

Análisis de Redes Complejas: Un paseo matemático entre Google y las redes sociales

Miguel Romance

miguel.romance@urjc.es

Dpto. de Matemática Aplicada
ESCET
Universidad Rey Juan Carlos
Madrid

Murcia, 25 de noviembre de 2010

Contenido

- 1 **Introducción**
 - Motivación y objetivos
 - Un paseo por el análisis de redes complejas

- 2 **Centralidad y poder**
 - Redes Sociales: Centralidad Bonacich
 - Buscadores y Centralidad PageRank
 - Algunos problemas sobre centralidad

Poder en redes sociales

Si tenemos un grupo social



Poder en redes sociales

Si tenemos un grupo social o una red social,



facebook

Poder en redes sociales

Si tenemos un grupo social o una red social,



facebook

¿cómo saber la importancia (o poder) de cada miembro?

Ordenación de páginas web en un buscador

Si hemos realizado una búsqueda en internet,

Ordenación de páginas web en un buscador

Si hemos realizado una búsqueda en internet,



Ordenación de páginas web en un buscador

Si hemos realizado una búsqueda en internet,



¿cómo se **ordenan** las páginas web para presentar los resultados?

Ordenación de páginas web en un buscador

Si hemos realizado una búsqueda en internet,



¿cómo se **ordenan** las páginas web para presentar los resultados?

Para poder resolver matemáticamente estos problemas, emplearemos técnicas del **análisis de redes complejas**.

¿Qué es una red compleja?

Definición

Una **red compleja** (dirigida o no) $G = (X, E)$ es un grafo,

¿Qué es una red compleja?

Definición

Una **red compleja** (dirigida o no) $G = (X, E)$ es un grafo, es decir un par de conjuntos $G = (X, E)$, donde

¿Qué es una red compleja?

Definición

Una **red compleja** (dirigida o no) $G = (X, E)$ es un grafo, es decir un par de conjuntos $G = (X, E)$, donde

- X es finito (lo llamamos conjunto de **nodos**).

¿Qué es una red compleja?

Definición

Una **red compleja** (dirigida o no) $G = (X, E)$ es un grafo, es decir un par de conjuntos $G = (X, E)$, donde

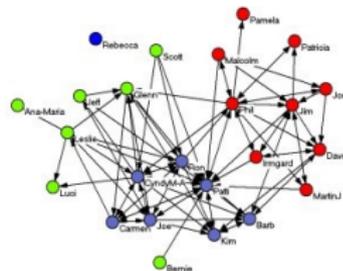
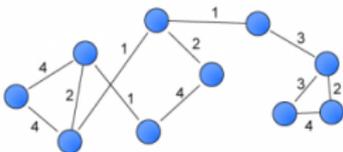
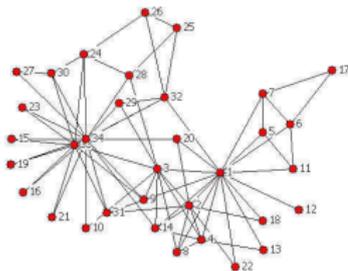
- X es finito (lo llamamos conjunto de **nodos**).
- E es un conjunto de parejas (ordenadas o no) de X (**aristas**).

¿Qué es una red compleja?

Definición

Una **red compleja** (dirigida o no) $G = (X, E)$ es un grafo, es decir un par de conjuntos $G = (X, E)$, donde

- X es finito (lo llamamos conjunto de **nodos**).
- E es un conjunto de parejas (ordenadas o no) de X (**aristas**).



¿Por qué son interesantes las redes complejas?

Gracias a los avances de la informática de los últimos años se pueden calcular propiedades de redes (muy grandes) que corresponden a **aplicaciones reales**, como por ejemplo:

¿Por qué son interesantes las redes complejas?

Gracias a los avances de la informática de los últimos años se pueden calcular propiedades de redes (muy grandes) que corresponden a **aplicaciones reales**, como por ejemplo:

- **Redes de Comunicaciones:**
 - Redes de transporte (redes de metro, de tren, de líneas aéreas).
 - Redes informáticas (redes de servidores, Internet, ...).

¿Por qué son interesantes las redes complejas?

Si consideramos, por ejemplo, el Metro de Madrid,

¿Por qué son interesantes las redes complejas?

Gracias a los avances de la informática de los últimos años se pueden calcular propiedades de redes (muy grandes) que corresponden a **aplicaciones reales**, como por ejemplo:

- **Redes de Comunicaciones:**
 - Redes de transporte (redes de metro, de tren, de líneas aéreas).
 - Redes informáticas (redes de servidores, Internet, ...).

¿Por qué son interesantes las redes complejas?

Gracias a los avances de la informática de los últimos años se pueden calcular propiedades de redes (muy grandes) que corresponden a **aplicaciones reales**, como por ejemplo:

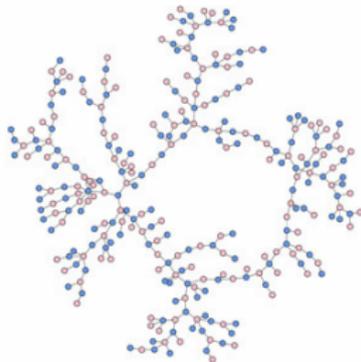
- **Redes de Comunicaciones:**
 - Redes de transporte (redes de metro, de tren, de líneas aéreas).
 - Redes informáticas (redes de servidores, Internet, ...).
- **Redes Biológicas:**
 - Neurociencias (interacción neuronal, redes corticales,...).
 - Redes Bioquímicas (redes metabólicas, genómicas).
 - Redes epidémicas.

