

BIG DATA

DIPLOMADO DE DATOS 2021

Clase 6: Recuperación de Información a gran escala (II)

Aidan Hogan

aidhog@gmail.com

Con Elasticsearch

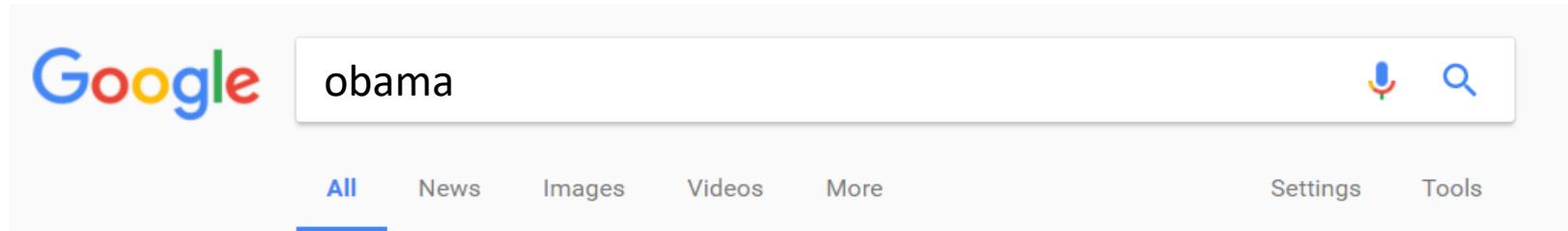
Búsqueda: "obama"

Resultados:

1. http://es.wikipedia.org/wiki/Obama_Republican
2. [http://es.wikipedia.org/wiki/Obama_\(Fukui\)](http://es.wikipedia.org/wiki/Obama_(Fukui))
3. http://es.wikipedia.org/wiki/Michelle_Obama
4. http://es.wikipedia.org/wiki/Barack_Obama
5. http://es.wikipedia.org/wiki/Republicanos_por_Obama
6. http://es.wikipedia.org/wiki/Cartel_%22Hope%22_de_Obama Cartel
7. http://es.wikipedia.org/wiki/Ricardo_Mangue_Obama_Nfubea
8. http://es.wikipedia.org/wiki/Transición_presidencial_de_Barack_Obama
9. http://es.wikipedia.org/wiki/Primera_investidura_presidencial_de_Barack_Obama
10. http://es.wikipedia.org/wiki/Campaña_presidencial_de_Barack_Obama_de_2008



¿Cómo Google obtiene tan buenos resultados?



About 462,000,000 results (0.71 seconds)

Barack Obama (@BarackObama) · Twitter

<https://twitter.com/BarackObama> 

Well said, Jimmy. That's exactly why we fought so hard for the ACA, and why we need to protect it for kids like Billy. And congratulations! [twitter.com/jimmykimmel...](https://twitter.com/jimmykimmel)

11 hours ago · Twitter

The Office of Barack and Michelle Obama

<https://www.barackobama.com/> 

Welcome to the Office of Barack and Michelle **Obama**. We Love You Back. Play video. The Office of Barack and Michelle **Obama**. © 2017 | Legal & Privacy.

Barack Obama - Wikipedia

https://en.wikipedia.org/wiki/Barack_Obama 

Barack Hussein **Obama** II is an American politician who served as the 44th President of the United States from 2009 to 2017. He is the first African American to ...



¿Cómo Google obtiene tan buenos resultados?

Google that one movie where the guy breaks his leg and spies on his neighbor

Web Videos News Images Shopping More Search tools

About 64,700,000 results (0.91 seconds)

Rear Window (1954) - IMDb
www.imdb.com/title/tt0047396/ - Internet Movie Database
★★★★★ Rating: 8.6/10 - 274,497 votes

Google da da da dum symphony

Web Videos News Shopping Images More Search tools

About 107,000 results (0.36 seconds)



Beethoven - Symphony No. 5 in C Minor (1) - YouTube
www.youtube.com/watch?v=W2qW6fOtAMY

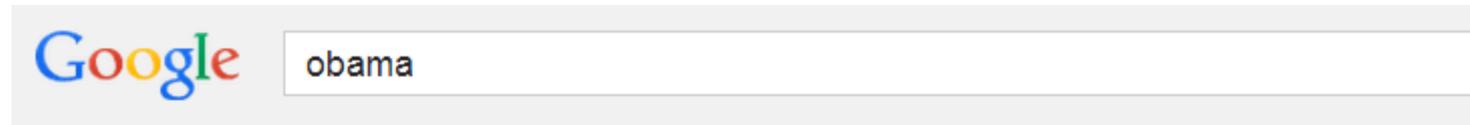


sometimes when i'm

- sometimes when i'm **alone i use comic sans**
- sometimes when i'm **alone i google myself**
- sometimes when i'm **alone i cry**
- sometimes when i'm **all alone**
- sometimes when i'm **dreaming**
- sometimes when i'm **sad i like to cut myself**
- sometimes when i'm **dreaming lyrics**
- sometimes when i'm **alone**
- sometimes when i'm **driving on the road at night**
- sometimes when i'm **alone i wonder**

Google Search I'm Feeling Lucky

Las dos caras del ranking: Relevancia



Web Images News Videos More ▾ Search tools

About 16,700,000 results (0.23 seconds)

Broccoli - Wikipedia, the free encyclopedia

en.wikipedia.org/wiki/Broccoli ▾

Broccoli is an edible green plant in the cabbage family, whose large flowering head is used as a vegetable. The word **broccoli** comes from the Italian plural of ...

[Cauliflower](#) - [Romanesco broccoli](#) - [Broccoli \(disambiguation\)](#) - [Broccolini](#)

Broccoli - The World's Healthiest Foods

www.whfoods.com/genpage.php?tname=foodspice&dbid=9 ▾

Broccoli can provide you with some special cholesterol-lowering benefits if you will cook it by steaming. The fiber-related components in **broccoli** do a better job ...

News for broccoli

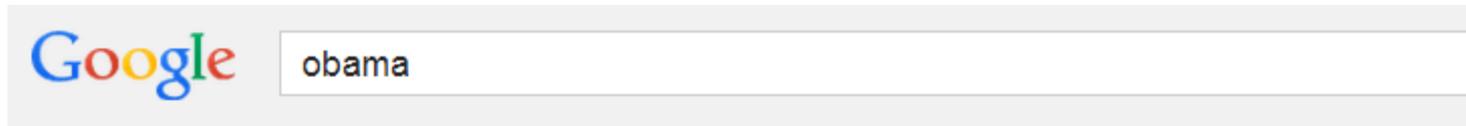
Mistakes We All Make With Spaghetti, Steak And ...

Huffington Post - 2 days ago

But in her new book *Brassicas: Cooking the World's Healthiest Vegetables*, she says plunking **broccoli**, cauliflower or brussels sprouts into ...



Las dos caras del ranking: Importancia



Web Images News Videos More ▾ Search tools

About 48,100,000 results (0.26 seconds)

Mount Obama - Wikipedia, the free encyclopedia

en.wikipedia.org/wiki/Mount_Obama ▾

Mount Obama (known as **Boggy Peak** until August 4, 2009) is the highest point in the nation of Antigua and Barbuda and on the island of Antigua. It lies in the far ...

Images for mount obama

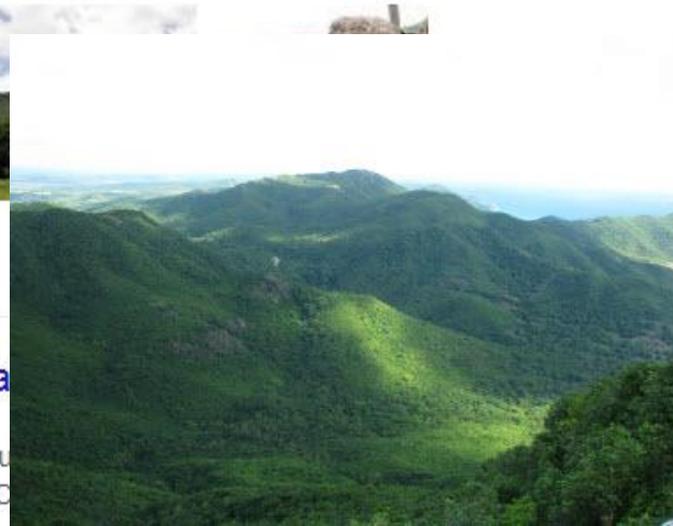
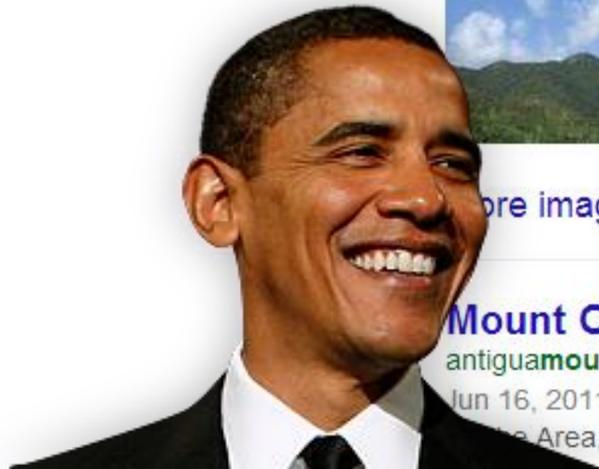
Report images



More images for mount obama

Mount Obama National Park | Antigua and Barbuda
antiguamountobama.com/

Jun 16, 2011 - As the **Mount Obama** Committee continues its work in the Mount Obama National Park Area, the committee organized a site visit to the O...



Apache Lucene usa puntaje de relevancia

Búsqueda: "obama"

Resultados:

1. http://es.wikipedia.org/wiki/Obama_Republican
2. [http://es.wikipedia.org/wiki/Obama_\(Fukui\)](http://es.wikipedia.org/wiki/Obama_(Fukui))
3. http://es.wikipedia.org/wiki/Michelle_Obama
4. http://es.wikipedia.org/wiki/Barack_Obama
5. http://es.wikipedia.org/wiki/Republicanos_por_Obama
6. http://es.wikipedia.org/wiki/Cartel_%22Hope%22_de_Obama Cartel
7. http://es.wikipedia.org/wiki/Ricardo_Mangue_Obama_Nfubea
8. http://es.wikipedia.org/wiki/Transición_presidencial_de_Barack_Obama
9. http://es.wikipedia.org/wiki/Primera_investidura_presidencial_de_Barack_Obama
10. http://es.wikipedia.org/wiki/Campaña_presidencial_de_Barack_Obama_de_2008

Entonces ¿cómo podemos definir un proceso de ranking en el cual Barack Obama salga como el primer resultado?



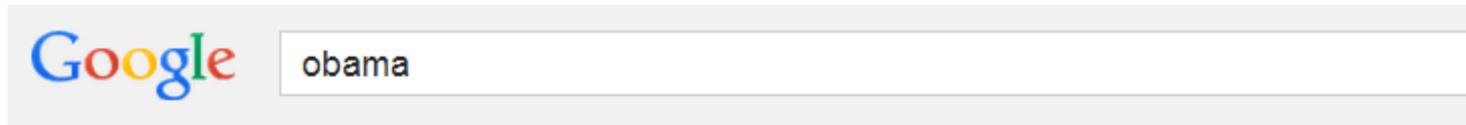
Veremos ...



