

# Análisis de Redes

George G Vega<sup>1</sup>

Superintendencia de Pensiones

20 de junio, 2014

---

<sup>1</sup><mailto:gvegayon@caltech.edu>

# Contenidos

- 1 Introducción
- 2 Definiciones
- 3 Estadísticos
- 4 Ejemplos
- 5 Referencias

# Introducción

## Objetivos de esta sección

Esta parte del curso se centrará en lo siguiente:

- Familiarizarse con el lenguaje
- Entender gráficos desde mirada estadística
- Realizar análisis de redes utilizando software especializado (Gephi)

# Alternativas de software

- **UCINET**

<https://sites.google.com/site/ucinetsoftware/home>

- **Pajek** <http://pajek.imfm.si/doku.php>

- **NodeXL** <http://nodexl.codeplex.com/>

- **NetworkX** (Python) <http://networkx.lanl.gov/>

- **igraph** (R y otros) <http://igraph.org/redirect.html>

- ...

- **Gephi** <http://gephi.org/>

# Introducción

## La ciencia de redes

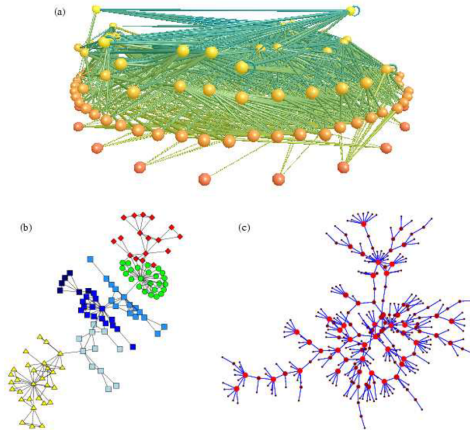
Cómo se estudian las redes?

- **Redes biológicas** Redes metabólicas, redes regulatorias, etc.
- **Redes tecnológicas** Redes de telecomunicaciones, energía, transporte, etc.
- **Redes sociales** Amigos, familia, parejas sexuales, etc.
- **de Información** Citas académicas, World Wide Web (Red), etc.

En las ciencias sociales es donde se lleva más tiempo desarrollando investigación cuantitativa al respecto.

# Introducción

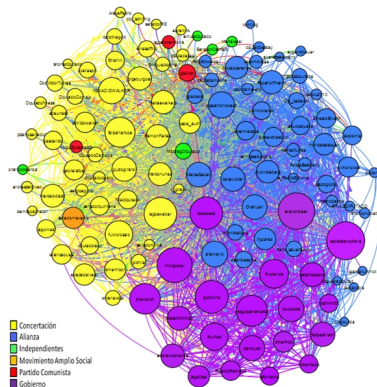
## La ciencia de redes



**a)** Cazador-presa, **b)** Red de colaboración, y **c)** Red de contactos sexuales Fuente: [3]

# Definiciones

Gráfico 2: Red de parlamentarios, ministros y el presidente de Chile en Twitter (al 28/04/2011)



Fuente: [1]

# Definiciones

## Conceptos fundamentales

- **Vertice** (aka nodo) Unidad fundamental de un grafo.
- **Arista** (aka arco) Puede ser dirigida o no.
- **Grado** (entrada/salida) Número de conexiones de un nodo.
- **Geodésica** Distancia más corta entre dos nodos.
- **Diametro** Distancia más larga en un grafo.
- **Diada** Par de nodos conectados.
- **Triada** Trio de nodos conectados.
- **Componente** (gigante) Porción de un grafo desconectado.



# Definiciones

- Definición de gráfo:  $G = (V, E)$ , donde  $V$  (vertex) representa el conjunto de nodos y  $E$  (edges) representa el número de aristas.
- Tipos de grafo:
  - **Grafo vs Bigraf** En un bigrafo existen dos tipos de nodos distintos (personas e instituciones por ejemplo)
  - **Dirigido** Los arcos tienen sentido,  $v_i \rightarrow v_j$
  - **No dirigido** Los arcos no tienen sentido (dirección),  $v_i - v_j$
  - **Cíclico** Existen bucles,  $v_i \rightarrow v_j \rightarrow v_k \rightarrow v_i$
  - **a-cíclico** No existen bucles

# Definiciones

## Caracterizando una red

Máximo número de conexiones en un grafo **dirigido**

$$n(n - 1) \quad (1)$$

Máximo número de conexiones en un grafo **no dirigido**

$$\frac{1}{2}n(n - 1) \quad (2)$$

# Definiciones

## Mundos Pequeños

**efecto mundo-pequeño** el hecho de que la mayoría de los pares de nodos está conectado por un camino pequeño a lo largo del grafo. La primera evidencia de esto se encuentra en el experimento de Stanley Milgram [2]

Se puede medir en función de la geodésica media

$$\ell = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij} \quad (3)$$

Notar que  $\frac{1}{2}n(n+1)$  corresponde al máximo número de links posibles. Como  $d_{ij} \rightarrow \infty$ , se recomienda calcular

$$\ell^{-1} = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij}^{-1} \quad (4)$$