¿Por qué son interesantes las redes complejas?

¿Podemos simular (y predecir) la evolución de una **epidemia**?

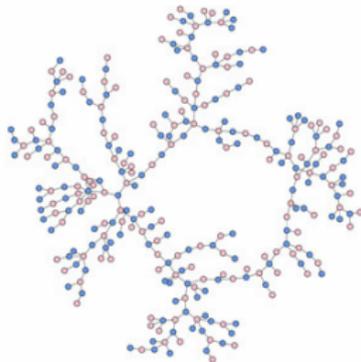
¿Por qué son interesantes las redes complejas?

¿Podemos simular (y predecir) la evolución de una **epidemia**?
Por ejemplo, ¿podemos pronosticar el desarrollo de una **enfermedad sexual**?



¿Por qué son interesantes las redes complejas?

¿Podemos simular (y predecir) la evolución de una **epidemia**?
Por ejemplo, ¿podemos pronosticar el desarrollo de una **enfermedad sexual**?



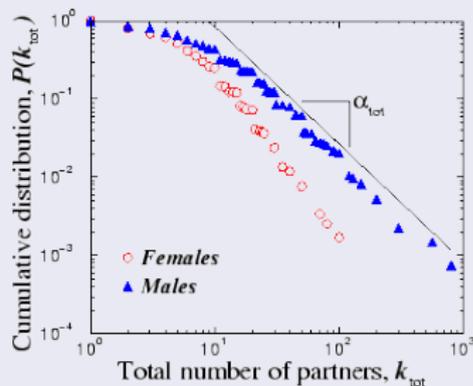
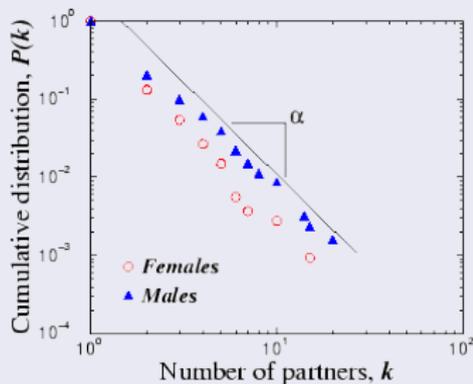
...en muchas situaciones, por privacidad, **no disponemos de la red...**

¿Por qué son interesantes las redes complejas?

Si sólo disponemos de **información local**, como la distribución de los contactos sexuales,

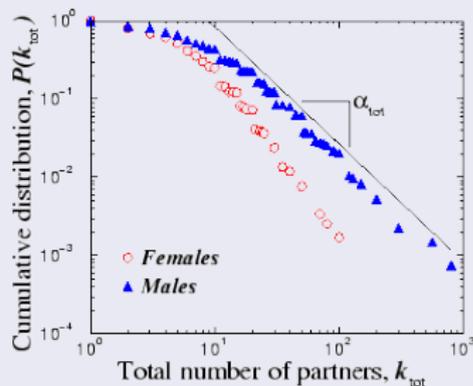
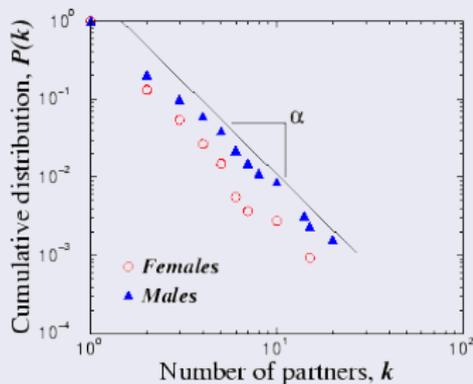
¿Por qué son interesantes las redes complejas?

Si sólo disponemos de **información local**, como la distribución de los contactos sexuales, ¿podemos simular la evolución de una enfermedad sexual?



¿Por qué son interesantes las redes complejas?

Si sólo disponemos de **información local**, como la distribución de los contactos sexuales, ¿podemos simular la evolución de una enfermedad sexual?



¿podemos ahora dar resultados realistas?

¿Por qué son interesantes las redes complejas?

Gracias a los avances de la informática de los últimos años se pueden calcular propiedades de redes (muy grandes) que corresponden a **aplicaciones reales**, como por ejemplo:

- **Redes de Comunicaciones:**
 - Redes de transporte (redes de metro, de tren, de líneas aéreas).
 - Redes informáticas (redes de servidores, Internet, ...).
- **Redes Biológicas:**
 - Neurociencias (interacción neuronal, redes corticales,...).
 - Redes Bioquímicas (redes metabólicas, genómicas).
 - Redes epidémicas.

¿Por qué son interesantes las redes complejas?

Gracias a los avances de la informática de los últimos años se pueden calcular propiedades de redes (muy grandes) que corresponden a **aplicaciones reales**, como por ejemplo:

- **Redes de Comunicaciones:**
 - Redes de transporte (redes de metro, de tren, de líneas aéreas).
 - Redes informáticas (redes de servidores, Internet, ...).
- **Redes Biológicas:**
 - Neurociencias (interacción neuronal, redes corticales,...).
 - Redes Bioquímicas (redes metabólicas, genómicas).
 - Redes epidémicas.
- **Redes Sociales y Económicas:**
 - Redes Sociales.
 - Redes de información.
 - Redes de recomendación y teoría de juegos.

