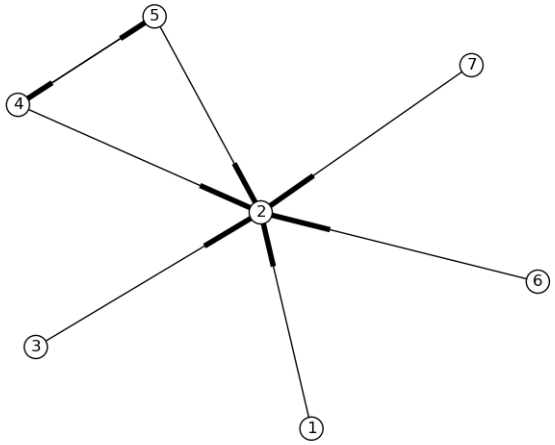


Figura 7: Red de 7 nodos

5.3.2. HITS

Resultado obtenido:

$$\begin{bmatrix} & Autoridad & Hub \\ Nodo1 & 0{,}000000 & 0{,}383092 \\ Nodo2 & 0{,}967054 & 0{,}000000 \\ Nodo3 & 0{,}000000 & 0{,}383092 \\ Nodo4 & 0{,}180008 & 0{,}454401 \\ Nodo5 & 0{,}180008 & 0{,}454401 \\ Nodo6 & 0{,}000000 & 0{,}383092 \\ Nodo7 & 0{,}000000 & 0{,}383092 \end{bmatrix}$$

Efectivamente podemos observar que en la columna de autoridades el nodo 2 es el mayor ya que es el que mas apuntado esta y todos aquellos que tienen 0 es porque no son apuntados por ninguno. Por otro lado en la columna de hubs podemos ver que los nodos 4 y 5 son los que mayor valor tienen ya que son los que mas apuntan a otros nodos con 2 salidas.

5.3.3. Indeg

Veamos el comportamiento dada esa pequeña red.

Resultado obtenido:

$$\begin{bmatrix} & Puntaje \\ Nodo1 & 0{,}000000 \\ Nodo2 & 0{,}714285 \\ Nodo3 & 0{,}000000 \\ Nodo4 & 0{,}142857 \\ Nodo5 & 0{,}142857 \\ Nodo6 & 0{,}000000 \\ Nodo7 & 0{,}000000 \end{bmatrix}$$

Este algoritmo naive es bastante claro de interpretar, y el resultado es claramente el esperado. El nodo 2 posee mucha más calidad de sitio ya que es el más apuntado, y en el segundo puesto empatando el nodo 4 y 5, por ser aquellos con más sitios apuntándolos, excpetuando el nodo 2. El resto de los nodos no reciben ningún tipo de link hacia ellos, por lo que su puntaje es de cero.

5.4. Análisis cualitativo

En esta sección procederemos a discutir sobre la calidad de resultados que obtenemos de cada algoritmo y luego los compararemos entre si.
Como el objetivo de este trabajo práctico esta enfocado al ranking web que se le asigna a los distintos sitios de internet, consideramos como buenos resultados aquellos que aparecerían en la primer página de los buscadores, es decir, los primero 10 resultados serán los que consideraremos para el análisis.

5.4.1. PageRank

Según el paper de Bryan y Leise, quienes proponen el algoritmo, lo más común es que el valor del navegante aleatorio sea de 0.15. Por lo tanto creemos que con este valor es donde aparecerán los mejores resultados, pero también veremos que sucede con valores de 0.5 y 0.85, ya que estos valores indican por un lado que la probabilidad del navegante entre quedarse e irse es equiprobable y por otro lado es el inverso de lo que ellos consideran como el valor más común. En valores de 0 y 1 no tendrían sentido el análisis ya que por un lado daría la matriz original y por el otro una matriz equiprobable.
El caso de prueba que utilizaremos es el dado por la cátedra, **Abortion**, y lo elegimos ya que es un tema bastante discutido donde se pueden encontrar resultados interesantes.