RANKING:

IMPORTANCIA

Las dos caras del ranking: Importancia



Web Images News Videos More ▾ Search tools

About 48,100,000 results (0.26 seconds)

Mount Obama - Wikipedia, the free encyclopedia

en.wikipedia.org/wiki/Mount_Obama ▾

Mount Obama (known as **Boggy Peak** until August 4, 2009) is the highest point in the nation of Antigua and Barbuda and on the island of Antigua. It lies in the far ...

Images for mount obama

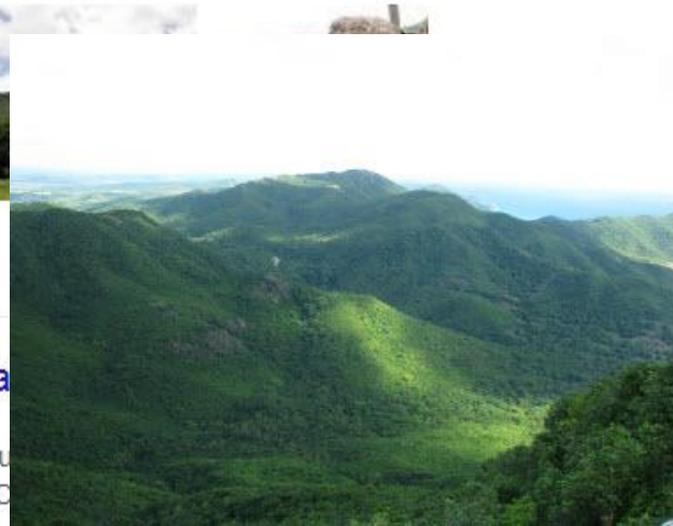
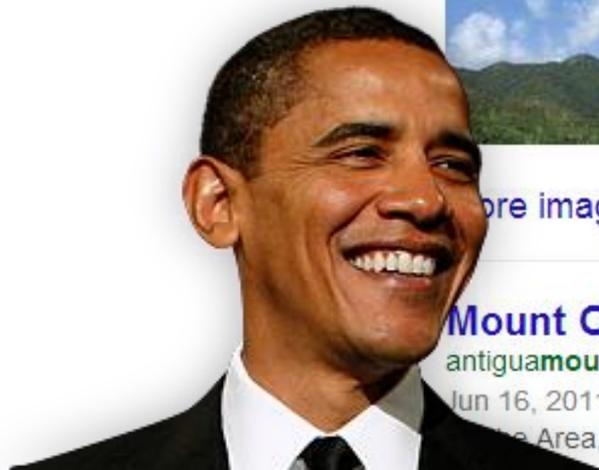
Report images



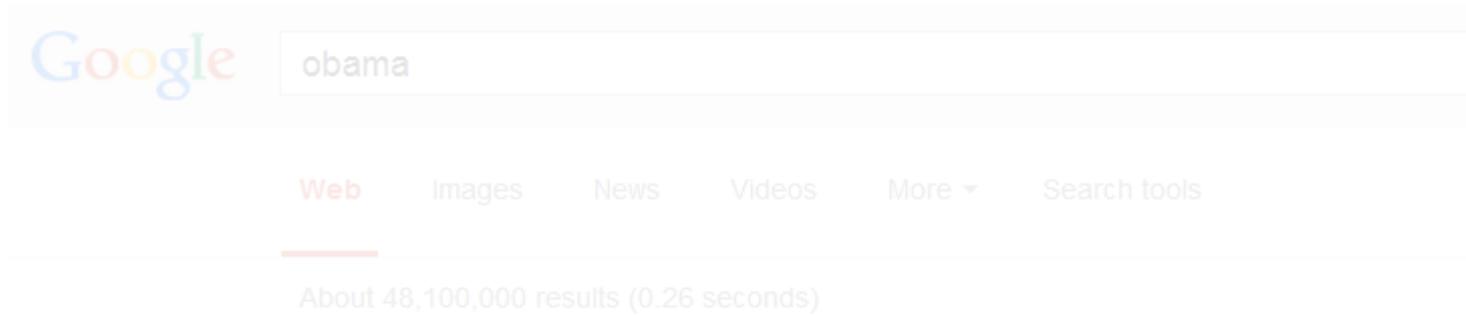
More images for mount obama

Mount Obama National Park | Antigua and Barbuda
antiguamountobama.com/

Jun 16, 2011 - As the **Mount Obama** Committee continues its work in the Mount Obama National Park Area, the committee organized a site visit to the O...



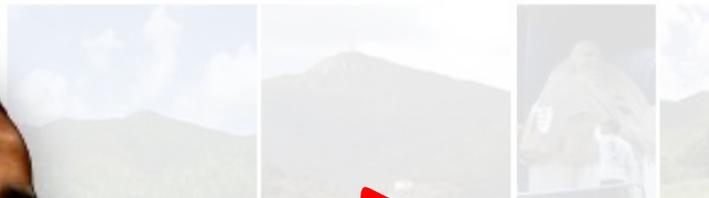
Las dos caras del ranking: Importancia



¿Cómo podemos determinar que Barack Obama es un resultado más importante que Mount Obama? 

Images for mount obama

Report images



More images for mount obama
Mount Obama National Park | Antigua a
antiguamountobama.com/

Jun 16, 2011 - As the Mount Obama Committee continu
Area, the committee organized a site visit to the C



Análisis de enlaces ("Links analysis")

¿A qué documento veremos más enlaces?
¿La página de Wikipedia de Mount Obama? o
¿La página de Wikipedia de Barack Obama?



Análisis de enlaces ("Links analysis")

- Considera los enlaces como "votos" para una página
- Un enlace es como un "like" de la Web abierta ...



(... incluso si la página es enlazada de manera negativa.)

Análisis de enlaces ("Links analysis")

Entonces, si contamos el número de enlaces entrantes a una página web, conocemos su importancia, ¿cierto?



"Spamming" de enlaces



semanticweb.com™

The Voice of Semantic Technology Business
Big Data, Linked Data, Smart Data

Home

Events

Media

Industry Verticals

Answers

Jobs

Questions

Tags

Users

Badges

[Claritin](#) [Clomid](#) [Combivent](#) [Confido](#) [Copegus](#) [Cordarone](#) [Coreg](#) [Coumadin](#) [Cozaar](#) [Crestor](#) [Cyklokapron](#) [Cymbalta](#) [Cystone](#) [Cytotec](#) [Danazol](#) [Deltasone](#) [Depakote](#) [Desyrel](#) [Detrol](#) [Diabecor](#) [Diakof](#) [Diarex](#) [DidroneI](#) [Differin](#) [Dilantin](#) [Diovan](#) [Dostinex](#) [Elavil](#) [Elimite](#) [Emsam](#) [Endep](#) [Eurax](#) [Evecare](#) [Evista](#) [Exelon](#) [Famvir](#) [Feldene](#) [Femara](#) [Femcare](#) [Flomax](#) [Flonase](#) [Flovent](#) [Fosamax](#) [Gasex](#) [Geodon](#) [Geriforte](#) [Herbolax](#) [High Love](#) [Himcocid](#) [Himcolin](#) [Himcospaz](#) [Himplasia](#) [Hoodia](#) [Hytrin](#) [Hyzaar](#) [Imdur](#) [Imitrex](#) [Inderal](#) [Ismo](#) [Isoptin](#) [Isordil](#) [Kamagra](#) [Karela](#) [Keftab](#) [Koflet](#) [Kytril](#) [Lamictal](#) [Lamisil](#) [Lanoxin](#) [Lariam](#) [Lasix](#) [Lasuna](#) [Leukeran](#) [Levaquin](#) [Levlen](#) [Levothroid](#) [Lincocin](#) [Lioresal](#) [Lisinopril](#) [Liv.52](#) [Lopid](#) [Lopressor](#) [Loprox](#) [Lotensin](#) [Lotrisone](#) [Loxitane](#) [Lozol](#) [Lukol](#) [Lynoral](#) [Maxaquin](#) [Menosan](#) [Mentat](#) [Mentax](#) [Mevacor](#) [Mexitil](#) [Miacalcin](#) [Micardis](#) [Mobic](#) [Monoket](#) [Motrin](#) [Myambutol](#) [Mycelex-G](#) [Mysoline](#) [Naprosyn](#) [Neurontin](#) [Nicotinell](#) [Nimotop](#) [Nirdosh](#) [Nizoral](#) [Nolvadex](#) [Nonoxinol](#) [Noroxin](#) [Omnicef](#) [Ophthalmicare](#) [Oxytrol](#) [Pamelor](#) [Parlodol](#) [Paxil](#) [Penisole](#) [Phentermine](#) [Pilex](#) [Plan B](#) [Plavix](#) [Plendil](#) [Pletal](#) [Prandin](#) [Pravachol](#) [Prednisone](#) [Premarkin](#) [Prevacid](#) [Prilosec](#) [Prinivil](#) [Procardia](#) [Prograf](#) [Prometrium](#) [Propecia](#) [Proscar](#) [Protonix](#) [Proventil](#) [Prozac](#) [Purin](#) [Purinethol](#) [Quibron-T](#) [Relafen](#) [Renalka](#) [Reosto](#) [Requip](#) [Retin-A](#) [Revvia](#) [Rhinocort](#) [Rimonabant](#) [Risperdal](#) [Rocaltrol](#) [Rogaine](#) [Rumalaya](#) [Sarafem](#) [Septilin](#) [Serevent](#) [Seropheue](#) [Seroquel](#) [Shallaki](#) [Shoot](#) [Sinequan](#) [Singular](#) [Snoroff](#) [Sorbitrate](#) [Speman](#) [Starlix](#) [StretchNil](#) [Stromectol](#) [Styplon](#) [Sumycin](#) [Superman](#) [Sustiva](#) [Synthroid](#) [Tenormin](#) [Topamax](#) [Trandate](#) [Tricor](#) [Trimox](#) [Triphala](#) [Tulasi](#) [Urispas](#) [V-Gel](#) [Vantin](#) [Vasodilan](#) [Vasotec](#) [Ventolin](#) [Viramune](#) [Vytorin](#) [Xeloda](#) [Xenacore](#) [Zanaflex](#) [Zantac](#) [Zebeta](#) [Zelnorm](#) [Zerit](#) [Yerba Diet](#) [Wellbutrin SR](#) [Women Attracting Pheromones](#) [Women's Intimacy Enhancer](#) [Women's Intimacy Enhancer Cream](#) [Virility Gum](#) [Vitamin A & D](#) [Viagra + Cialis](#) [Viagra + Cialis + Levitra](#) [Viagra Jelly](#) [Viagra Soft + Cialis Soft](#) [Viagra Soft Tabs](#) [Ultimate Male Enhancer](#) [Toprol XL](#) [Touch-Up Kit](#) [Tentex Royal](#) [Tentex Forte](#) [Tiberius Erectus](#) [Zero Nicotine 2 Complete Professional Whitening Kits 2 Sets Of Moldable Mouth Trays 36 Beauty Acne-n-Pimple Cream ActoPlus Met Superloss Multi SleepWell \(Herbal XANAX\) Shuddha Guggulu Rythmol SR Rumalaya Forte Pulmicort Inhaler Professional Plasma Tooth Whitening Kit Premium Diet Patch Penis Growth Oil Penis Growth Pack Penis Growth Patch Penis Growth Pills Orgasm Enhancer Norpace CR Mental Booster Men Attracting Pheromones Menopause Gum Male Enhancement Oil Male Enhancement Patch Male Enhancement Pills Male Sexual Tonic InnoPran XL Hoodia Weght Loss Gum Hoodia Weight Loss Patch Human Growth Hormone](#)

[deleted] Kala Jada Specialist +91961



black magic specialist baba ji call now +919610897260



<http://www.blackmagicspecialist.net.in>



java

edit | close | undelete | more ▼

¿Cómo podríamos combatir el "spam" así?