Herramientas matemáticas en análisis de redes complejas

Las herramientas matemáticas empleadas son muy diversas y proceden de campos muy diferentes, como por ejemplo:

Herramientas matemáticas en análisis de redes complejas

Las herramientas matemáticas empleadas son muy diversas y proceden de campos muy diferentes, como por ejemplo:

- **Matemática Discreta** (Teoría de grafos y combinatoria)

Herramientas matemáticas en análisis de redes complejas

Las herramientas matemáticas empleadas son muy diversas y proceden de campos muy diferentes, como por ejemplo:

- **Matemática Discreta** (Teoría de grafos y combinatoria)
- **Análisis Matricial** (Álgebra lineal y teoría de Perron-Frobenius)
- **Ciencias de la computación**

Herramientas matemáticas en análisis de redes complejas

Las herramientas matemáticas empleadas son muy diversas y proceden de campos muy diferentes, como por ejemplo:

- **Matemática Discreta** (Teoría de grafos y combinatoria)
- **Análisis Matricial** (Álgebra lineal y teoría de Perron-Frobenius)
- **Ciencias de la computación**
- **Probabilidad** (matrices aleatorias, procesos estocásticos)
- **Ecuaciones diferenciales no lineales**
- **Mecánica Estadística**

Herramientas matemáticas en análisis de redes complejas

Las herramientas matemáticas empleadas son muy diversas y proceden de campos muy diferentes, como por ejemplo:

- **Matemática Discreta** (Teoría de grafos y combinatoria)
- **Análisis Matricial** (Álgebra lineal y teoría de Perron-Frobenius)
- **Ciencias de la computación**
- **Probabilidad** (matrices aleatorias, procesos estocásticos)
- **Ecuaciones diferenciales no lineales**
- **Mecánica Estadística**

e involucra a científicos de diferentes áreas, como por ejemplo:

- Físicos teóricos,
- Informáticos,
- Biólogos y neuro-científicos,
- Sociólogos and psicólogos,

Herramientas matemáticas en análisis de redes complejas

Las herramientas matemáticas empleadas son muy diversas y proceden de campos muy diferentes, como por ejemplo:

- **Matemática Discreta** (Teoría de grafos y combinatoria)
- **Análisis Matricial** (Álgebra lineal y teoría de Perron-Frobenius)
- **Ciencias de la computación**
- **Probabilidad** (matrices aleatorias, procesos estocásticos)
- **Ecuaciones diferenciales no lineales**
- **Mecánica Estadística**

e involucra a científicos de diferentes áreas, como por ejemplo:

- Físicos teóricos,
- Informáticos,
- Biólogos y neuro-científicos,
- Sociólogos and psicólogos,
- ...y, por supuesto, **¡también matemáticos!**

Principales líneas de trabajo

Los principales campos de trabajo en la actualidad son:

Principales líneas de trabajo

Los principales campos de trabajo en la actualidad son:

- **Análisis estructural** de redes complejas.

Principales líneas de trabajo

Los principales campos de trabajo en la actualidad son:

- **Análisis estructural** de redes complejas.
 - Estructura global vs. estructura local.
 - Análisis espectral y centralidad.

Principales líneas de trabajo

Los principales campos de trabajo en la actualidad son:

- **Análisis estructural** de redes complejas.
 - Estructura global vs. estructura local.
 - Análisis espectral y centralidad.
 - Detección de comunidades.
 - Problemas de diseño y mejora óptima.

Principales líneas de trabajo

Los principales campos de trabajo en la actualidad son:

- **Análisis estructural** de redes complejas.
 - Estructura global vs. estructura local.
 - Análisis espectral y centralidad.
 - Detección de comunidades.
 - Problemas de diseño y mejora óptima.
- **Modelos aleatorios** en redes complejas.
 - Modelo de Erdős-Rényi y redes aleatorias clásicas.
 - Modelo de Albert-Barábasi y el **preferential attachment**.

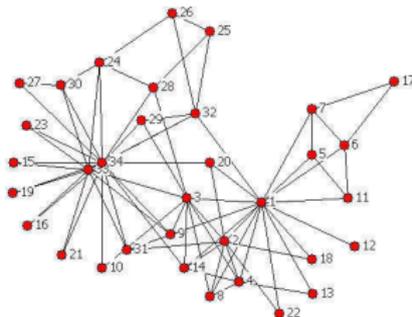
Principales líneas de trabajo

Los principales campos de trabajo en la actualidad son:

- **Análisis estructural** de redes complejas.
 - Estructura global vs. estructura local.
 - Análisis espectral y centralidad.
 - Detección de comunidades.
 - Problemas de diseño y mejora óptima.
- **Modelos aleatorios** en redes complejas.
 - Modelo de Ęrdos-Rénji y redes aleatorias clásicas.
 - Modelo de Albert-Barábasi y el **preferential attachment**.
- **Dinámica** en redes complejas.
 - Sincronización.
 - Propiedades estructurales vs. propiedades dinámicas.