Resultados con un c=0.15

1. **No relacionado con el tema**
   http://www.allexperts.com/about.asp
   AllExperts.com
2. http://www.nrlc.org
   National Right to Life Organization
3. **No relacionado con el tema**
   http://www.phone-soft.com/at/cyber-world/international/o1480i.htm
   PHONE-SOFT INTERNET DIRECTORY INTERNATIONAL:HERB THERAPY LINKS
4. http://www.lm.com/ jdehullu
   Ariadne's Thread: On abortion, affirmative action, hate speech
5. http://www.plannedparenthood.org
   Planned Parenthood Federation of America
6. http://www.gynpages.com
   Abortion Clinics OnLine
7. http://www.care-net.org/link.htm
   CareNet Links
8. http://www.naral.org
   NARAL: Abortion and Reproductive Rights: Choice For Women

9. http://www.crosswalk.com/ftr/1,,17,00.htm
   Crosswalk.com Forums - Welcome
10. http://www.cais.com/agm/main
    The Abortion Rights Activist Home Page

Resultados con un c=0.5

1. http://www.allexperts.com/about.asp
   AllExperts.com
2. http://www.nrlc.org
   National Right to Life Organization
3. **No relacionado con el tema**
   http://home.about.com
   About - The Human Internet
4. **No relacionado con el tema**
   http://www.phone-soft.com/at/cyber-world/international/o1480i.htm
   PHONE-SOFT INTERNET DIRECTORY INTERNATIONAL:HERB THERAPY LINKS
5. http://www.lm.com/ jdehullu
   Ariadne's Thread: On abortion, affirmative action, hate speech
6. http://www.plannedparenthood.org
   Planned Parenthood Federation of America
7. http://www.care-net.org/link.htm
   CareNet Links
8. http://www.gynpages.com
   Abortion Clinics OnLine
9. http://www.marchforlife.org
   The March For Life Fund Home Page
10. **No relacionado con el tema**
    http://www.jbs.org
    The John Birch Society

Resultados con un c=0.85

1. **No relacionado con el tema**
   http://www.jbs.org
   The John Birch Society
2. **No relacionado con el tema**
   http://home.about.com
   About - The Human Internet
3. **No relacionado con el tema**
   http://www.allexperts.com/about.asp
   AllExperts.com

4. **No relacionado con el tema**
   http://www.aobs-store.com
   American Opinion Book Services Online Store

5. http://www.nrlc.org
   National Right to Life Organization

6. **No relacionado con el tema**
   http://www.trimonline.org
   TRIMonline - Lower Taxes Through Less Government

7. http://www.marchforlife.org
   The March For Life Fund Home Page

8. **No relacionado con el tema**
   http://www.phone-soft.com/at/cyber-world/international/o1480i.htm
   PHONE-SOFT INTERNET DIRECTORY INTERNATIONAL:HERB THERAPY LINKS

9. **No relacionado con el tema**
   http://www.reagan.com
   The Reagan Information Interchange

10. **No relacionado con el tema**
    http://www.pregnancycenters.org
    Pregnancy Centers Online

En base a los resultados se puede ver como a medida que aumenta el c empiezan a aparecer resultados que poco tienen que ver con el tema directamente, ya que puede estar relacionado de alguna forma o no diferenciarse tanto del eje temático.
Nos pareció extraño que aparece siempre muy bien posicionado el sitio web All Experts, que nada tiene que ver con el tema de los abortos, por lo tanto decidimos hacer un foco especial en este para ver porque sucedía esto y llegamos a la conclusión que es debido a que el factor mas determinante es que gran cantidad de sitios referidos al tema y a su vez bien posicionados (aunque fuera del top 10) apuntaban al mismo, y por lo tanto le daban bastante peso a All Experts.
Sucede algo parecido con otro sitio de venta de software que aparece pero no nos pareció importante su análisis ya que es un claro caso de publicidad paga en anuncios de los sitios que hablan sobre el tema.
Aunque se pueda ver que con un c menor los resultados tienen relación con el tema nos pareció que no son lo suficiente buenos como para considerarlos excelente resultados cuando se busca sobre un tema tan discutido como el aborto, esperando quizás más definiciones sobre el tema y luchas por su legalización/penalización.

5.4.2. HITS

Este analásis de calidad lo haremos sobre el tema death penalty. Veremos cuales son los primeros 5 resultados que devuelve HITS en cuanto a autoridades y hubs.

