# Representación, Medición y Conceptos Básicos de Redes[1]

Alvaro J. Riascos Villegas

Agosto, 2023

[1] Basado en Balakrishnan, R y Ranganathan, K (2000). A Textbook of Graph Theory y Jackson, M (2008). Social and Economic Networks. Princeton University Press. Todas las figuras Jackson (2008).

# Contenido

1. Grafos
    - Definiciones básicas*
    - Grado
    - Homomorfismos e isomorfismos
    - Cohesión y aglomeración
    - Centralidad*
    - Otros conceptos
2. Resultados Básicos
    - Teorema de Hall y grafos bipartitos
    - Coberturas y conjuntos independientes
    - Colorear
    - Tours Eulerianos y Ciclos Hamiltonianos

# Grafos simples

- Un grafo es una pareja $G = (N, g)$ donde $N = \{1, ...n\}$ es un conjunto de nodos y $g$ es una matriz $n \times n$. $n$ se llama el orden del grafo.
- $g_{ij} = 1$ representa que existe un enlace entre $i$ y $j$.
- Un grafo es no dirigido si $g_{ij} = g_{ji}$. Dirigido en caso contrario.
- Los grafos sin loops se representan con una matriz con ceros en la diagonal, $g_{ii} = 0$ (excepto cuando se diga lo contrario).
- Un grafo es simple si es no dirigido y no tiene loops.

# Grafos simples*

- Excepto cuando se diga lo contrario, los grafos van a ser no dirigidos, sin loops.
- Los grafos también sulen representarse como una lista: $G = \{\{1, 2\}, \{1, 4\}\}$ lo que significa que $g_{12} = g_{21} = g_{14} = g_{41} = 1$ y todos los demás cero. Abusando un poco de la notación tambien se escribe $G = \{12, 14\}$.
- Si todos los vertices son adjacentes con todos los demas vertices se llama un grafo completo. Si $n$ es el orden, $K_n$ denota el grafo completo de orden $n$.

# Grafos dirigidos, con pesos, multigrafos*

- En un grafo dirigido, $g_{ij} = 1$ significa que existe un enlace de $i$ hacia $j$. $g$ no tiene que ser simétrica.
- $g_{ij} > 1$ se usa para representar el peso del enlace.
- Los elementos distintos de uno también se pueden utilizar para representar multigráfos (i.e., grafos en los que dos vertices tiene varios enlaces que los conectan).

# Subgrafos

- Dado $G$, algunas relaciones básicas entre grafos son la de **subgrafo** $G'$ (o supergrafo): si los vértices y enlaces de $G'$ son un subconjunto de $G$; **subgrafo que expande el grafo** (spanning): si $G'$ es un subgrafo y además tiene los mismos vértices de $G$; y $G'$ es un **subgrafo inducido** si todo enlace en $G$ con vértices en $G'$ es también un enlace de $G'$.

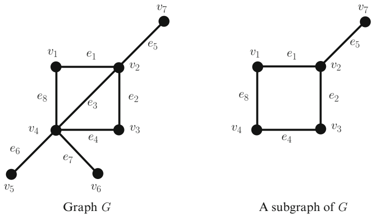
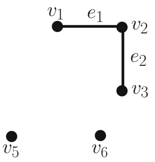
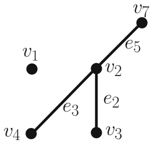