Centralidad y poder

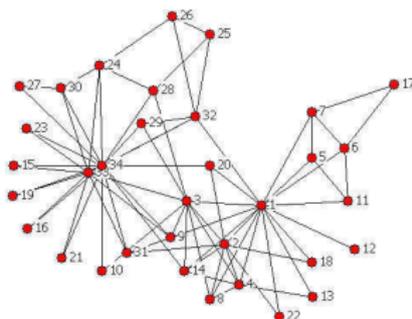
Si tenemos una red compleja $G = (X, E)$,



...no todos los nodos juegan el mismo papel ni tienen la misma **influencia** en toda la red.

Centralidad y poder

Si tenemos una red compleja $G = (X, E)$,



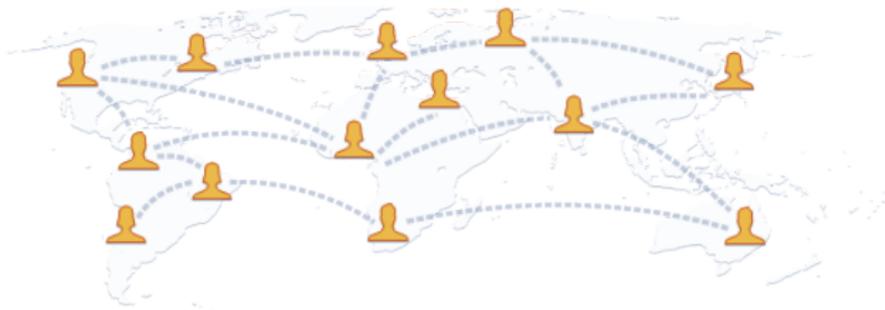
...no todos los nodos juegan el mismo papel ni tienen la misma **influencia** en toda la red.

Pregunta

¿Cómo podemos medir la **importancia** de cada nodo de la red?

Centralidad y poder

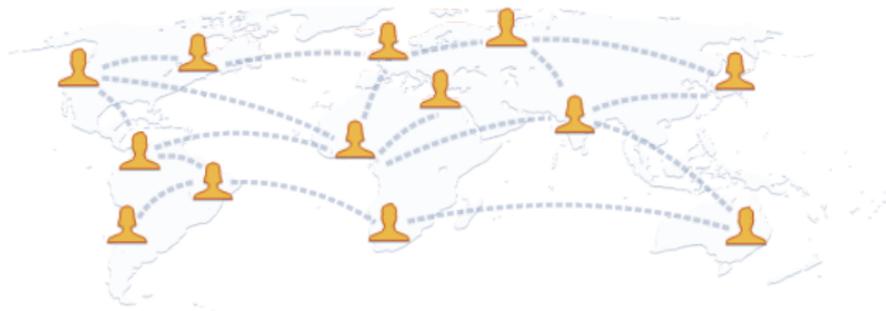
Si tenemos una **red social**



podemos modelarla como una red $G = (X, E)$, donde los **nodos** $X = \{1, \dots, n\}$ son los **miembros de la red** y los **enlaces** son las **interacciones** (amistad, relación, etc.) entre los miembros.

Centralidad y poder

Si tenemos una **red social**



podemos modelarla como una red $G = (X, E)$, donde los **nodos** $X = \{1, \dots, n\}$ son los **miembros de la red** y los **enlaces** son las **interacciones** (amistad, relación, etc.) entre los miembros.

¿Cómo podemos definir $C : X \rightarrow [0, +\infty)$ que mida la **influencia** de cada nodo en la red?

Centralidad y poder

En sociología la centralidad (y el poder) está asociado con la **cantidad y calidad de los contactos** de cada individuo.

Centralidad y poder

En sociología la centralidad (y el poder) está asociado con la **cantidad y calidad de los contactos** de cada individuo.

Definición

Si $G = (X, E)$ es una red de n nodos, y tomamos $i \in X = \{1, \dots, n\}$, se define el **grado (de entrada)** de i como el número de enlaces que terminan en i (es decir de la forma $j \rightarrow i$). Este valor se denota como $gr_{in}(i)$.

Centralidad y poder

En sociología la centralidad (y el poder) está asociado con la **cantidad y calidad de los contactos** de cada individuo.

Definición

Si $G = (X, E)$ es una red de n nodos, y tomamos $i \in X = \{1, \dots, n\}$, se define el **grado (de entrada)** de i como el número de enlaces que terminan en i (es decir de la forma $j \rightarrow i$). Este valor se denota como $gr_{in}(i)$.



Este concepto solo mide el **número de contactos**, pero no su **calidad**.

Centralidad y poder

Pregunta de sociología

Por ejemplo, si consideramos la red social **Facebook**, ¿podemos **ordenar** las siguientes personas por su **influencia** (poder) en la red social?

Figura: ¿Quién es el **jefe** de la red?

Centralidad y poder

Pregunta de sociología

Por ejemplo, si consideramos la red social **Facebook**, ¿podemos **ordenar** las siguientes personas por su **influencia** (poder) en la red social?



Figura: ¿Quién es el **jefe** de la red?

Centralidad y poder

Pregunta de sociología

Por ejemplo, si consideramos la red social Facebook, ¿podemos **ordenar** las siguientes personas por su **influencia** (poder) en la red social?



Figura: ¿Quién es el **jefe** de la red?

Centralidad y poder

Pregunta de sociología

Por ejemplo, si consideramos la red social Facebook, ¿podemos **ordenar** las siguientes personas por su **influencia** (poder) en la red social?