[Maxx Herbal](#)
[Deluxe](#)
[Calcium](#)

[Cialis Jelly](#) [Cialis Soft Tabs](#) [Calcium Carbonate Dust Enhancer](#) [Deconase AQ](#) [Anarium Diet Pills](#) [Advair Diskus](#) [Advanced Gain Pro](#) [Breast Augmentation](#) [Breast Enhancement](#) [Breast Enhancement Gel](#) [Breast Enhancement Gum](#) [Breast Intense](#) [Buy Trazodone](#) [Buy Celebrex](#) [Buy Alprazolam](#) [Buy Tramadol](#) [Buy Fioricet](#) [Buy Soma](#) [Buy Cialis](#) [Buy Carisoprodol](#) [Buy Levitra](#) [Buy Ultram](#) [Buy Ambien](#) [Buy Viagra](#) [Buy Xanax](#) [Buy Phentermine](#) [Buy Valium](#) [Buy Diazepam](#) [Generic Celebrex](#) [Generic Alprazolam](#) [Generic Tramadol](#) [Generic Fioricet](#) [Generic Soma](#) [Generic Cialis](#) [Generic Carisoprodol](#) [Generic Levitra](#) [Generic Ultram](#) [Generic Ambien](#)

La importancia de un enlace

¿Cuál debería pesar más?

¿Un enlace desde http://en.wikipedia.org/wiki/Barack_Obama? o

¿Un enlace desde <http://blackmagicspecialist.net.in>?



¿Pero por qué?

Y ¿cómo podríamos medir esta intuición?

PageRank



PageRank

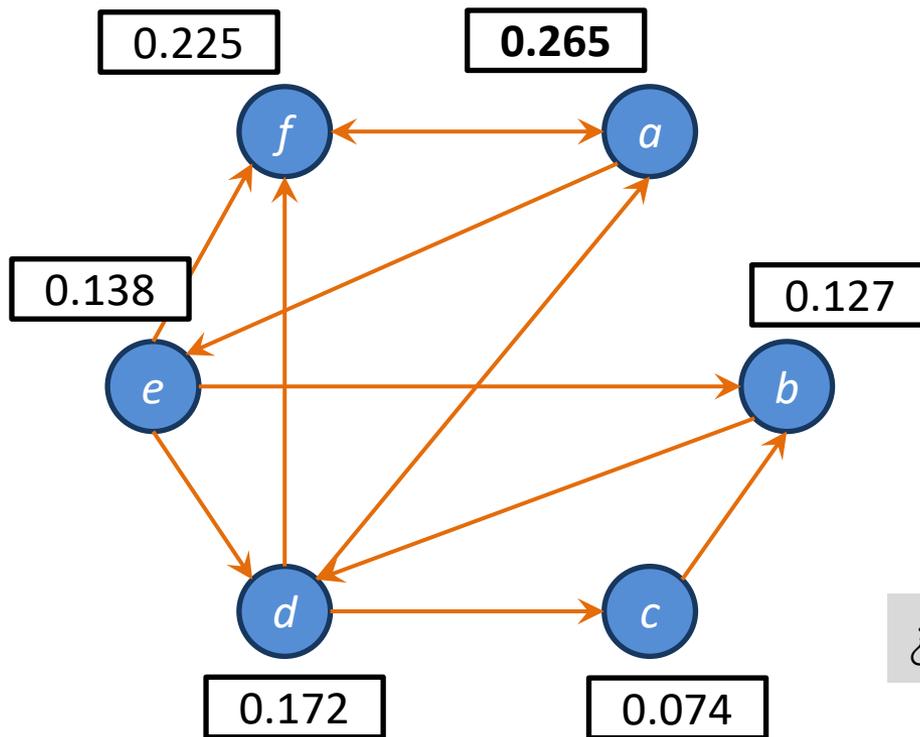
- No tan solo un conteo de enlaces entrantes
 - Un enlace de una página más importante pesa más
 - Un enlace de una página con pocos enlaces pesa más
 - ∴ Una página con más enlaces entrantes desde páginas más importantes (las cuales tienen menos enlaces salientes) es más importante

PageRank es recursivo



La versión básica de PageRank

- La Web: un grafo dirigido



$$G = [V, E]$$

Vértices
(páginas)

Arcos
(enlaces)

¿Cuál es el vértice más importante?

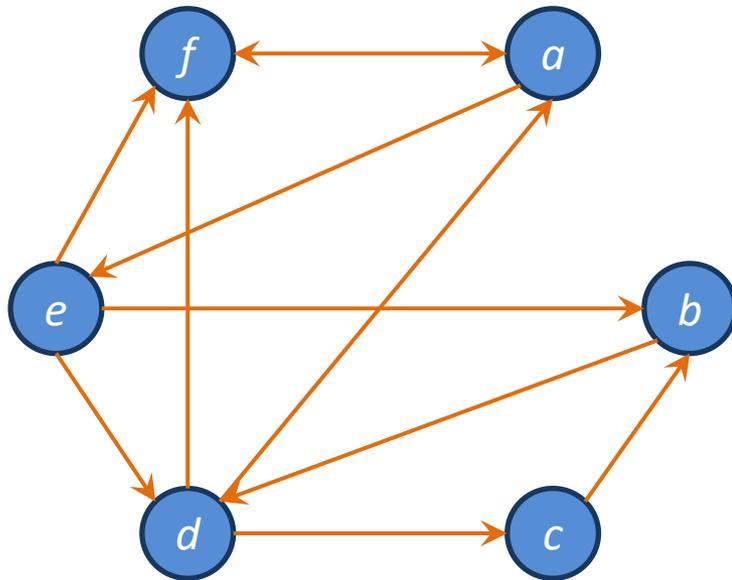


$$V = \{a, b, c, d, e, f\}$$

$$E = \{(a, e), (a, f), (b, d), (c, b), (d, a), (d, c), (d, f), (e, b), (e, d), (e, f), (f, a)\}$$

La versión básica de PageRank

- La Web: un grafo dirigido



$$G = \boxed{V} \boxed{E}$$

Vértices
(páginas)

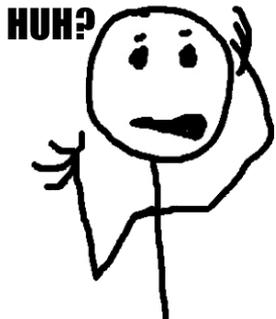
Arcos
(enlaces)

$$\text{out}(v) := \{v' \in V \mid (v, v') \in E\}$$

$$\text{in}(v) := \{v' \in V \mid (v', v) \in E\}$$

$$\text{rank}_0(v) := \frac{1}{|V|}$$

$$\text{rank}_i(v) := \sum_{v' \in \text{in}(v)} \frac{\text{rank}_{i-1}(v')}{|\text{out}(v')|}$$



La versión básica de PageRank

$$G = [V, E]$$

Vértices
(páginas)

Arcos
(enlaces)

$$\text{rank}_1(f) = \frac{1}{6} \times \frac{1}{3}$$

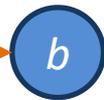


$$\text{rank}_0(e) = \frac{1}{6}$$

$$|\text{out}(e)| = 3$$



$$\text{rank}_1(b) = \frac{1}{6} \times \frac{1}{3}$$



$$\text{rank}_1(d) = \frac{1}{6} \times \frac{1}{3}$$

$$\text{out}(v) := \{v' \in V \mid (v, v') \in E\}$$

$$\text{in}(v) := \{v' \in V \mid (v', v) \in E\}$$

$$\text{rank}_0(v) := \frac{1}{|V|}$$

$$\text{rank}_i(v) := \sum_{v' \in \text{in}(v)} \frac{\text{rank}_{i-1}(v')}{|\text{out}(v')|}$$

La versión básica de PageRank

$$G = [V, E]$$

Vértices
(páginas)

Arcos
(enlaces)

$$\text{rank}_1(f) = \frac{1}{6} \times \frac{1}{3}$$



$$\text{rank}_0(e) = \frac{1}{6}$$

$$|\text{out}(e)| = 3$$



$$\text{rank}_1(b) = \frac{1}{6} \times \frac{1}{3} + 1 \times \frac{1}{6}$$



$$\text{rank}_1(d) = \frac{1}{6} \times \frac{1}{3}$$



$$\text{rank}_0(c) = \frac{1}{6}$$

$$|\text{out}(c)| = 1$$

$$\text{out}(v) := \{v' \in V \mid (v, v') \in E\}$$

$$\text{in}(v) := \{v' \in V \mid (v', v) \in E\}$$

$$\text{rank}_0(v) := \frac{1}{|V|}$$

$$\text{rank}_i(v) := \sum_{v' \in \text{in}(v)} \frac{\text{rank}_{i-1}(v')}{|\text{out}(v')|}$$

La versión básica de PageRank

$$G = [V, E]$$

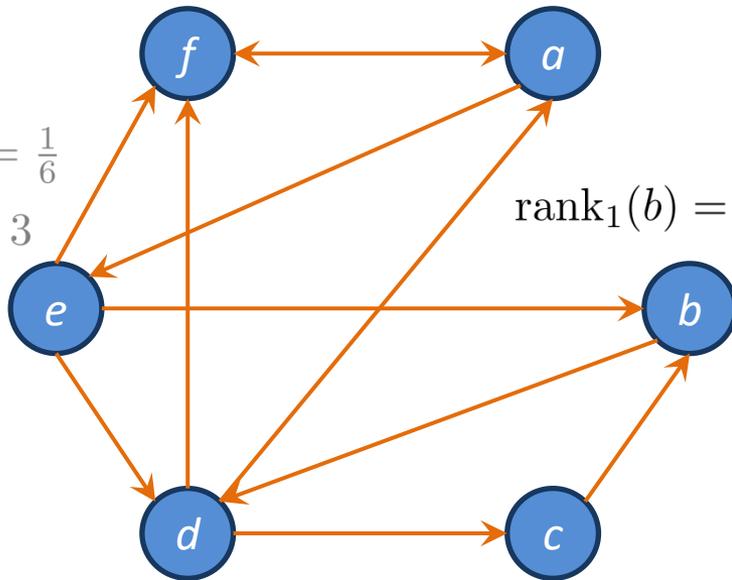
Vértices
(páginas)