**Fig. 1.13** Various subgraphs and cliques of $G$

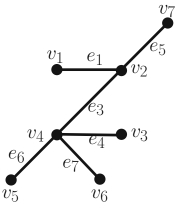
# Ejemplos



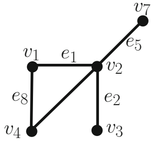

An induced subgraph of $G$




A subgraph of $G$,
but not an induced subgraph of $G$




A spanning subgraph of $G$




An edge-induced subgraph of $G$
induced by $E' = \{e1, e2, e3, e5, e8\}$

# Caminatas, caminos, etc.*

- Una caminata (*walk*) es un recorrido sin saltos por las aristas de un grafo. Las caminatas tienen un comienzo y un final. Pueden repetirse vertices y aristas.
- Un camino (*path*) es una caminata sin repetir vertices. Los caminos de un grafo $G$ son de la forma $P = (V, E)$ donde $V = \{x_0, ..., x_k\}$ y $E = \{x_0x_1, ..., x_{k-1}x_k\}$, donde todos los $x_i$ son distintos.
- Un ciclo $C$ de un grafo es un camino $P$ más una arista $x_kx_0$ y se denota por $P + x_kx_0$.
- Una geodésica entre los nodos ij es el un camino más corto entre los nodos.
- Obsérvese que, definimos $g_{ii} = 0$ entonces, las componentes de $g^k$ cuentan cuantas caminatas existen entre dos nodos incluyendo posibles cíclos en el camino (i.e. no son necesariamente caminos).

# Distancia, girth, etc.

- La distancia entre dos nodos es el número (o suma de los pesos) de las aristas de una geodésica entre los nodos (cuando no hay un camino entre los nodos es infinito).
- Una medida relevante es el promedio de la distancia del camino más corto (geodésicas) entre todo par de nodos.
- El diámetro es la mayor distancia entre cualquier par de nodos.
- *Girth* es la longuitud del ciclo más corto en la red (infinito si no hay ciclos).
- La circunferencia es la longuitud del ciclo más largo en la red (cero si no existen ciclos).

# Conexión*

- Una red es **fuertemente conexa** si existe un camino (dirigido) entre todo par de vértices.
- La matriz de producción de 15 sectores en Estados Unidos es fuertemente conexa. En contraste, la desagregación a 71 sectores no es fuertemente conexa. Por ejemplo "food and beverage stores" son un nodo absorvente, no proveen insumos a ningun sector aunque si provean de bienes a los consumidores finales.
- Notación: $g^k_{i,j} \equiv (g^k)_{i,j}$

**Theorem**

*Sea $(N, g)$ un grafo dirigido. Existe un camino dirigido entre $i, j$ de longuitud $k$ si y solo si $g^k_{i,j} > 0$*

- La prueba de este teorema es por inducción. Cuando $k = 1$ el teorema es cierto por definición. Supongamos que existe un camino dirigido entre $i, j$ de longuitud $k + 1$: $\{i_1, ..., i_{k+1}\}$. En particular, existe un camino dirigido entre $i_1, i_k$. Por la hipótesis de inducción, $g^k_{i_1,i_k} > 0$. Ahora:

$$g^k_{i_1,i_{k+1}} = \sum_{j=1}^{n} g^k_{i_1,j} g_{j,i_{k+1}} \tag{1}$$

y como, $g_{i_k,i_{k+1}} > 0$, luego, $g^{k+1}_{i_1,i_{k+1}} > 0$.

# Aperiodicidad*

- Un grafo es periodico si existe un $k > 1$ que divide la longuitud de todos los ciclos del grafo.
- Un grafo es aperiodico si no periodico.
- Ejemplo: Si un grafo tiene loops, entonces es aperiodico.

# Árboles

- Un grafo que no contiene cíclos (i.e., acíclico) es un bosque. Un bosque conexo es un árbol.
- Los vertices de grado uno se llaman hojas.
- Usualmente se marca un vértice específico que se denomina la raíz del árbol y que induce un orden parcial.

# Grado y distribución del grado*

- El grado de un vertice es el número de vertices que son adyacentes.
- Tres distribuciones importantes del grado de un vertice son: regular (todos los vertices tiene el mismo grado), Poisson y la distribución invariante de escala (power law):

$$P(d) = cd^{-\gamma}$$

  donde $c$ es una constante de normalización.
- La razón de las probabilides de diferentes grados no depende de la escala (i.e., solo depende de la razón de los grados).
- Tiene colas (en ambos lados) más gruesas que Poisson.