Figura: ¿Quién es el **jefe** de la red?

Centralidad y poder

Pregunta de sociología

Por ejemplo, si consideramos la red social **Facebook**, ¿podemos **ordenar** las siguientes personas por su **influencia** (poder) en la red social?



Figura: ¿Quién es el **jefe** de la red?

Centralidad y poder

...en este caso el número de contactos no parece buena idea...

Centralidad y poder

...en este caso el número de contactos no parece buena idea...



The image shows a screenshot of a Facebook profile for José Luis Rodríguez Zapatero. The profile includes a profile picture, a cover photo, and various sections such as 'About Me', 'Basic Info', 'Relationship Status', 'Children', 'Interested In', 'Looking For', and 'Bios'. The 'Friends' section shows 47 friends.

facebook

Search

José Luis Rodríguez Zapatero

Wall Info

People who aren't friends with José Luis see José Luis personally, send him a message or add him

About Me

Basic Info	Sex:	Male
	Children:	Laura Rodríguez Alba Rodríguez
	Relationship Status:	Married
	Interested In:	Women, Men
	Looking For:	Friendship, Networking

Bios

I studied Law at the University of Seville. I was above average but not a genius.

Friends

47 friends See All

Centralidad y poder

...en este caso el número de contactos no parece buena idea...

facebook

José Luis Rodríguez Zapatero

Wall Info

People who aren't friends with José Luis see José Luis personally, send him a message or add him

About Me

Basic Info	Sex:	Male
Children:	Laura Rodríguez Alba	Married
Relationship Status:	Married	
Interested In:	Women	Friendship Networking
Looking For:	Friendship Networking	
Bio:	I studied Law at the University student was above average be	

Information

Relationship Status: Married

Children: Laura Rodríguez Espinosa Alba Rodríguez Espinosa

Friends

47 friends See All

facebook

Mariano Rajoy

Info

People who aren't friends with Mariano see Mariano personally, send him a message or add him

About Me

Basic Info	Sex:
Current	

Friends

4,904 friends See All

Popular: San Juan, Chelo Fernández Palomares, David Portales del Rey

Likes and Interests

Interests: Clds

Centralidad y poder

...en este caso el número de contactos no parece buena idea...

The image shows two Facebook profiles side-by-side. The top profile is for José Luis Rodríguez Zapatero, with 47 friends. The bottom profile is for Barack Obama, with 699 friends. Both profiles show a 'Basic Info' section with fields for Sex, Children, Relationship Status, Interested In, and Looking For. The Obama profile also shows a 'Wall' post: 'Well I Am The 44th President Of United States Of America'.

The image shows a Facebook profile for Mariano Rajoy, with 4,904 friends. The profile includes a 'Basic Info' section with fields for Sex and Current City. There is also a 'Friends' section showing three friends: Popularis San Juan, Chelo Fernández Palomares, and David Portales del Rey.

Centralidad y poder

...en este caso el número de contactos no parece buena idea...



Idea Sociológica

En una red social tenemos que tener en cuenta tanto la **cantidad** como la **calidad** de los enlaces si queremos medir apropiadamente la importancia de un nodo, es decir

Idea Sociológica

En una red social tenemos que tener en cuenta tanto la **cantidad** como la **calidad** de los enlaces si queremos medir apropiadamente la importancia de un nodo, es decir

Heurística Sociológica

Si $G = (X, E)$ es una red compleja con nodos $X = \{1, \dots, n\}$, buscamos una función $c : X \rightarrow \mathbb{R}$ t.q.:

Idea Sociológica

En una red social tenemos que tener en cuenta tanto la **cantidad** como la **calidad** de los enlaces si queremos medir apropiadamente la importancia de un nodo, es decir

Heurística Sociológica

Si $G = (X, E)$ es una red compleja con nodos $X = \{1, \dots, n\}$, buscamos una función $c : X \rightarrow \mathbb{R}$ t.q.:

- $c(i) \geq 0$ para todos los nodos $i \in X$.

Idea Sociológica

En una red social tenemos que tener en cuenta tanto la **cantidad** como la **calidad** de los enlaces si queremos medir apropiadamente la importancia de un nodo, es decir

Heurística Sociológica

Si $G = (X, E)$ es una red compleja con nodos $X = \{1, \dots, n\}$, buscamos una función $c : X \rightarrow \mathbb{R}$ t.q.:

- $c(i) \geq 0$ para todos los nodos $i \in X$.
- $c(1) + \dots + c(n) = 1$.

Idea Sociológica

En una red social tenemos que tener en cuenta tanto la **cantidad** como la **calidad** de los enlaces si queremos medir apropiadamente la importancia de un nodo, es decir

Heurística Sociológica

Si $G = (X, E)$ es una red compleja con nodos $X = \{1, \dots, n\}$, buscamos una función $c : X \rightarrow \mathbb{R}$ t.q.:

- $c(i) \geq 0$ para todos los nodos $i \in X$.
- $c(1) + \dots + c(n) = 1$.
- existe $\lambda > 0$ t.q. $c(i) = \frac{1}{\lambda} \sum_{j \rightarrow i} c(j)$.