Arcos
(enlaces)

$$\text{rank}_1(f) = \frac{1}{6} \times \frac{1}{3} \quad \dots$$

$$\text{rank}_0(e) = \frac{1}{6}$$

$$|\text{out}(e)| = 3$$



$$\text{rank}_1(b) = \frac{1}{6} \times \frac{1}{3} + 1 \times \frac{1}{6}$$

$$\text{rank}_1(d) = \frac{1}{6} \times \frac{1}{3}$$

$$\text{rank}_0(c) = \frac{1}{6}$$

$$|\text{out}(c)| = 1$$

$$\text{out}(v) := \{v' \in V \mid (v, v') \in E\}$$

$$\text{in}(v) := \{v' \in V \mid (v', v) \in E\}$$

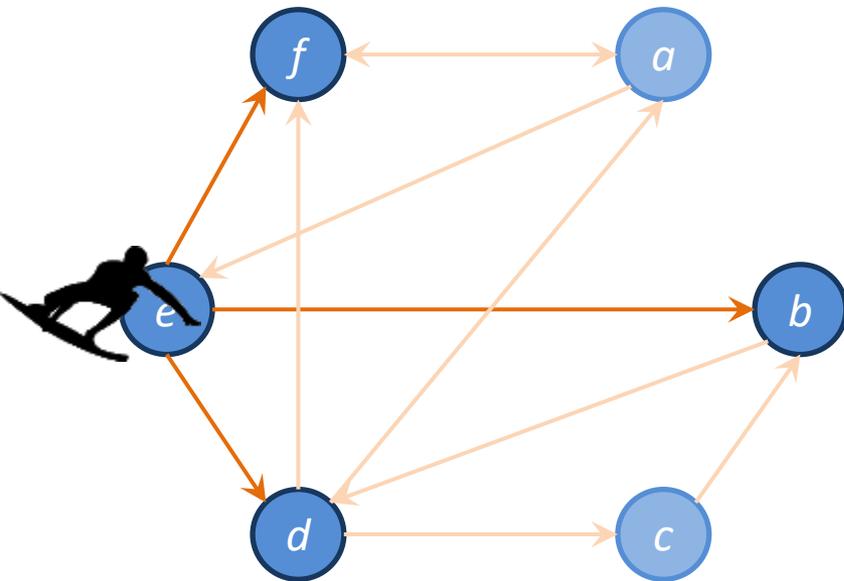
$$\text{rank}_0(v) := \frac{1}{|V|}$$

$$\text{rank}_i(v) := \sum_{v' \in \text{in}(v)} \frac{\text{rank}_{i-1}(v')}{|\text{out}(v')|}$$

PageRank: El modelo del "surfer" aleatorio



= alguien "surfeando" la Web, haciendo clics en enlaces aleatoriamente

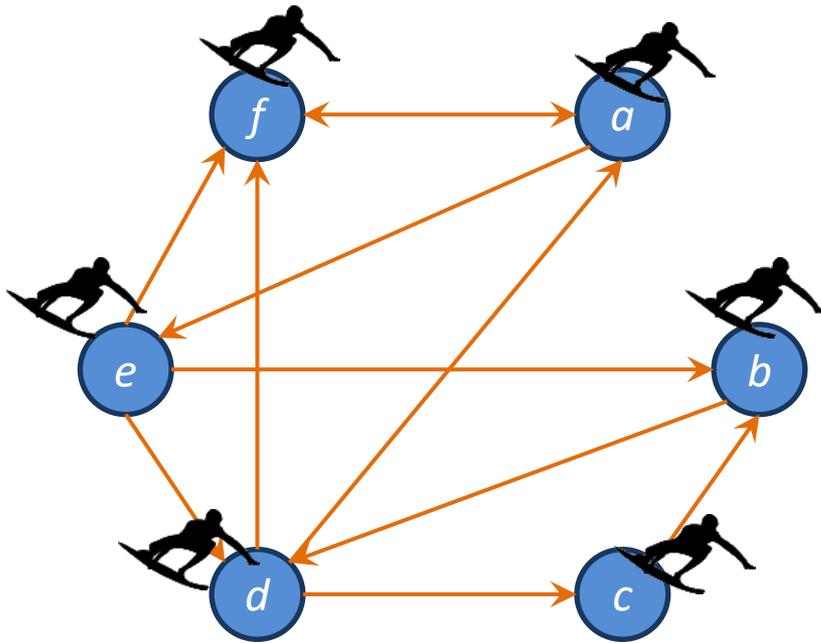


- ¿Cuál es la probabilidad de llegar a una página *x* después de *n* saltos?

PageRank: El modelo del "surfer" aleatorio



= alguien "surfeando" la Web, haciendo clics en enlaces aleatoriamente

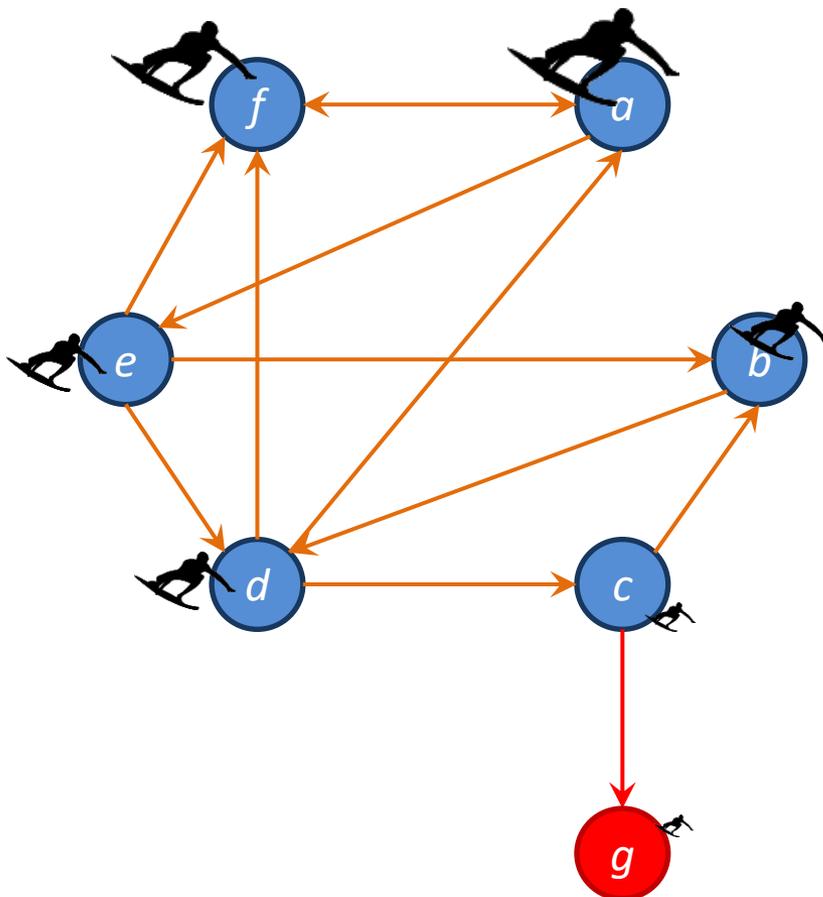


- ¿Cuál es la probabilidad de llegar a una página *x* después de *n* saltos?
- *Estado inicial*: El surfer puede empezar en cualquier nodo

PageRank: El modelo del "surfer" aleatorio



= alguien "surfeando" la Web, haciendo clics en enlaces aleatoriamente



- ¿Cuál es la probabilidad de llegar a una página x después de n saltos?
- *Estado inicial*: El surfer puede empezar en cualquier nodo
- PageRank se aplica iterativamente en cada salto: el puntaje indica la probabilidad de estar en esa página después de esa cantidad de saltos

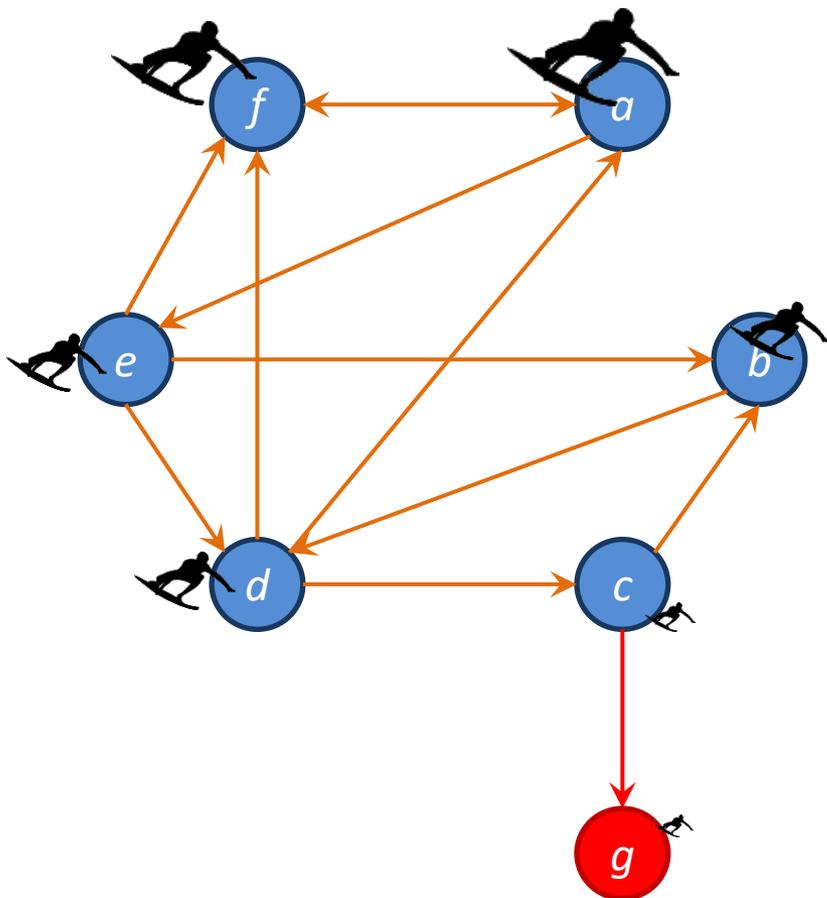
¿Qué pasará con g a medida que pasa el tiempo?