Resultados Hubs

1. http://www.clarkprosecutor.org/html/links/dplinks.htm
   Death Penalty Links

2. http://faculty.etsu.edu/blankenm/deathlinks.htm
   Death Penalty Links

3. http://coramnobis.com/portal/deathpen.html
   A Capital Defender's Toolbox: criminal defense death penalty litigation online resource center

4. http://info-s.com/deathpenalty.html
   The Info Service

5. http://members.xoom.com/ccadp/links.htm
   Canadian Coalition Against the Death Penalty - Collection of Links

Resultados Autoridades

1. http://sun.soci.niu.edu/ critcrim/dp/dp.html
   Death Penalty Information

2. http://www.aclu.org/issues/death/hmdp.html
   Death Penalty and the ACLU

3. http://www.ncadp.org
   National Coalition To Abolish the Death Penalty

4. http://www.smu.edu/ deathpen
   Death Penalty News and Updates

5. http://www.deathpenalty.org
   Death Penalty Focus

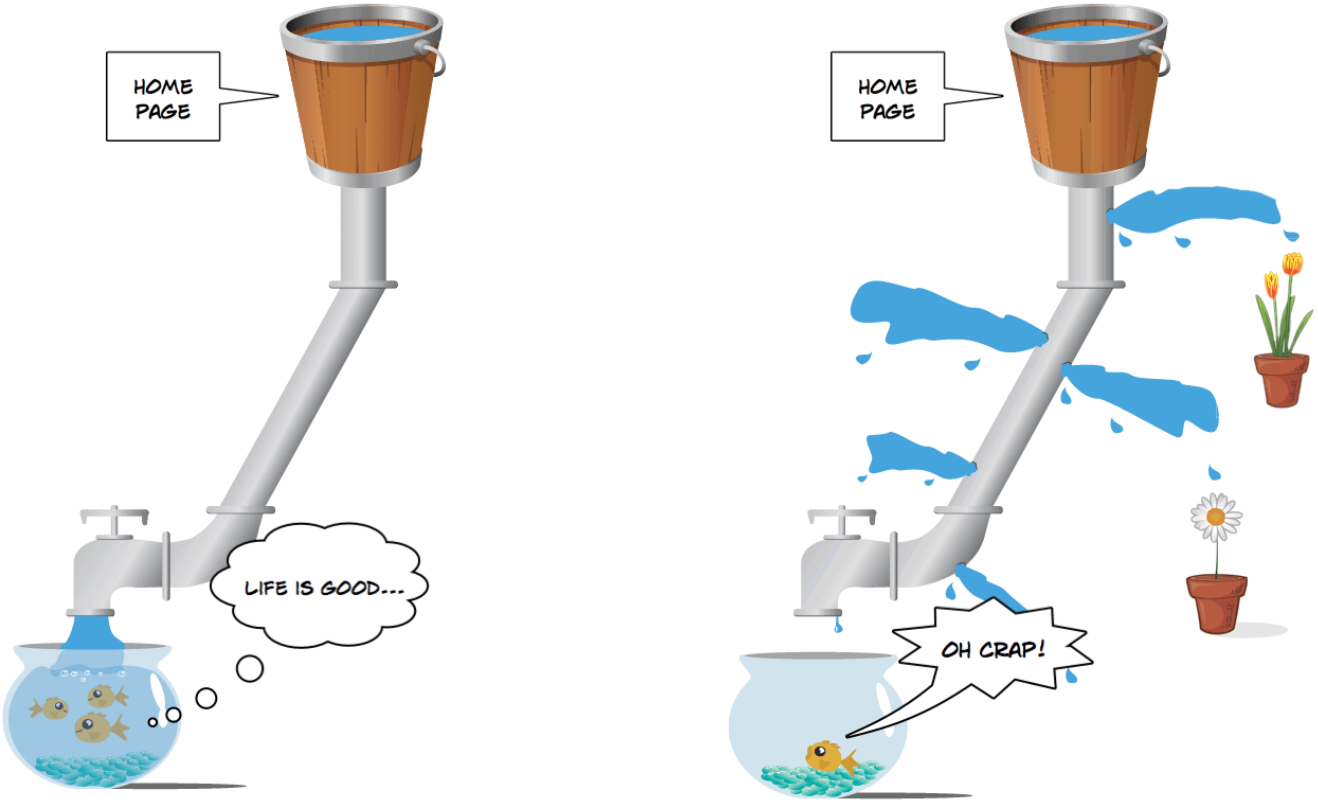
Aquí podemos observar claramente que la calidad además de ser, a priori, buena y correcta, además tiene mucho sentido que la mayoría de los hubs sobre el tema sean links sobre pena de muerta, o servicios de información o central de recursos sobre litigios en penas de muerte. Y que, por otro lado, las autoridades sean diarios con noticias y novedades, organizaciones enfocadas a eso o paginas institucionales (.edu). También es de destacar que ningún link pareciera ser spam, o sobre algo no relacionado.