# Distribución del grado*

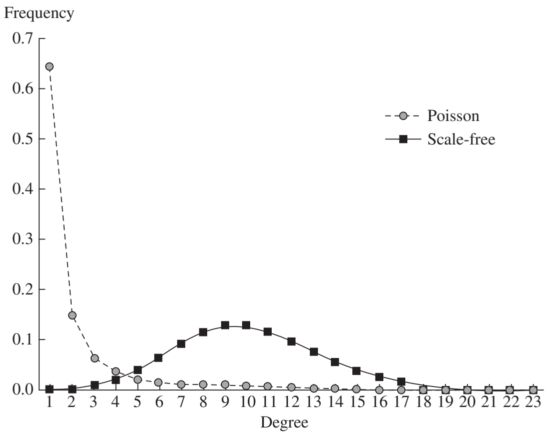



**FIGURE 2.8** Comparing a scale-free distribution to a Poisson distribution.

# Distribución del grado*

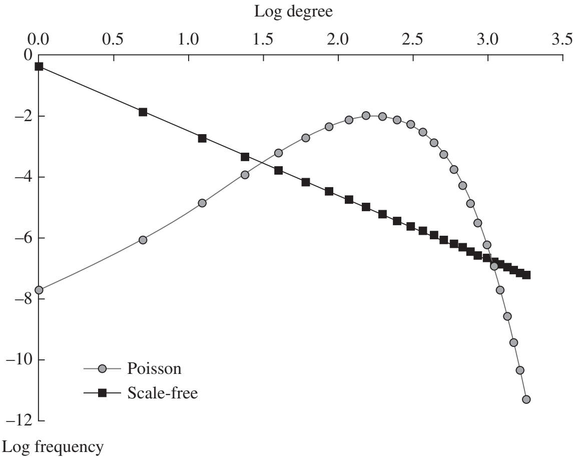



**FIGURE 2.9** Comparing a scale-free distribution to a Poisson distribution: log-log plot.

# Homomorfismos e isomorfismos

- Un homomorfismo entre grafos es una función uno a uno entre los vertices que respeta la relación de enlaces entre los vértices.
- Es un isomorfismo si además la función es sobreyectiva, invertible y la inversa preserva la relación de enlaces entre vertices.

- A proper coloring of $G$ is an assignment of colors to each vertex such that no two vertices in the same edge have the same color. The least number of colors for which there is a coloring of graph $G$ is called the chromatic number of $G$.
- Aplications: Exams scheduling. Each class is a vertex. There is an edge among to classes if there is at least one student taking both classes. Colors define time slots for each exam. A coloring of this graph is an exam scheduling such that no students is scheduled to take two exams at the same time.
- The complete graph of $n$ vertices is denoted by $K_n$

**Theorem**

*$G$ is r-colorable if and only if there is an homomorphism from $G$ to $K_r$.*

# Complementos

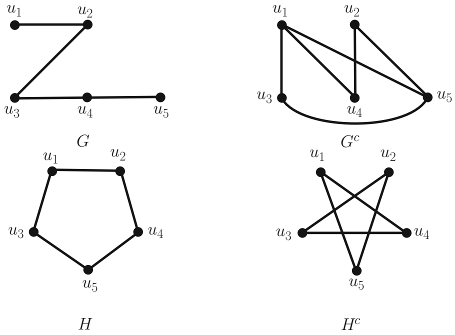
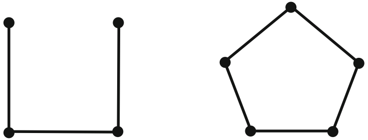



**Fig. 1.11** Two simple graphs and their complements

# Complementos isomorfos

**Fig. 1.12** Self-complementary graphs



# Cohesión

- Medidas de cohesión: Un *clique* Es una sub red completa maximal.

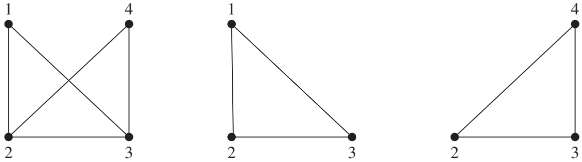



FIGURE 2.11 A network on four nodes and its two cliques.

  - Contar el número y tamaño de cliques es valioso pero muy sensible a cambios pequeños en la red. Por ejemplo eliminar un nodo de una red completa da origen a varios cliques de menor tamaño.

# Aglomeración

- La medida más común es buscar todos los nodos con por lo menos dos vecinos (i.e., $j, k$) y calcular qué tan frecuente existe un enlace entre los vecinos (i.e., un enlace entre $j, k$). La medida de aglomeración (global) $Cl(g)$ se define como:

$$Cl(g) = \frac{\sum_i \{jk \in g : k \neq j; k, j \in N_i(g)\}}{\sum_i \{jk : k \neq j; k, j \in N_i(g)\}} \qquad (2)$$

- Una medida análoga se puede definir (localmente) para cada nodo (no sumar sobre todos los nodos en el numerador y denominador).
- La aglomeración promedio se obtiene de promediar la aglomeración por nodo y es distinto al concepto de aglomeración global.