Idea Sociológica

En una red social tenemos que tener en cuenta tanto la **cantidad** como la **calidad** de los enlaces si queremos medir apropiadamente la importancia de un nodo, es decir

Heurística Sociológica

Si $G = (X, E)$ es una red compleja con nodos $X = \{1, \dots, n\}$, buscamos una función $c : X \rightarrow \mathbb{R}$ t.q.:

- $c(i) \geq 0$ para todos los nodos $i \in X$.
- $c(1) + \dots + c(n) = 1$.
- existe $\lambda > 0$ t.q. $c(i) = \frac{1}{\lambda} \sum_{j \rightarrow i} c(j)$.

¿Cómo podemos **localizar** una de estas funciones (si existen)?

Reformulación Matemática

Consideremos la **matriz de adyacencia** $A(G) = (a_{ij}) \in M_{n \times n}$ del grafo G dada por

$$a_{ij} = \begin{cases} 1, & \text{si existe una arista } i \rightarrow j, \\ 0, & \text{en otro caso.} \end{cases}$$

Entonces la heurística anterior se re-escribe de **forma vectorial** del siguiente modo:

Reformulación Matemática

Forma vectorial/matricial

Si $G = (X, E)$ es una red con nodos $X = \{1, \dots, n\}$, buscamos un vector $c = (c(1), \dots, c(n))$ tal que:

Reformulación Matemática

Forma vectorial/matricial

Si $G = (X, E)$ es una red con nodos $X = \{1, \dots, n\}$, buscamos un vector $c = (c(1), \dots, c(n))$ tal que:

- $c(i) \geq 0$ para todo $i \in X$, es decir $c \geq 0$.

Reformulación Matemática

Forma vectorial/matricial

Si $G = (X, E)$ es una red con nodos $X = \{1, \dots, n\}$, buscamos un vector $c = (c(1), \dots, c(n))$ tal que:

- $c(i) \geq 0$ para todo $i \in X$, es decir $c \geq 0$.
- $c(1) + \dots + c(n) = \|c\|_1 = 1$.

Reformulación Matemática

Forma vectorial/matricial

Si $G = (X, E)$ es una red con nodos $X = \{1, \dots, n\}$, buscamos un vector $c = (c(1), \dots, c(n))$ tal que:

- $c(i) \geq 0$ para todo $i \in X$, es decir $c \geq 0$.
- $c(1) + \dots + c(n) = \|c\|_1 = 1$.
- existe $\lambda > 0$ tal que para todo $i \in X$

$$c(i) = \frac{1}{\lambda} \sum_{j=1}^n a_{ji} c(j)$$

por tanto $c = \frac{1}{\lambda} A^t c$, es decir c es **vector propio** de A^t .

Existencia y unicidad

Teorema de Perron-Frobenius

Si $B \in M_{n \times n}$ es una matriz con todas sus entradas **no negativas**, entonces, existe vector $c \geq 0$ t.q. $\|c\|_1 = 1$ y además

$$B c = \rho(B)c.$$

Existencia y unicidad

Teorema de Perron-Frobenius

Si $B \in M_{n \times n}$ es una matriz con todas sus entradas **no negativas**, entonces, existe vector $c \geq 0$ t.q. $\|c\|_1 = 1$ y además

$$B c = \rho(B)c.$$

Es más, si

Existencia y unicidad

Teorema de Perron-Frobenius

Si $B \in M_{n \times n}$ es una matriz con todas sus entradas **no negativas**, entonces, existe vector $c \geq 0$ t.q. $\|c\|_1 = 1$ y además

$$B c = \rho(B)c.$$

Es más, si

- O bien todas sus entradas son **positivas** ($B > 0$),

Existencia y unicidad

Teorema de Perron-Frobenius

Si $B \in M_{n \times n}$ es una matriz con todas sus entradas **no negativas**, entonces, existe vector $c \geq 0$ t.q. $\|c\|_1 = 1$ y además

$$B c = \rho(B)c.$$

Es más, si

- O bien todas sus entradas son **positivas** ($B > 0$),
- O bien B es matriz de adyacencia de un grafo **conexo**,

Existencia y unicidad

Teorema de Perron-Frobenius

Si $B \in M_{n \times n}$ es una matriz con todas sus entradas **no negativas**, entonces, existe vector $c \geq 0$ t.q. $\|c\|_1 = 1$ y además

$$B c = \rho(B)c.$$

Es más, si

- O bien todas sus entradas son **positivas** ($B > 0$),
- O bien B es matriz de adyacencia de un grafo **conexo**,

entonces este c es **único**.

¿Cómo funciona un buscador?



¿Cómo funciona un buscador?



De forma simplificada, un **buscador de páginas web** se compone de:

¿Cómo funciona un buscador?



De forma simplificada, un **buscador de páginas web** se compone de:

- una **araña** de páginas web,

¿Cómo funciona un buscador?



De forma simplificada, un **buscador de páginas web** se compone de:

- una **araña** de páginas web,
- un método para **ordenar** páginas web, de acuerdo con su **relevancia**.

¿Cómo funciona un buscador?



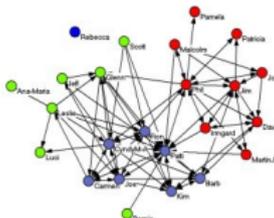
De forma simplificada, un **buscador de páginas web** se compone de:

- una **araña** de páginas web,
- un método para **ordenar** páginas web, de acuerdo con su **relevancia**.

¿Cómo sabe **Google** qué páginas son **más importantes**?