# Definiciones

## Distribución de grado

Una característica importante de los grafos *naturales* es la distribución del grado ( $k_i$ ) de sus nodos ( $v_i$ )

$$P(\text{grado} \geq k) : P_k = \sum_{k'=k}^{\infty} p'_k$$

donde  $p_k$  es la fracción de nodos que tiene grado  $k$

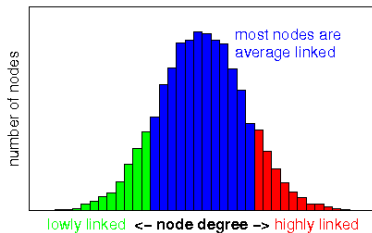
Ahora, dado que se da  $p_k \sim k^{-\alpha}$

$$P_k \sim \sum_{k'=k}^{\infty} k'^{-\alpha} \sim k^{-(\alpha-1)}$$

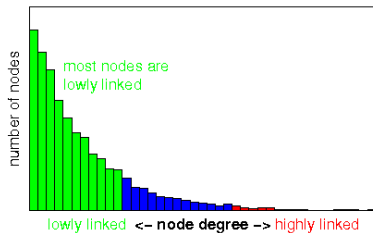
# Definiciones

## Distribución de grado

random networks



real networks (power-law, scale-free)



www.network-science.org

# Definiciones

## Clusterización

Indica en qué medida existen triadas conectadas en un grafo.

- **Watts y Strogatz, 1998**

$$C_i = \frac{2E_i}{k_i(k_i - 1)}; C = \frac{1}{n} \sum_{i \in V} C_i \quad (5)$$

donde  $E_i$  corresponde al número de arcos,  $k_i$  el grado del nodo  $i$ .

- **Barrat y Weight, 2000; Newman, Strogatz y Watts, 2000**

$$C = \frac{3 \times \# \text{ de triángulos en el grafo}}{\# \text{ de triadas}} \quad (6)$$

El número 3 se utiliza para asegurar que el índice se encuentre entre 0 y 1

# Definiciones

## Clusterización

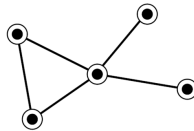


FIG. 5 Illustration of the definition of the clustering coefficient  $C$ , Eq. (3). This network has one triangle and eight connected triples, and therefore has a clustering coefficient of  $3 \times 1/8 = \frac{3}{8}$ . The individual vertices have local clustering coefficients, Eq. (5), of 1, 1,  $\frac{1}{6}$ , 0 and 0, for a mean value, Eq. (6), of  $C = \frac{13}{30}$ .

**(3.3) Nuestra ecuación (6); (3.5) y (3.6) Nuestra ecuación (5)**

# Ejemplos

## Medidas en la literatura

10

	network	type	$n$	$m$	$z$	$\ell$	$\alpha$	$C^{(1)}$	$C^{(2)}$	$r$	Ref(s).
social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208	20, 416
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127	311, 313
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1				8, 9
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16		136
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092	321
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	–0.029	45
	sexual contacts	undirected	2 810				3.2				265, 266
information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	–0.067	14, 34
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7				74
	citation network	directed	783 339	6 716 198	8.57		3.0/–				351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157	244
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44		119, 157
technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	–0.189	86, 148
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	–0.003	416
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	–0.033	366
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	–0.016	318
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	–0.119	395
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	–0.154	155
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	–0.366	6, 354
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	–0.240	214
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	–0.156	212
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	–0.263	204
	freshwater food web	directed	92	997	10.84	1.90	–	0.20	0.087	–0.326	272
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	–0.226	416, 421

TABLE II Basic statistics for a number of published networks. The properties measured are: type of graph, directed or undirected; total number of vertices  $n$ ; total number of edges  $m$ ; mean degree  $z$ ; mean vertex-vertex distance  $\ell$ ; exponent  $\alpha$  of degree distribution if the distribution follows a power law (or “–” if not; in/out-degree exponents are given for directed graphs); clustering coefficient  $C^{(1)}$  from Eq. (3); clustering coefficient  $C^{(2)}$  from Eq. (6); and degree correlation coefficient  $r$ , Sec. III.F. The last column gives the citation(s) for the network in the bibliography. Blank entries indicate unavailable data.



# Temas que revisaremos en la próxima sesión

- Centralidad de Grado (in/out)
- Centralidad de Cercanía
- Centralidad de Intermediación

Otras medidas de centralidad...

- Número de Erdős
- El oráculo de Kevin Bacon <http://oracleofbacon.org/>

# Referencias



Jorge Fábrega and George Vega.

The network of scientific production in chile since 1909.  
2011.



Stanley Milgram.

The small world problem.  
*Psychology today*, 2(1):60–67, 1967.



Mark EJ Newman.

The structure and function of complex networks.  
*SIAM review*, 45(2):167–256, 2003.