# Centralidad

- Las medidas de centralidad se pueden clasificar en:
  1. Centralidad de grado $= \frac{d_i(g)}{n-1}$.
  2. Distancia $\sum_{j \neq i} \delta^{l(i,j)}$, donde $\delta \in (0, 1)$.
  3. Posicionamiento (*betweenness*).
  4. Características de los vecinos (prestigio, poder).
- El grado no es una buena medida de centralidad.

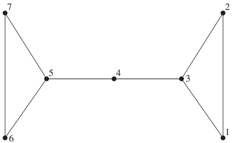



FIGURE 2.13 A central node with low degree centrality.

# Centralidad: Posicionamiento*

- Medida de posicionamiento (*betweenness* Freeman) de cada nodo $k$:

$$\sum_{i,j:i\neq j} \frac{\frac{P_k(ij)}{P(ij)}}{(n-1)(n-2)/2} \tag{3}$$

donde $\frac{P_k(ij)}{P(ij)} = 0$ si no hay caminos entre $i$ y $j$. El coeficiente $(n-1)(n-2)/2$ es el número máximo de pares de nodos que incluirían al nodo $k$.

# Centralidad: Prestigio Katz*

- Características de los vecinos: La idea fundamental es que la importancia de un nodo depende de la importancia de sus vecinos.
- El prestigio de Katz se define como:

$$P_i^K(g) = \sum_{j \neq i} g_{ij} \frac{P_j^K(g)}{d_j(g)} \tag{4}$$

- Sea $\hat{g}_{ij} = \frac{g_{ij}}{d_j(g)}$. Entonces el prestigio de Katz se puede escribir como:

$$P^K(g) = \hat{g} P^K(g) \tag{5}$$

- Lo que se reduce a calcular el vector propio asociado al valor propio 1.

# Centralidad: PageRank (Google)

- Es un modificación del prestigio de Katz que garantiza que exista el valor propio.

$$P_i^K(g) = (1-d)I + d\sum_{j\neq i} g_{ij}\frac{P_j^K(g)}{d_j(g)} \tag{6}$$

Esto garantiza que toda página tenga un prestigio mínimo y que se pueda utilizar el teorema fundamental de cadenas de Markov para calcularlo.

# Centralidad: Vectores propios (Bonacich)

- Bonacich propuso la siguiente variante (centralidad de vectores propios). La centralidad de un nodo $C_i^e(g)$ es proporcional a la suma de la centralidad de los nodos vecinos:

$$\lambda C_i^e(g) = \sum_j g_{ij} C_j^e(g) \tag{7}$$

  donde $\lambda$ es la constante de proporcionalidad.
- Este problema se reduce a encontrar los valores propios y vectores propios de $g$ (i.e., a diferencia de la medida de Katz, no se normaliza la matriz de incidencia).
- La convención es usar el vector propio correspondiete al mayor valor propio $\lambda$ (que en general es positivo).

# Grafos: Centralidad*

- En la siguiente tabla se muestra el resultado de aplicar las diferentes medidas de centralidad al grafo anterior.

**TABLE 2.1**
Centrality comparisons for Figure 2.13

| Measure of centrality | Nodes 1, 2, 6, and 7 | Nodes 3 and 5 | Node 4 |
|---|---|---|---|
| Degree (and Katz prestige $P^K$) | .33 | .50 | .33 |
| Closeness | .40 | .55 | .60 |
| Decay centrality ($\delta = .5$) | 1.5 | 2.0 | 2.0 |
| Decay centrality ($\delta = .75$) | 3.1 | 3.7 | 3.8 |
| Decay centrality ($\delta = .25$) | .59 | .84 | .75 |
| Betweenness | .0 | .53 | .60 |
| Eigenvector centrality | .47 | .63 | .54 |
| Katz prestige-2 $P^{K2}$, $a = 1/3$ | 3.1 | 4.3 | 3.5 |
| Bonacich centrality $b = 1/3$, $a = 1$ | 9.4 | 13.0 | 11.0 |
| Bonacich centrality $b = 1/4$, $a = 1$ | 4.9 | 6.8 | 5.4 |

# Grafos: Hubs y Autoridades*

- En grafos dirigidos se pueden diferenciar dos tipos de centralidad de los vertices.
- Los hubs son vertices que tienen un **grado de salida** alto.
- Las autoridades son vertices que tienen un **grado de entrada** alto.