PageRank: El modelo del "surfer" aleatorio



= alguien "surfeando" la Web, haciendo clics en enlaces aleatoriamente

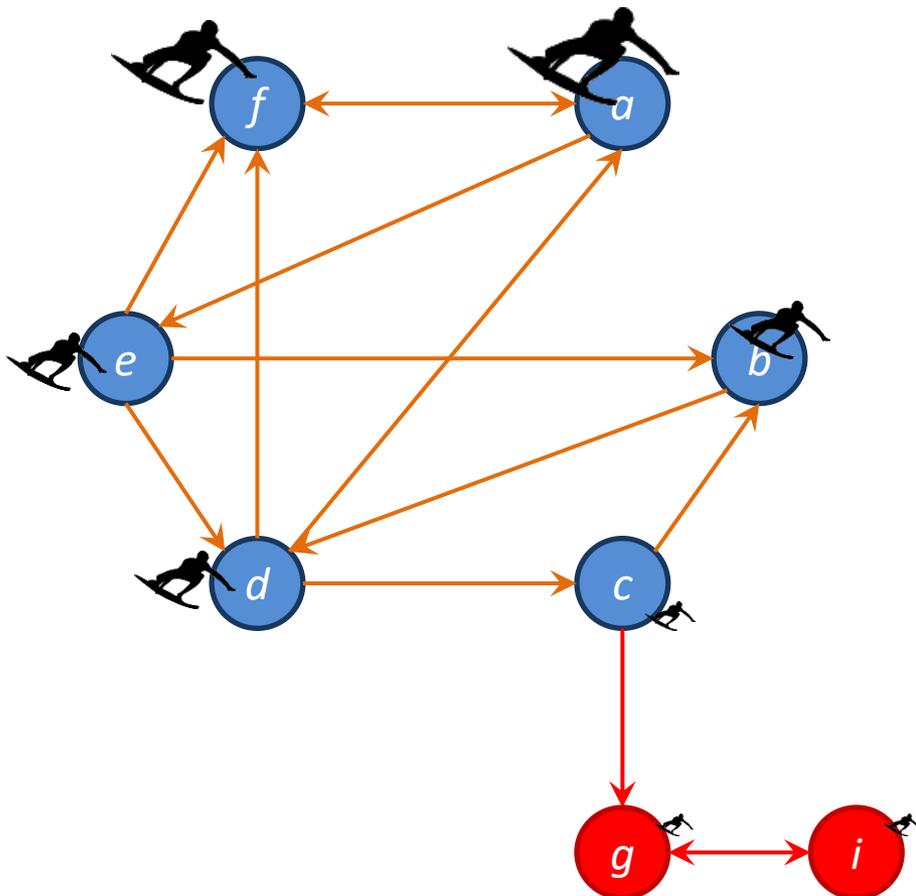


- ¿Cuál es la probabilidad de llegar a una página x después de n saltos?
- *Estado inicial*: El surfer puede empezar en cualquier nodo
- PageRank se aplica iterativamente en cada salto: el puntaje indica la probabilidad de estar en esa página después de esa cantidad de saltos
- Si el surfer llega a una página sin enlaces, salta a otra aleatoriamente

PageRank: El modelo del "surfer" aleatorio



= alguien "surfeando" la Web, haciendo clics en enlaces aleatoriamente



- ¿Cuál es la probabilidad de llegar a una página x después de n saltos?
- *Estado inicial*: El surfer puede empezar en cualquier nodo
- PageRank se aplica iterativamente en cada salto: el puntaje indica la probabilidad de estar en esa página después de esa cantidad de saltos
- Si el surfer llega a una página sin enlaces, salta a otra aleatoriamente

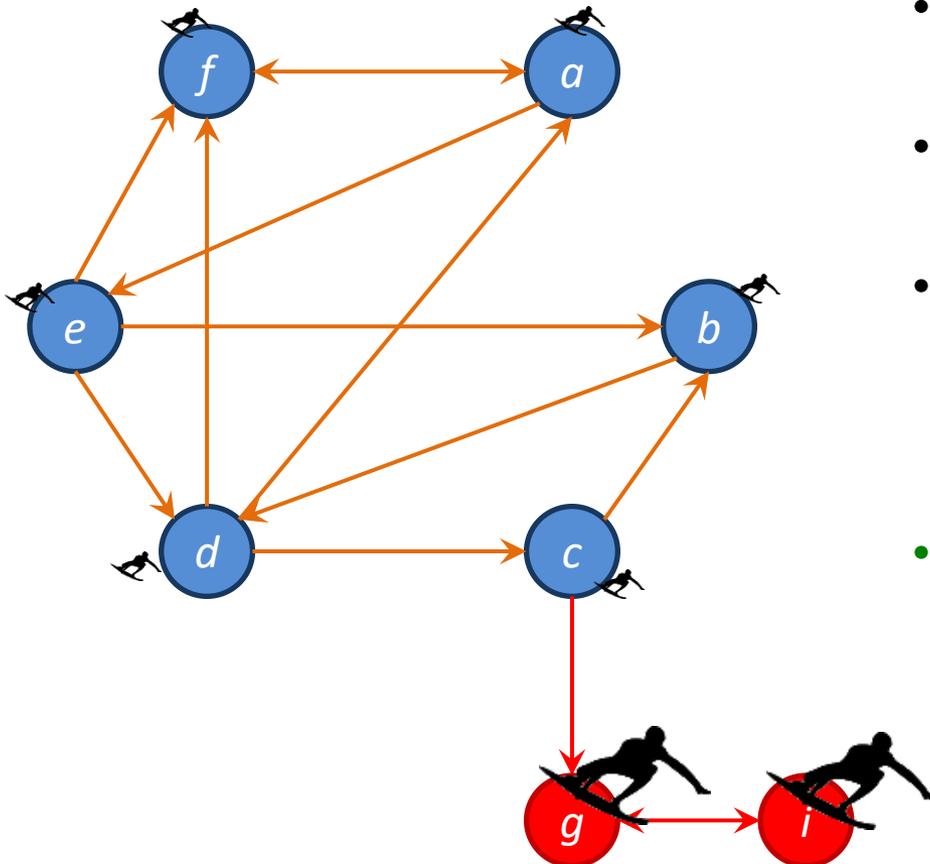
¿Qué pasará con g e i a medida que pasa el tiempo?



PageRank: El modelo del "surfer" aleatorio



= alguien "surfeando" la Web, haciendo clics en enlaces aleatoriamente



- ¿Cuál es la probabilidad de llegar a una página x después de n saltos?
- *Estado inicial*: El surfer puede empezar en cualquier nodo
- PageRank se aplica iterativamente en cada salto: el puntaje indica la probabilidad de estar en esa página después de esa cantidad de saltos
- Si el surfer llega a una página sin enlaces, salta a otra aleatoriamente

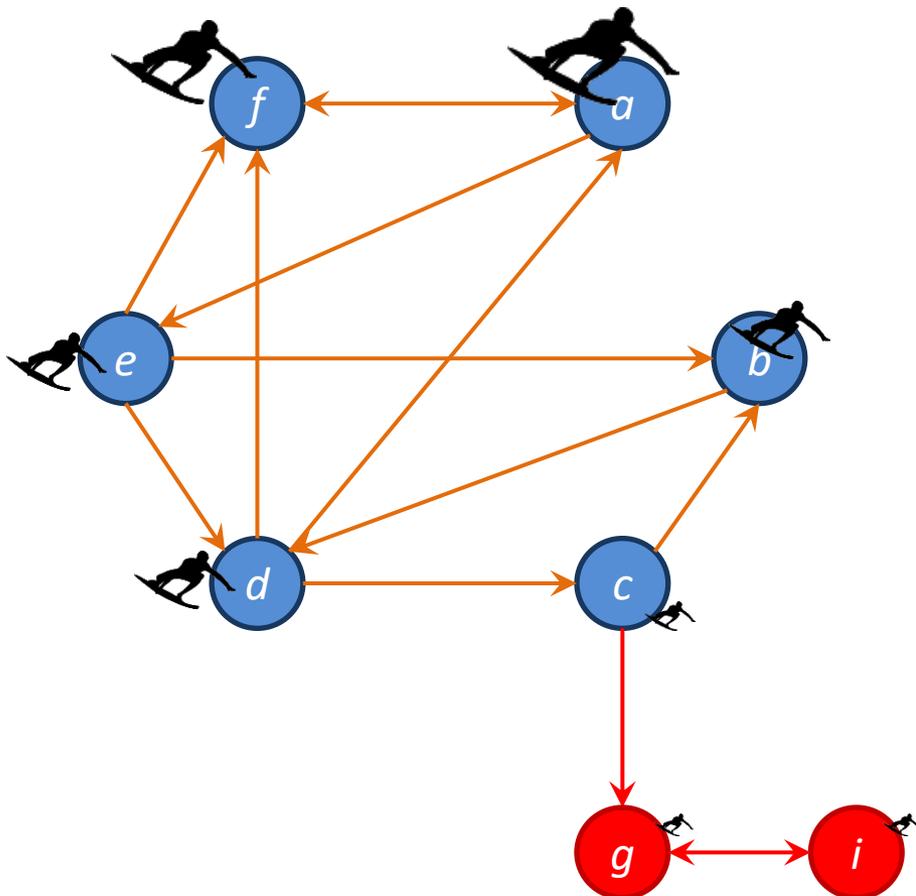
¿Qué pasará con g e i a medida que pasa el tiempo?



PageRank: El modelo del "surfer" aleatorio



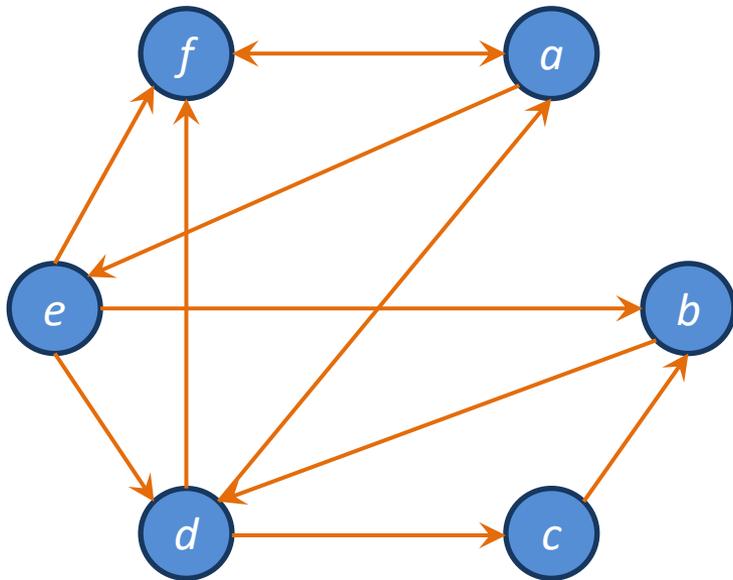
= alguien "surfeando" la Web, haciendo clics en enlaces aleatoriamente



- ¿Cuál es la probabilidad de llegar a una página x después de n saltos?
- *Estado inicial*: El surfer puede empezar en cualquier nodo
- PageRank se aplica iterativamente en cada salto: el puntaje indica la probabilidad de estar en esa página después de esa cantidad de saltos
- Si el surfer llega a una página sin links, salta a otra aleatoriamente
- El surfer saltará a una página aleatoria con probabilidad $1 - d$... *esto evita trampas y asegura convergencia*

La versión final de PageRank

- La Web: Un grafo dirigido



$$G = \boxed{V} \boxed{E}$$

Vértices
(páginas)

Arcos
(enlaces)

$$\text{out}(v) := \{v' \in V \mid (v, v') \in E\}$$

$$\text{in}(v) := \{v' \in V \mid (v', v) \in E\}$$

$$\text{rank}_0(v) := \frac{1}{|V|}$$

$$V' := \{v \in V : |\text{out}(v)| = 0\}$$

$$V'' := \{v \in V : |\text{out}(v)| \neq 0\}$$

d es la probabilidad de seguir un enlace (típicamente $d = 0,85$)