Modelo Matemático

WWW puede modelarse matemáticamente como una **red compleja dirigida** $G = (X, E)$, t.q. cada **nodo** $X = \{1, \dots, n\}$ es una **página web** y hay una **arista** entre dos páginas si hay un **hyper-enlace** entre ellas.



Idea: Usar caminos aleatorios

Si navegamos de forma **aleatoria**, con **mayor frecuencia** pasaremos por los nodos de **mayor importancia**.

Caminos aleatorios y centralidad

- En el instante $t = 0$ elegimos un nodo **al azar**.

Caminos aleatorios y centralidad

- En el instante $t = 0$ elegimos un nodo **al azar**.
- Si en el instante t estamos en el vértice j , en el instante $t + 1$ nos movemos de forma **equiprobable** a uno de los vecinos de j , es decir para cada $1 \leq i \leq n$, la **probabilidad** de que pasemos **de j a i** es

$$p_{ji} = \frac{a_{ji}}{gr_{out}(j)}.$$

Caminos aleatorios y centralidad

- En el instante $t = 0$ elegimos un nodo **al azar**.
- Si en el instante t estamos en el vértice j , en el instante $t + 1$ nos movemos de forma **equiprobable** a uno de los vecinos de j , es decir para cada $1 \leq i \leq n$, la **probabilidad** de que pasemos **de j a i** es

$$p_{ji} = \frac{a_{ji}}{gr_{out}(j)}.$$

- Por tanto para cada $t > 0$ tenemos un vector $p_t = (p_t(1), \dots, p_t(n))$, de forma que cada $p_t(i)$ nos dice la probabilidad de que en el instante t estemos en el nodo i . A la vista de lo anterior

$$p_t(i) = \sum_{j=1}^n p_{t-1}(j) p_{ji} = \sum_{j=1}^n p_{t-1}(j) \frac{a_{ji}}{gr_{out}(j)}.$$

Caminos aleatorios y centralidad

Es decir si definimos la matriz $B = (b_{ij})$ dada por $b_{ij} = p_{ji}$, entonces tenemos que

$$p_t = B p_{t-1}.$$

Caminos aleatorios y centralidad

Es decir si definimos la matriz $B = (b_{ij})$ dada por $b_{ij} = p_{ji}$, entonces tenemos que

$$p_t = B p_{t-1}.$$

Si $p_t = (p_t(1), p_t(2), \dots, p_t(n))$, entonces la **frecuencia de pasar por el nodo i** es

$$\lim_{t \rightarrow \infty} p_t(i).$$

Caminos aleatorios y centralidad

Es decir si definimos la matriz $B = (b_{ij})$ dada por $b_{ij} = p_{ji}$, entonces tenemos que

$$p_t = B p_{t-1}.$$

Si $p_t = (p_t(1), p_t(2), \dots, p_t(n))$, entonces la **frecuencia de pasar por el nodo i** es

$$\lim_{t \rightarrow \infty} p_t(i).$$

Teoría de Perrón-Frobenius

Por ser B una matriz estocástica, $p_t \rightarrow v$, cuanto $t \rightarrow \infty$, donde v es el **valor propio** de B asociado a $\rho(B)$ tal que $v > 0$ y $\|v\|_1 = 1$.

Centralidad PageRank

En este caso, tomando $0 \leq q \leq 1$ (que sera el **factor de salto imprevisto**), consideramos la matriz $\Phi = (\phi_{ij})$ dada por

$$\phi_{ij} = (1 - q) \frac{a_{ji}}{gr_{out}(j)} + \frac{q}{n}.$$

Centralidad PageRank

En este caso, tomando $0 \leq q \leq 1$ (que sera el **factor de salto imprevisto**), consideramos la matriz $\Phi = (\phi_{ij})$ dada por

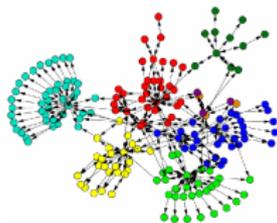
$$\phi_{ij} = (1 - q) \frac{a_{ji}}{gr_{out}(j)} + \frac{q}{n}.$$

Centralidad Page-Rank (tipo Google)

Usando de nuevo la teoría de Perrón-Frobenius, si tenemos la sucesión de vectores $(p_t)_{t \geq 0}$, entonces p_t tiende al valor propio de Φ asociado a $\rho(\Phi)$ tal que $w > 0$ y $\|w\|_1 = 1$. Este w es la **centralidad Page-Rank** de G con factor de salto q .

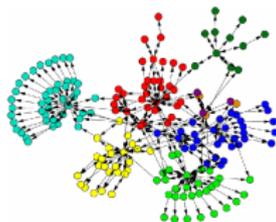
Algunos problemas relacionados

Si tenemos una red de páginas web



Algunos problemas relacionados

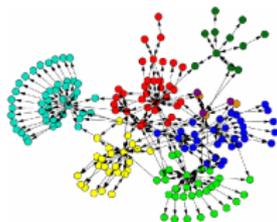
Si tenemos una red de páginas web



- ¿Qué hiper-enlaces debemos **añadir** para **mejorar lo máximo posible** el PageRank de un nodo fijo j ?