# Grafos: Hubs y Autoridades*

- En una red dirigida con pesos, la centralidad de vector propio definida arriba refleja un concepto de centralidad de hubs.
- Si se define la centralidad de vector propio como:

$$\lambda C_j^e(g) = \sum_i g_{ij} C_i^e(g) \tag{8}$$

se obtiene una medida de centralidad de autoridad.

# Hubs y Autoridades*

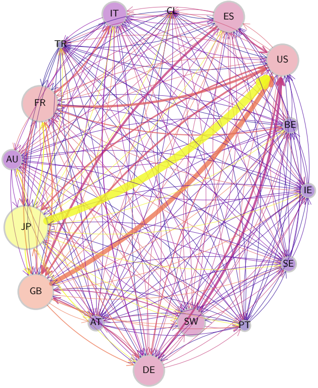



Figura: Red de crédito entre países.

# Hubs y Autoridades*

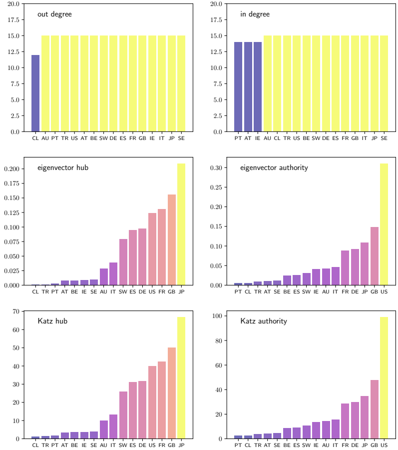



Figura: Red de crédito entre países.

# Reducción de un grafo dirigido a no dirigido

- La forma más trivial es ignorar la dirección.
- Grafo de cocitaciones: se construye la matriz de cocitaciones $C$. Donde $C_{ij}$ es el número de vertices que apuntan a $i$ y $j$ simultáneamente. $C$ define un grafo no dirigido con pesos.
- Grafo de referencias comúnes (bibliographical coupling): se construye la matriz $B$, $B_{ij}$ es el número de vertices a los que ambos vertices $i$ y $j$ apuntan de forma simultánea.
- Estos grafos se utilizan para hacer busquedas como por ejemplo en el algoritmo HITS.

# Hipergrafos y grafos bipartitos

- Los hipergrafos son grafos en los que los vertices pertenecen a una o varias clases. En este caso los enlaces conectan varios vertices. La forma de representarlos es mediante un grafo bipartito.
- Los vertices del grafo original se denotan como vertices de tipo uno, y se introducen nuevos vertices, todos de tipo cero, uno por cada clase. Los enlaces denotan la relación de pertenencia de las diferentes clases.
- Los grafos bipartitos se pueden representar como proyecciones en un grafo con vertices de tipo uno o alternativamente como un grafo con vertices de tipo cero.

# Contenido

1. Grafos
   - Definiciones básicas*
   - Grado
   - Homomorfismos e isomorfismos
   - Cohesión y aglomeración
   - Centralidad*
   - Otros conceptos
2. Resultados Básicos
   - Teorema de Hall y grafos bipartitos
   - Coberturas y conjuntos independientes
   - Colorear
   - Tours Eulerianos y Ciclos Hamiltonianos

# Resultados Básicos

- En esta parte discutimos:
  1. Teorema de Hall y grafos bipartitos
  2. Coberturas con conjuntos y conjuntos independientes
  3. Colorear
  4. Toros Eulerianos y Ciclos Hamiltonianos

# Teorema de Hall y grafos bipartitos

- Un grafo bipartito es uno en el que los nodos se pueden particionar en dos conjuntos tal que enlace del grafo tiene un extremo en un conjunto y el otro extremo en otro conjunto.
- Un problema típico es, dados $A, B$ una partición de un grafo bipartito y $S \subset A$ decimos que $\mu : S \to B$ inyectiva es un emparejamiento relativo a al grafo $g$ si para todo $i \in S, i\mu(i) \in g$.