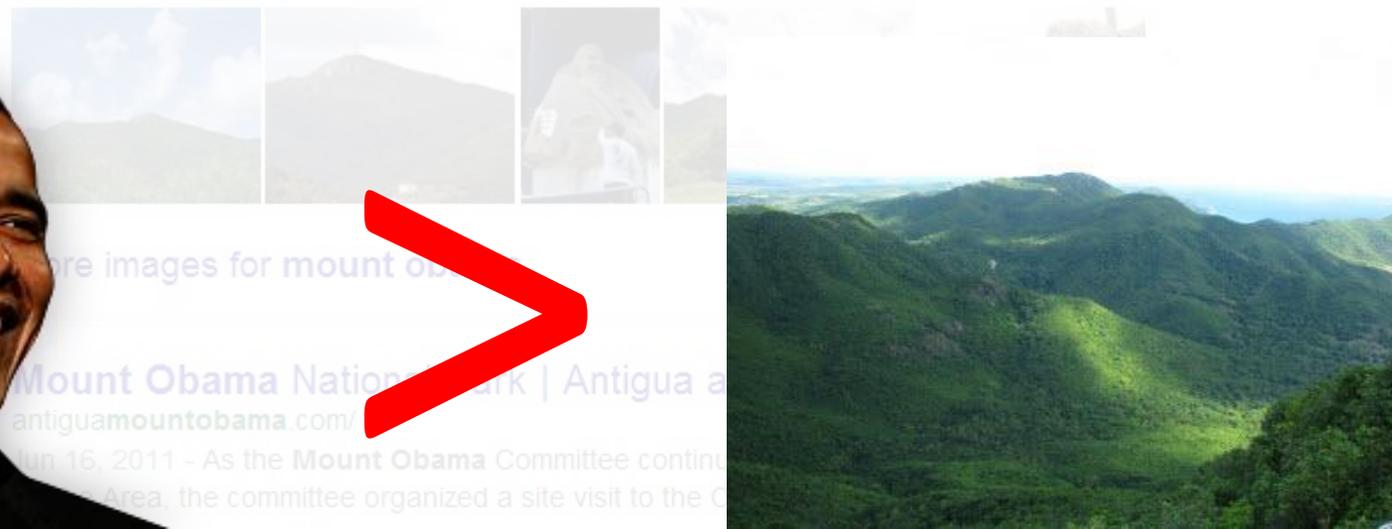
$$\text{rank}_i(v) := d \times \sum_{u \in \text{in}(v)} \frac{\text{rank}_{i-1}(u)}{|\text{out}(u)|} + \sum_{v' \in V'} \frac{\text{rank}_{i-1}(v')}{|V|} + (1-d) \times \sum_{v'' \in V''} \frac{\text{rank}_{i-1}(v'')}{|V|}$$

PageRank: Beneficios



- ✓ Más robusto que un conteo simple de enlaces
- ✓ Menos empates
- ✓ Aproximaciones escalables (para grafos dispersos)
- ✓ Su convergencia es garantizada

Las dos caras del ranking: Importancia



Ranking: ¿Ciencia o arte?



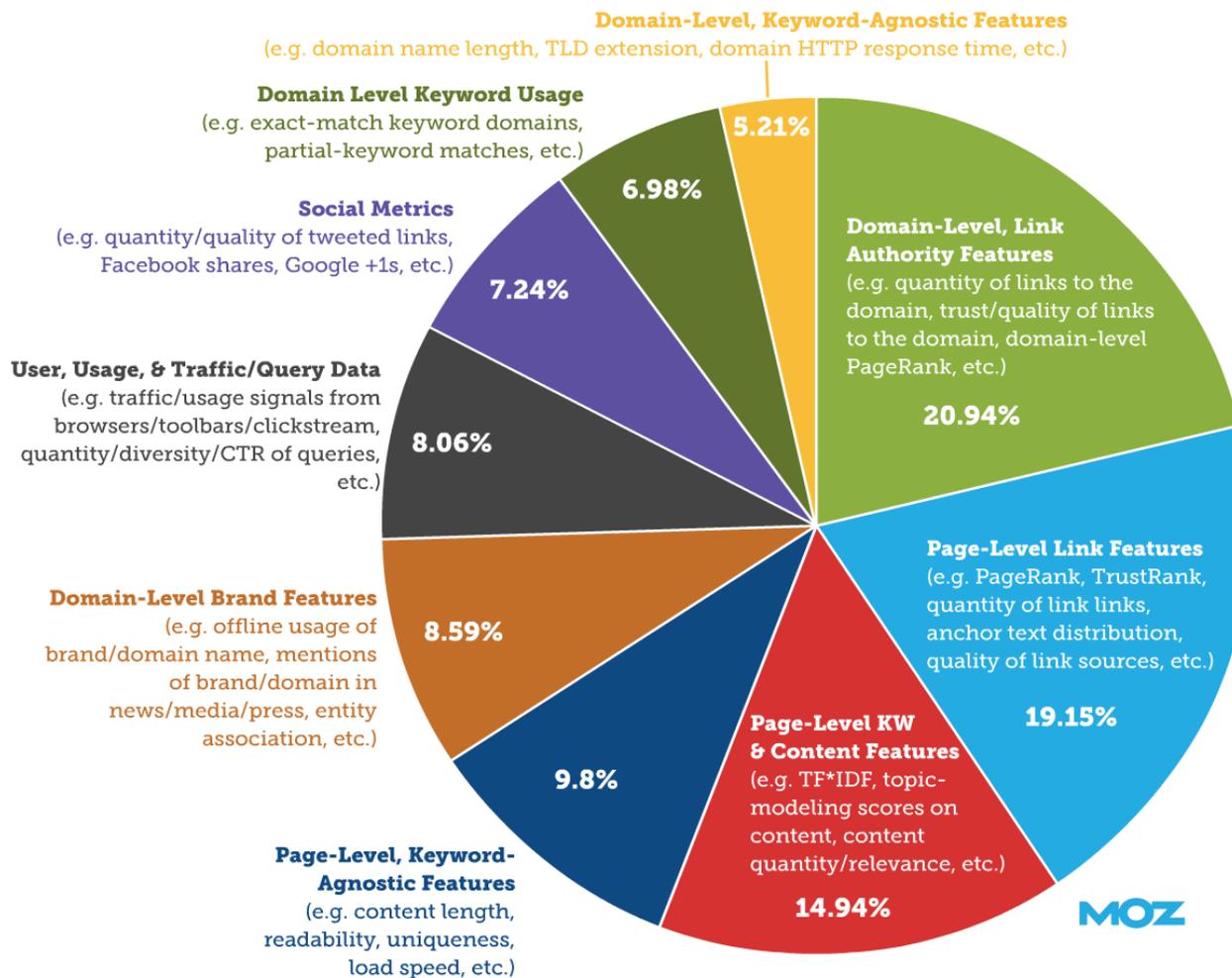
¿CÓMO LO HACE GOOGLE?

LO QUE SABEMOS

¡Según algunos expertos de SEO (optimización de motores de búsqueda), no Google! Cómo Google rankea sus resultados (quizás)

Weighting of Thematic Clusters of Ranking Factors in Google

(based on survey responses by 128 SEO professionals in June 2013)



Cómo Google rankea sus resultados (quizás)

Weighting of Thematic Clusters of Ranking Factors in Google

(based on survey responses by 128 SEO professionals in June 2013)



Ley de Goodhart:

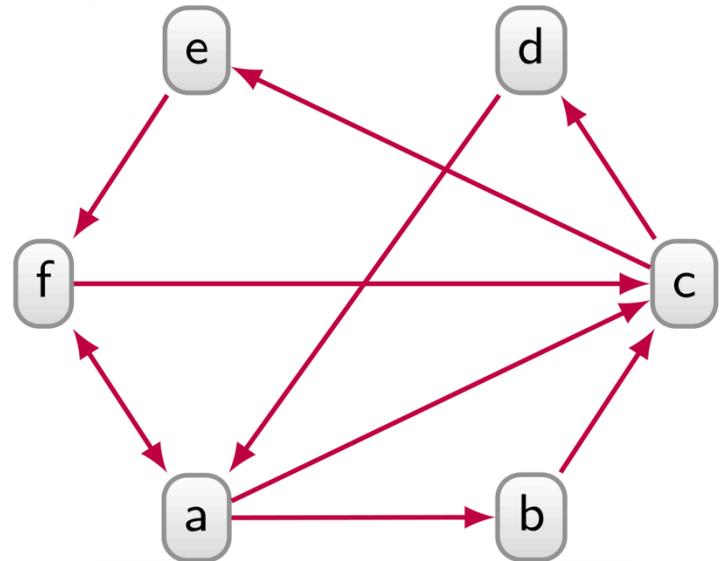
"Cuando una medida se convierte en una meta, deja de ser una buena medida"

¡Según algunos expertos de SEO (optimización de motores de búsqueda), no Google!