5.4.3. Comparación

6. Conclusiones

6.1. PageRank

A medida que fuimos investigando y probando el algoritmo del PageRank nos fue quedando cada vez más claro como es que funciona y que se necesita para obtener un buen resultado para un sitio en particular. Lo que nos pareció interesante es explicar el algoritmo con la siguiente interpretación metafórica:



Tomando como al sitio a analizar en cuestión como la pecera, y a otro sitio que tiene un link a nuestro sitio como el balde se puede observar que cuando el *recurso*, en este caso el agua, se reparte equitativamente a todos los destinatarios, por lo tanto, si mi pecera es la única que recibe agua voy a obtener más que si tiene otras bocas la canilla con la cual compartir. Esto es lo mismo que sucede en la web y tiene el cuenta el PageRank, cada sitio le distribuye equitativamente una probabilidad a cada salida, cuya suma total es 1. Por lo tanto, me conviene más que me linkee un sitio con pocas salidas que uno con gran cantidad, pero suponiendo que sus respectivos PageRank son similares, ya que mi PageRank también va a depende del de mis entradas, por lo tanto también hay que tener esto en cuenta, ya que es un factor bastante influyente. Por consiguiente, no solo depende la cantidad de sitios que apuntan a si no también el PageRank de cada uno (la *calidad*)

6.2. HITS

En el gráfico que nos muestra el tiempo de computo en función del tamaño de los grafos podemos observar que para grafos grandes este algoritmo se vuelve bastante ineficiente. Sin embargo no debemos olvidar que en su paper[2] Kleinberg habla de que este algoritmo debe ser aplicado no sobre toda la red sin sobre un subconjunto de la misma (*root set*) obtenido de una busqueda incial. Por lo tanto si acotamos el análisis a los grafos mas acotados podemos ver que el tiempo de computo es aceptable y hasta muy parecido al de page rank.

6.3. INDEG

Este algoritmo es bastante simple y en una red chica y confiable puede llegar a valer. Igualmente tiene mucho peso la confiabilidad, ya que es muy simple de crecer tu puntaje, simplemente comprando un lugar mínimo en la mayor cantidad de páginas posibles. Volviendo al ejemplo anterior, notar que al todas las páginas tener el mismo peso, el nodo dos gana más puntaje simplemente por comprar espacio en las páginas

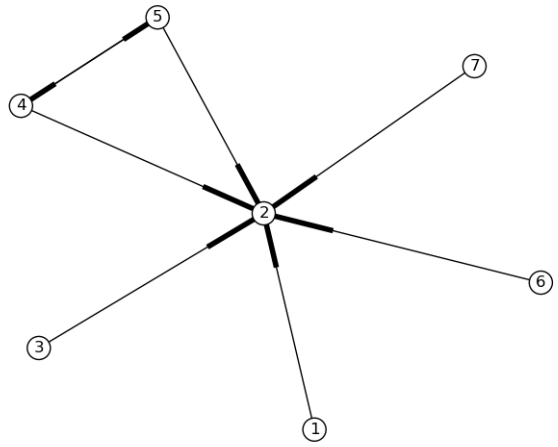


Figura 8: Red de 7 nodos

6.4. Mejor estrategia para comprar links

6.4.1. PageRank

6.4.2. HITS

Referencias

[1] $http://personales.upv.es/\ pedroche/inv/_docs/fpedrochev4(sema).pdf$

[2] $JonM.Kleinberg.Authoritativesourcesinahyperlinkedenvironment.J.ACM, 46(5) : 604, 632, September 1999.$