**Theorem (Hall)**

*Sea $(N, g)$ un grafo bipartito con partición $\{A, B\}$. Existe un emparejamiento de $S \subset$ relativo a $g$ si y solo si para todo $S' \subset A$, $Card(N_{S'}(g)) \geq Card(S')$*

# Coberturas y conjuntos independientes

- Un conjunto de nodos es independiente si entre ellos no existe ningún enlace.
- Si $g'$ es un subgrafo de $g$ entonces todo conjunto independiente de $g'$ lo es de $g$. Sin embargo, para subgrafos propios, siempre existe un conjunto independiente (maximal) de $g$ que no es independiente (maximal) de $g'$.

# Coberturas y conjuntos independientes: Ejemplo

- Cosidere el siguiente juego en una red: En una red cada nodo representa un jugador. Cada jugador puede comprar un libro o pedirlo prestado sin costo solo de vecinos directos. En caso de que ningin vecino le preste el libro es mejor comprarlo.
- Un equilibrio es una situación en la que nadie que haya comprado el ibro se arrepiente y nadie que no haya comprado el libro le gustaria comprarlo.
- Un equilibrio de este juego es uno en el que los compradores del libro son un conjunto independiente maximal.
- Para ver esto, obsérvese que la primera condición implica que el conjunto es independiente y la segunda que es maximal.

# Colorear

- Queremos colorear los nodos de tal forma que ningún enlace relacione a dos nodos del mismo color. El mínimo número de colores necesarios se llama el número cromático.
- Aplicación: Un grafo donde los nodos son conferencistas y los enlaces reflejan el deseo de los investigaodres estar en la conferencia del otro. Los colores representarian diferentes espacios horarios para las presentaciones. El número cromático sería el menor número de espacios para que no haya interferencias.
- Un problema famosos es el el problema de los cuatro colores que es equivalente a que todo grafo planar tiene número cromático menor o igual a 4.

# Colorear

- Un grafo que necesita cuatro colores.

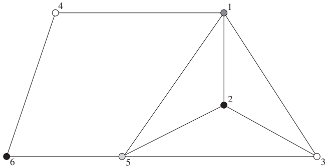



FIGURE 2.16 A planar network on six nodes with chromatic number 4.

# Tours Eulerianos y Ciclos Hamiltonianos

- Un caminata cerrada que incluye cada enlace del grafo solo una vez se llama Tour Euleriano o circuito.

**Theorem**

*Una red conexa tiene una caminata cerrada si y sólo si el grado de cada nodo es par.*

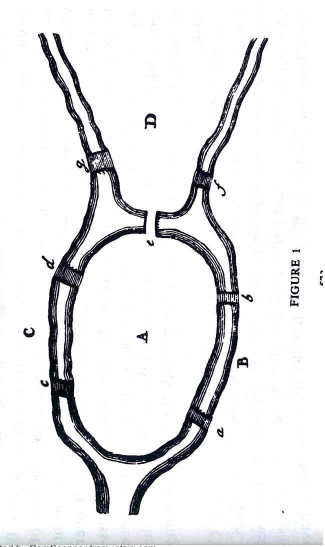
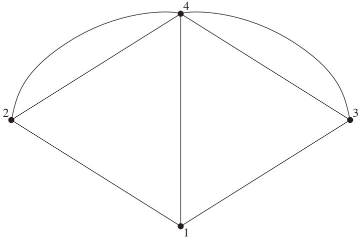



FIGURE 1

# Tours Eulerianos y Ciclos Hamiltonianos



- Una pregunta análoga es si es posible encontrar una caminata cerrada en el que se visite cada nodo exactamente una vez (debe ser un ciclo y se llama ciclo Hamiltoniano).
- Existe un camino que involucre a cada nodo exactamente una vez (camino Hamiltoniano)?

**Theorem (Dirac)**

*Si la red tiene $n \geq 3$ y cada nodo tiene grado por lo menos $\frac{n}{2}$ entonces la red tiene un ciclo Hamiltoniano.*