DISTRIBUCIÓN DE ALGORITMOS DE GRAFOS

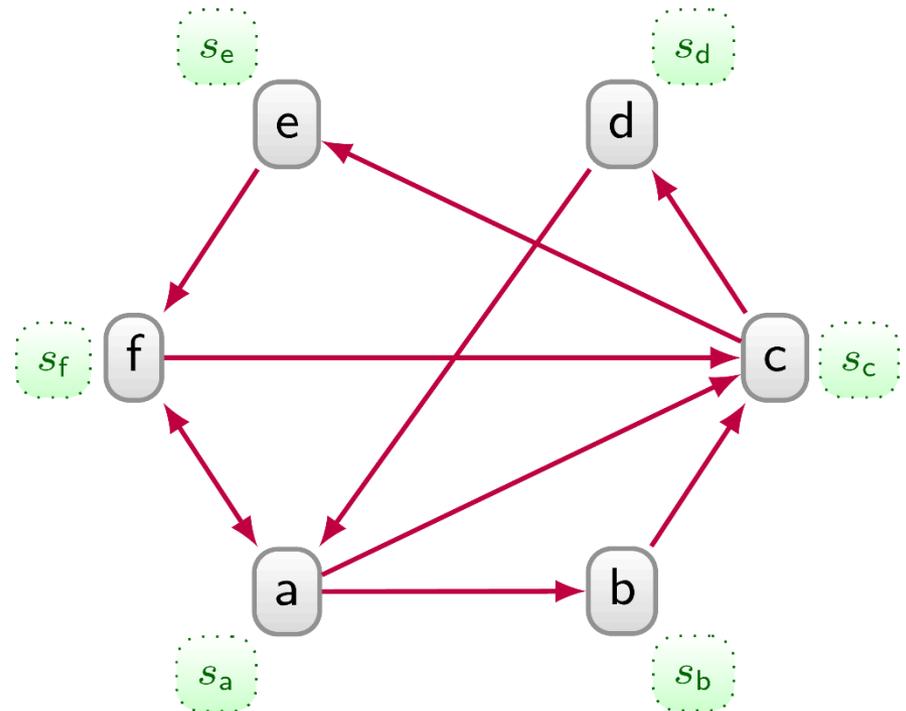
Computación Paralela sobre Grafos

- La entrada es un grafo



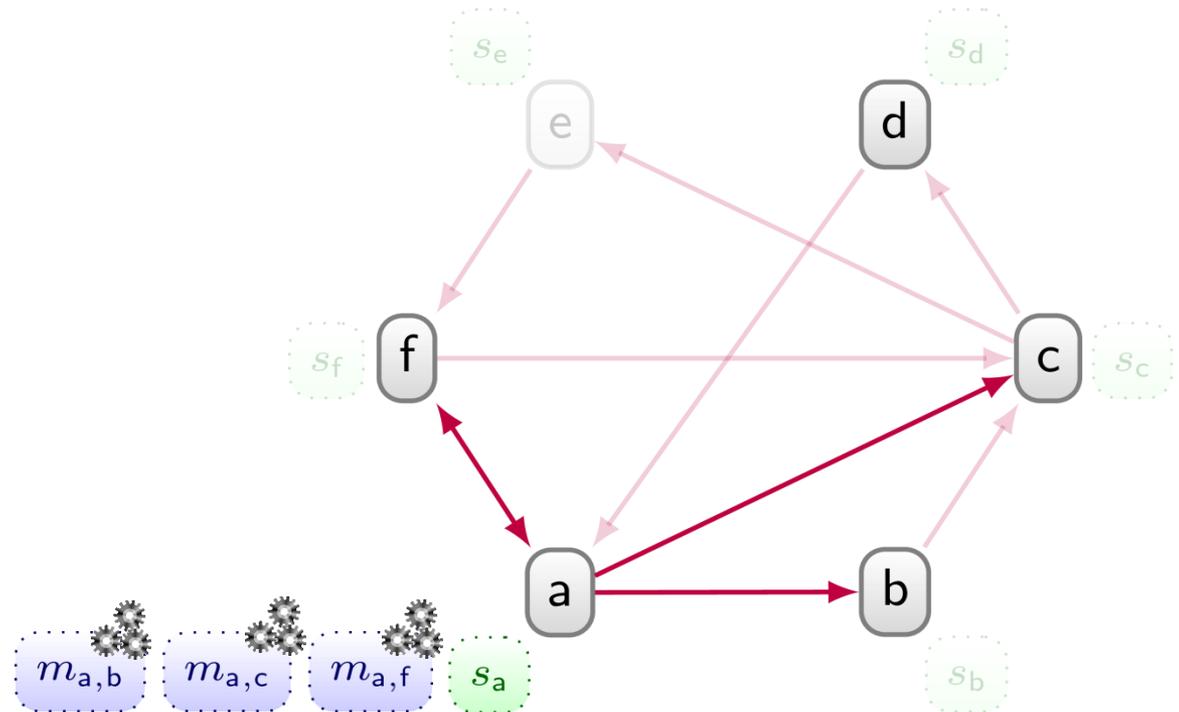
Computación Paralela sobre Grafos

- La entrada es un grafo
- Nodos tienen un estado



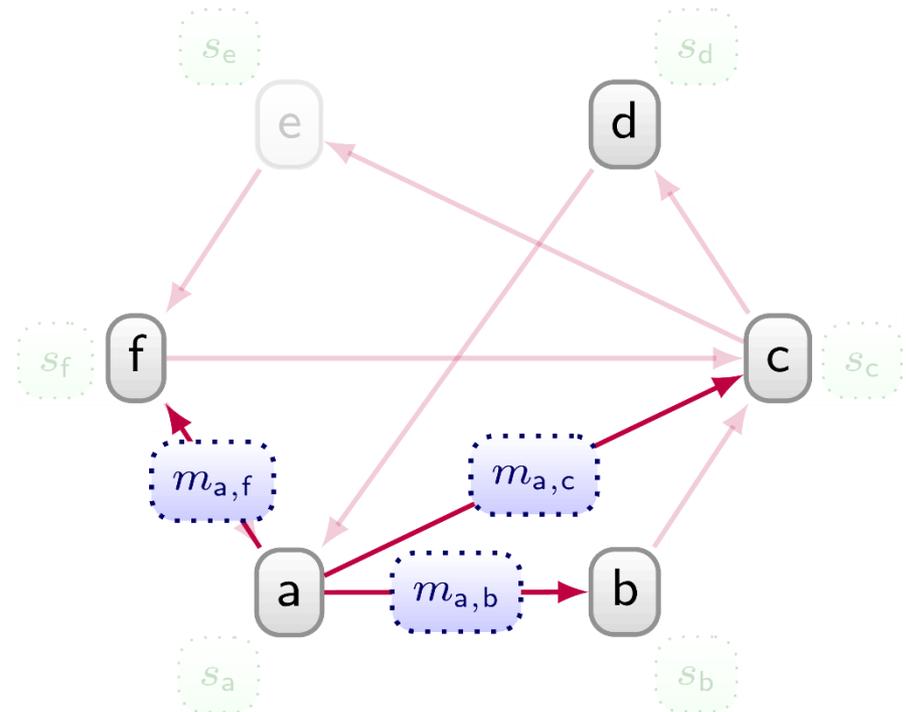
Computación Paralela sobre Grafos

- La entrada es un grafo
- Nodos tienen un estado
- Nodos computan mensajes



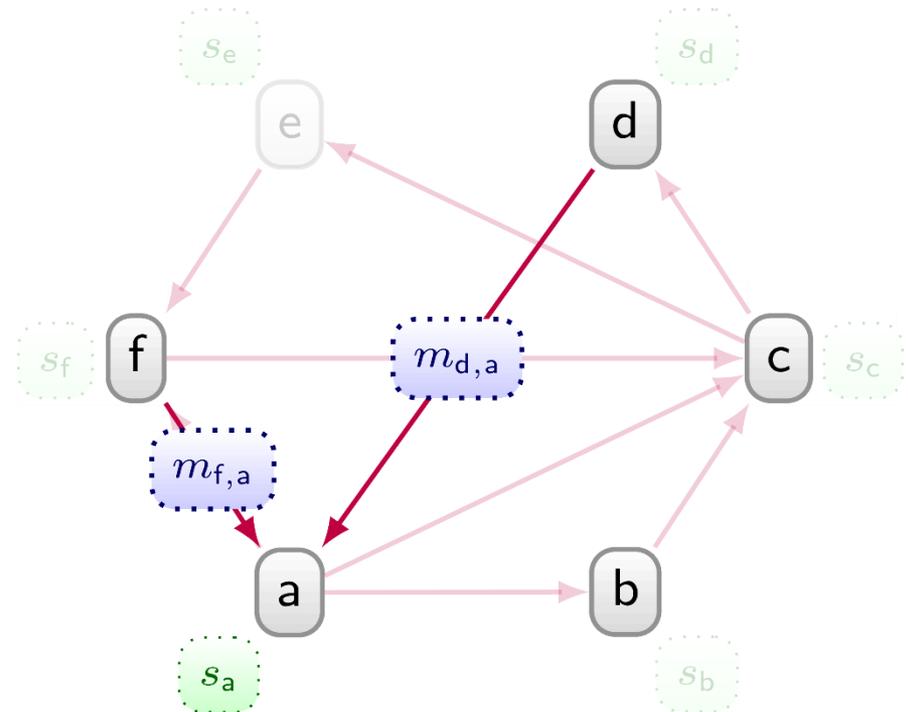
Computación Paralela sobre Grafos

- La entrada es un grafo
- Nodos tienen un estado
- Nodos computan mensajes
- Nodos mandan mensajes a sus vecinos



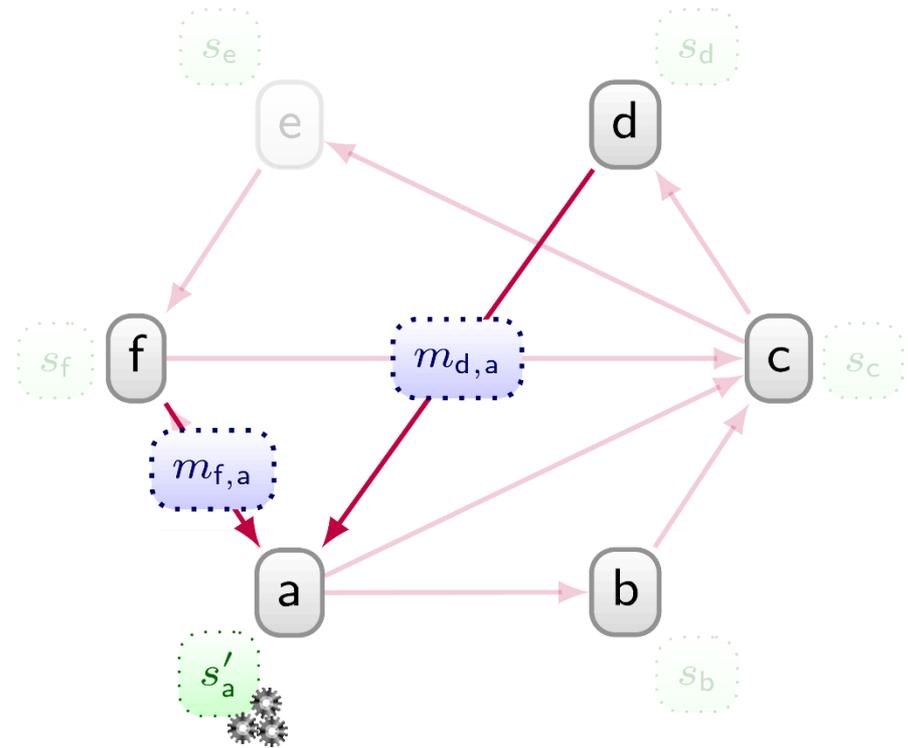
Computación Paralela sobre Grafos

- La entrada es un grafo
- Nodos tienen un estado
- Nodos computan mensajes
- Nodos mandan mensajes a sus vecinos
- Nodos reciben mensajes y computan un nuevo estado



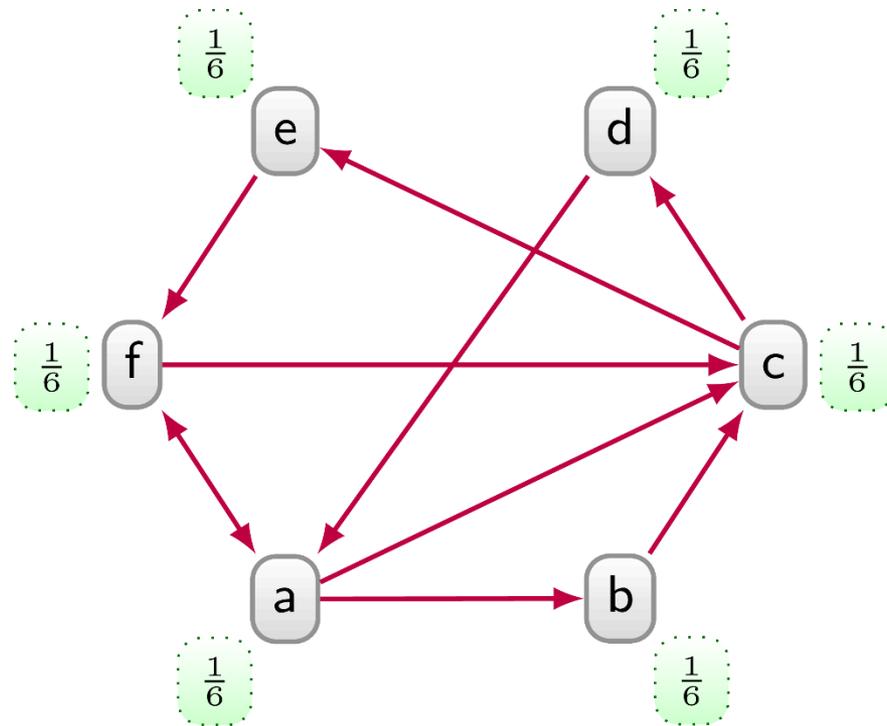
Computación Paralela sobre Grafos

- La entrada es un grafo
- Nodos tienen un estado
- Nodos computan mensajes
- Nodos mandan mensajes a sus vecinos
- Nodos reciben mensajes y computan un nuevo estado
 - ... recursivamente
 - ... (en paralelo)