Algunos problemas relacionados

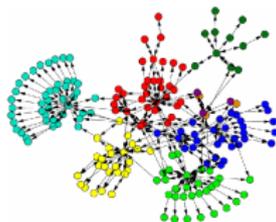
Si tenemos una red de páginas web



- ¿Qué hiper-enlaces debemos **añadir** para **mejorar lo máximo posible** el PageRank de un nodo fijo j ?
- ¿Cómo afectan los **nuevos hiper-enlaces** en el PageRank de las páginas?

Algunos problemas relacionados

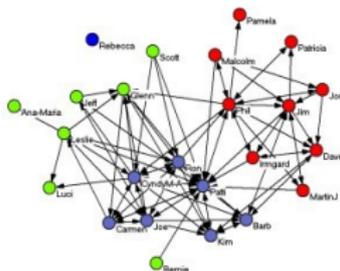
Si tenemos una red de páginas web



- ¿Qué hiper-enlaces debemos **añadir** para **mejorar lo máximo posible** el PageRank de un nodo fijo j ?
- ¿Cómo afectan los **nuevos hiper-enlaces** en el PageRank de las páginas?
- ¿Pueden algunos nodos de la red **controlar** la centralidad de toda la red?

Controlabilidad de Redes

Si tenemos una red $G = (X, E)$ con **pesos** positivos (w_{ij}) en las aristas,



podemos definir su **centralidad Bonacich** y **Pagerank** tomando como matriz de adyacencia

$$a_{ij} = \begin{cases} w_{ij}, & \text{si existe una arista } i \rightarrow j, \\ 0, & \text{en otro caso.} \end{cases}$$

Controlabilidad de Redes

Por fijar ideas, trabajemos con la **centralidad Bonacich**.

Definición

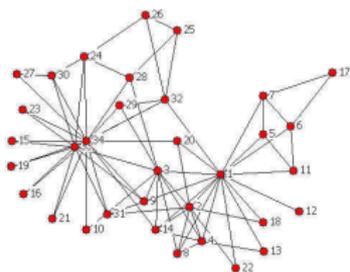
Si tenemos una red $G = (X, E)$ con **pesos** positivos en las aristas, un subconjunto de nodos $Y \subseteq X$ es un **conjunto controlador** de G si modificando el peso de las aristas que salen de Y , podemos obtener la centralidad que queramos **para todos los nodos** de G .

Controlabilidad de Redes

Por fijar ideas, trabajemos con la **centralidad Bonacich**.

Definición

Si tenemos una red $G = (X, E)$ con **pesos** positivos en las aristas, un subconjunto de nodos $Y \subseteq X$ es un **conjunto controlador** de G si modificando el peso de las aristas que salen de Y , podemos obtener la centralidad que queramos **para todos los nodos** de G .



Controlabilidad de Redes

Teorema de Existencia

Si tenemos una red $G = (X, E)$ con **pesos** positivos en las aristas, tal que $gr_{in} > 0$ para cualquier $i \in X$, entonces **siempre existe** al menos un **conjunto controlador** de nodos.

Controlabilidad de Redes

Teorema de Existencia

Si tenemos una red $G = (X, E)$ con **pesos** positivos en las aristas, tal que $gr_{in} > 0$ para cualquier $i \in X$, entonces **siempre existe** al menos un **conjunto controlador** de nodos.

Pregunta

¿Qué **tamaño relativo** tienen los conjuntos controladores?

Controlabilidad de Redes

Teorema de Existencia

Si tenemos una red $G = (X, E)$ con pesos positivos en las aristas, tal que $gr_{in} > 0$ para cualquier $i \in X$, entonces **siempre existe** al menos un **conjunto controlador** de nodos.

Pregunta

¿Qué **tamaño relativo** tienen los conjuntos controladores?

Red Real	Nodos	Aristas	Tamaño relativo
astro	13259	123838	16 %
Internet AS	11174	23409	8 %
US Airports	500	2980	14 %
ENRON	2351	139579	7 %
Jazz	198	2742	8 %

-  R. ALBERT AND A.L. BARABÁSI, “*Statistical Mechanics of Complex Systems*” Rev. Mod. Phys. **E74**, (2002) 47–97.
-  P. BONACICH, “*Factoring and weighing approaches to status scores and clique identification*”, J. Math. Soc. **2** (2002), 113–141.
-  S. BRIN AND L. PAGE, “*The anatomy of a large-scale hypertextual Web search engine*”, Comput. Netw. **30** (1998), 107–117.
-  R. CRIADO, V. LATORA, V. NICOSIA, M. ROMANCE AND G. RUSSO, “*Analytical solutions for the inverse centrality problems*”, to appear in Chaos (201?).
-  V. NICOSIA, M. ROMANCE, V. LATORA, R. CRIADO AND G. RUSSO, “*Controlling centrality in weighted complex networks*”, to appear in Phys. Rev. Let. (201?).
-  M. ROMANCE, “*Local estimates for eigenvector-like centralities of complex networks*”, to appear in J. Comput. Appl. Math. (201?).

Análisis de Redes Complejas: Un paseo matemático entre Google y las redes sociales

Miguel Romance

miguel.romance@urjc.es

Dpto. de Matemática Aplicada
ES CET
Universidad Rey Juan Carlos
Madrid

Murcia, 25 de noviembre de 2010

Contenido

1 Introducción

- Motivación y objetivos
- Un paseo por el análisis de redes complejas

2 Centralidad y poder

- Redes Sociales: Centralidad Bonacich
- Buscadores y Centralidad PageRank
- Algunos problemas sobre centralidad