

PageRank: Bringing order to the Web

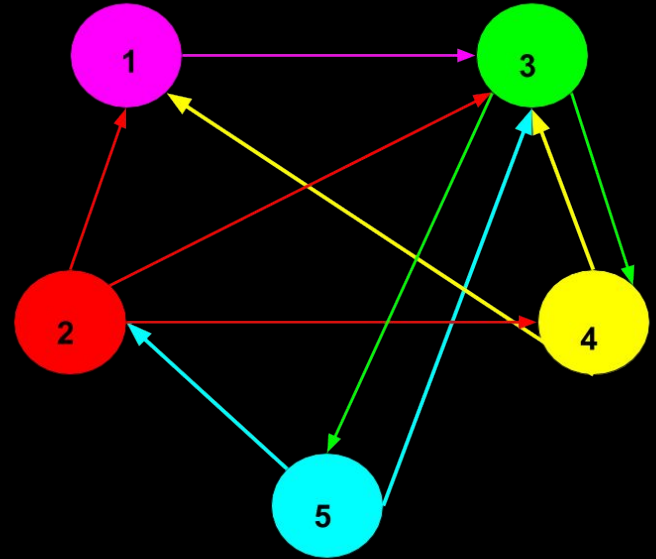
Diana Laura Nicolás Pavía

¿Cómo encontrar la
página web más
relevante dentro de
una Base de Datos?

Caminata Aleatoria

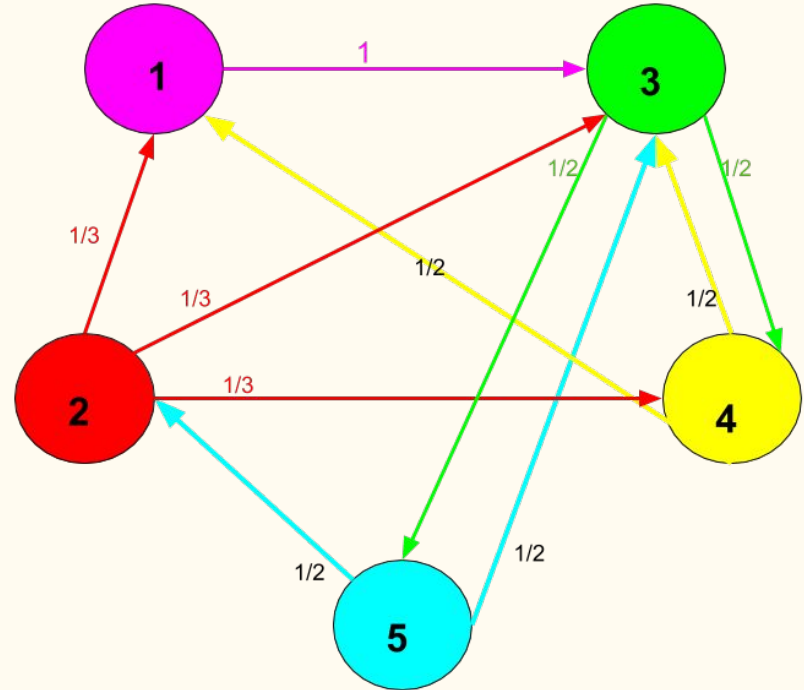
Modelo

Digráfica dirigida



Procedimiento

1. Asigna importancia a cada página.
2. Distribuye la importancia conforme cita a otras páginas.
3. Inicialmente todas las páginas tienen la misma probabilidad de ser elegidas.
4. Inicia una caminata aleatoria.



Hipótesis:

El vector de
distribución converge
a un punto llamado
distribución
estacionaria.

Ideas de refuerzo de la hipótesis

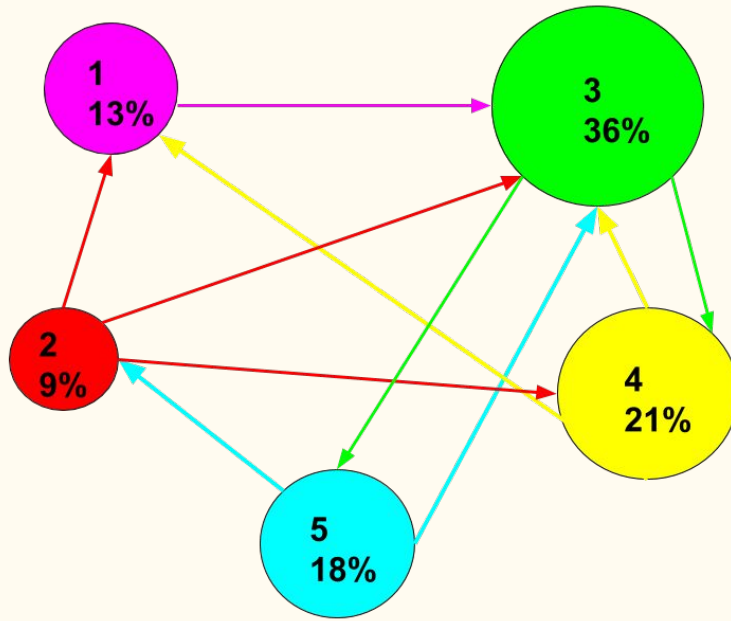
Teorema de existencia y unicidad de distribuciones estacionarias.

Teorema de convergencia de una cadena de Markov.

La gráfica que representa la base de datos traducida a una cadena de Markov cumple con las hipótesis del teorema de la convergencia.

Las variables que pueden influir en el resultado son...

- Las páginas no tienen la misma probabilidad de ser elegidas.
- Las páginas no distribuyen su importancia de forma uniforme.
- La gráfica puede no ser fuertemente conexa.
- La gráfica puede no ser aperiódica



- Algoritmo de Tarjan
- Algoritmo de Periodicidad de una gráfica
- Algoritmo para Exponenciación Binaria

Simulación de una Cadena de Markov

Conclusión

- Las aplicaciones de los procesos estocásticos es muy amplia.
- El modelo funciona bien bajo algunos supuestos... Y sin ellos también.
- El modelo usando cadenas de Markov es el que actualmente se usa junto con un montón de algoritmos más para cada búsqueda.
- La solución general se traduce un algoritmo “sencillo” que oculta toda la complejidad de las ideas que hay detrás de él.
- El poder de cómputo necesario “no” es tan grande como podríamos imaginar para una simulación real.
- Reflexionar sobre el algoritmo real.