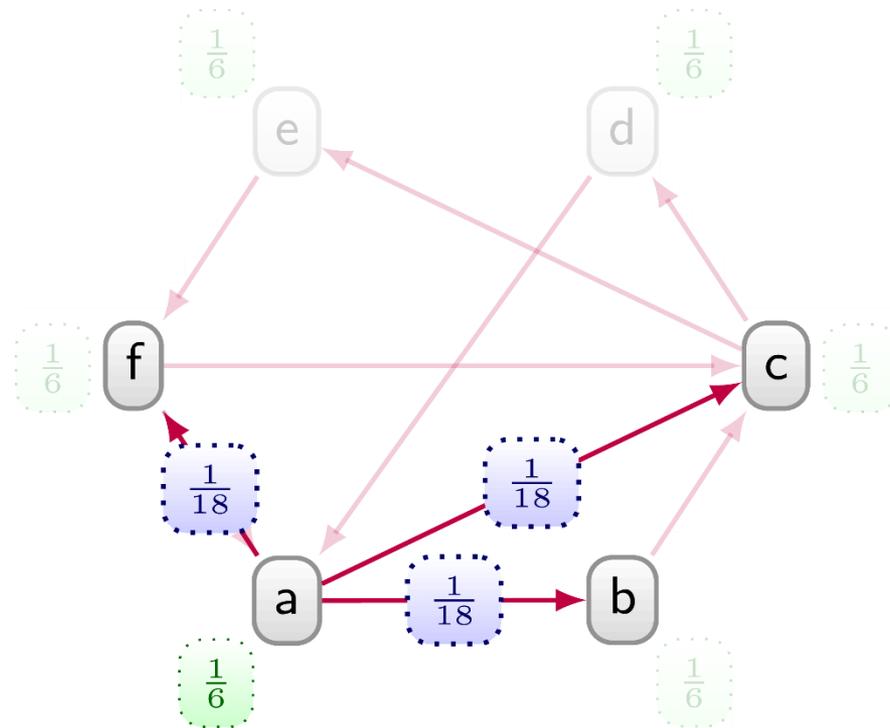
Ejemplo: PageRank (simplificado)

PageRank(G) :



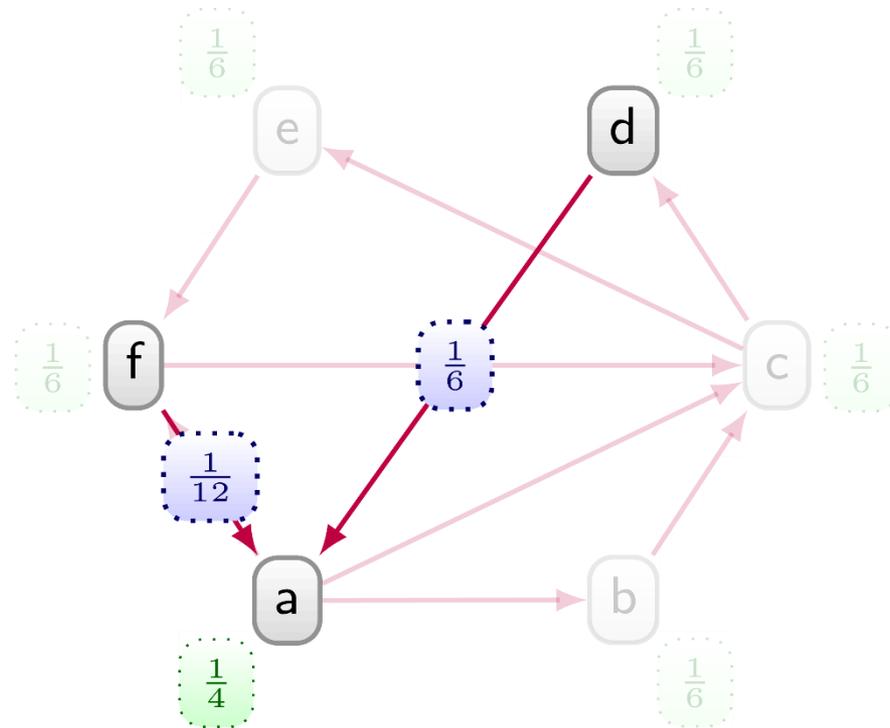
Ejemplo: PageRank (simplificado)

PageRank(G) :



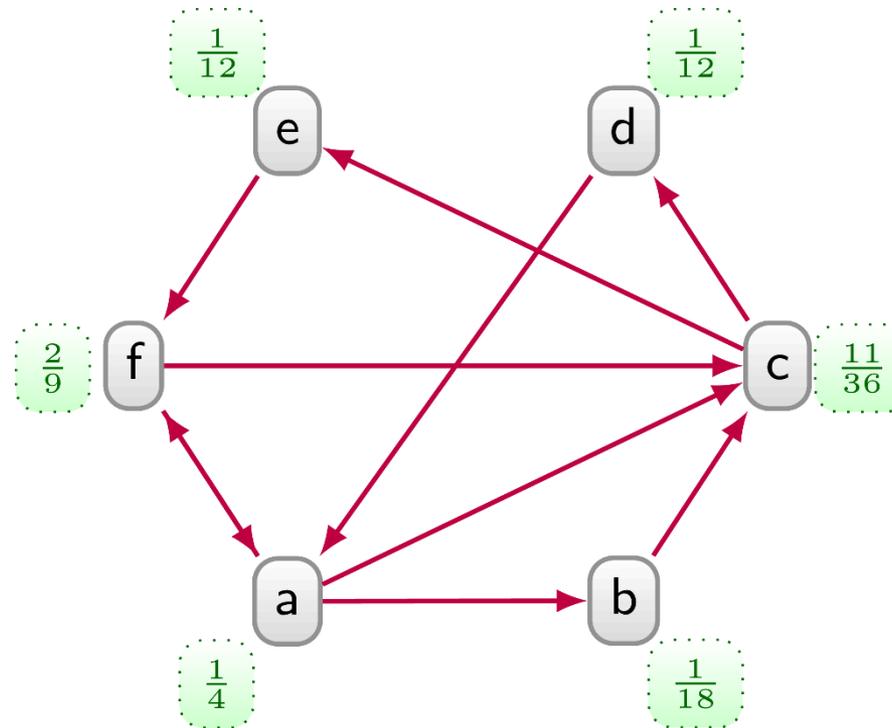
Ejemplo: PageRank (simplificado)

PageRank(G) :



Ejemplo: PageRank (simplificado)

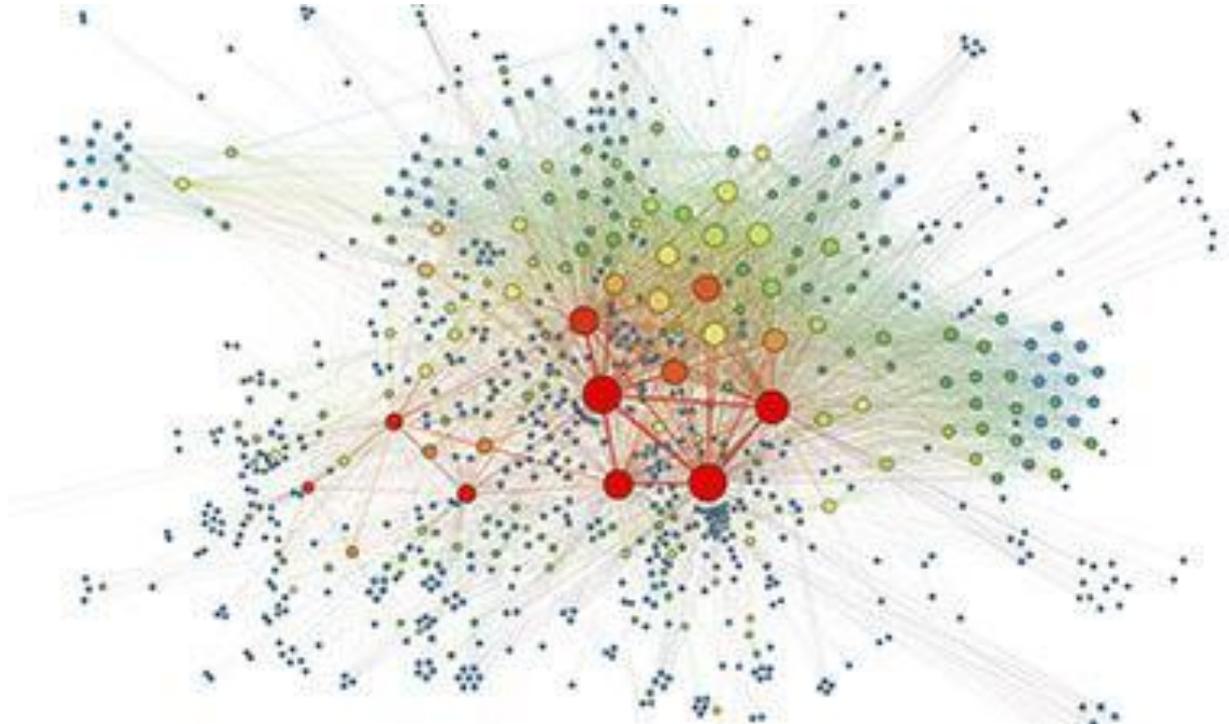
PageRank(G) :



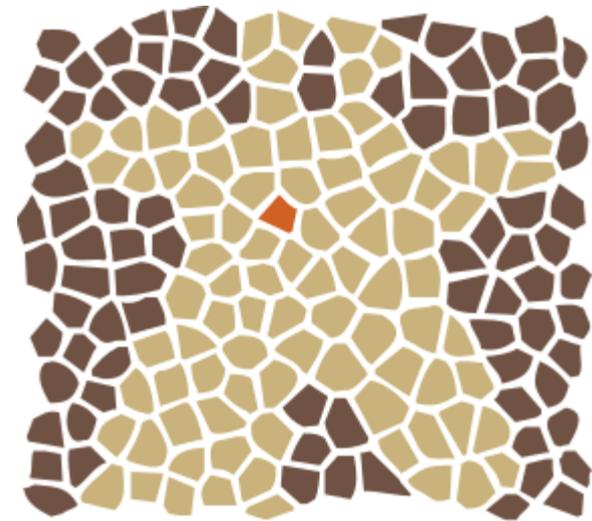
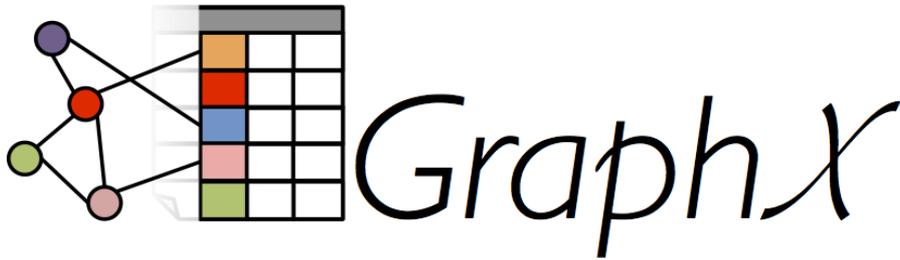
...

Otros algoritmos ...

- Caminos más cortos
- Clustering (k-means, label prop.)
- Centralidad (PageRank, ...)
- Cortes de grafo



Implementaciones distribuidas



A P A C H E
G I R A P H



¿Preguntas?