Introducción a la Teoría de Redes[1]

Alvaro J. Riascos Villegas

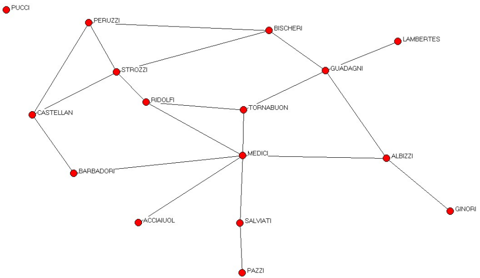
Enero, 2014

[1] Basado en Jackson, M (2008). Social and Economic Networks. Princeton University Press. Todas las figuras Jackson (2008).

Contenido

Ejemplos de Redes: Matrimonios Familia Medici, Florencia 1400

- Padgett, J.F y C.K. Ansell (1993). Robust action and the rise of the Medici, 1400-1434.
- Red de matrimonios entre familias (cada enlace representa un matrimonio entre miembros de dos familias).



Ejemplos de Redes: Matrimonios Familia Medici, Florencia 1400

- Basados en la riqueza y poder político es difícil explicar como los Medici surgieron como una familia tan importante (la familia Strozzi tenía más riqueza y poder político, sin embargo fueron opacados por los Medici).
- La estructura de relaciones puede ser un determinante.
- Si comparamos con cuántas familias se encuentra una familia específica relacionada y comparamos entre ellas, los Medici sobresalen (3 a 2).
- Una relacion de cercania resulta más sugestiva.

Ejemplos de Redes: Matrimonios Familia Medici, Florencia 1400

- Sea $P(ij)$ el número de caminos más cortos que conectan una familia i con j. Sea $P_k(ij)$ el número de estos caminos que incluyen a la familia k.
- Por ejemplo si $i =$Barbadori, $j =$Guadagni, entonces $P(ij) = 2$. Si $k =$Medici entonces $P_k(ij) = 2$ mientras que si $k =$ Strozzi o Albizzi $P_k(ij) = 2$ es cero o uno respectivamente.
- Si calculamos una medida de importancia (betweenness Freeman) de cada familia k como:

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

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 familias que incluirían a la familia k.

Ejemplos de Redes: Familia Medici, Florencia 1400

- Esta medida de poder para los Medici es 0.522. Esto significa que los Medici están en más de la mitad de los caminos más cortos entre todos los caminos más cortos entre cada par de familias.
- Este mismo cálculo para los Strozzi es 0.103. El segundo más alto es los Guadagni que es 0.255.
- En este sentido los Medici estaban mejor posicionados que cualquier otra familia.
- Esta estructura es endógena? Es ópima?

Ejemplos de Redes: Amistades y romances en estudiantes secundaria

- Datos de $90,000$ estuadiantes de la encuesta Add Health entrevistados en los años 90.
- A los estudiantes se les preguntaba con quién habian tenido relaciones romanticas en los últimos seis meses.

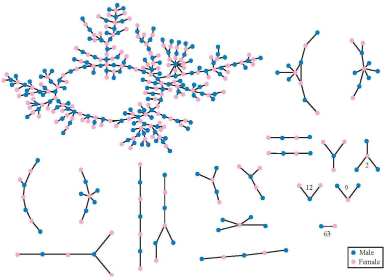


Figure 1.2: A Figure from Bearman, Moody and Stovel [47] based the Add Health Data Set. A Link Denotes a Romantic Relationship, and the Numbers by Some Components Indicate How Many Such Components Appear.

Ejemplos de Redes: Amistades y romances en estudiantes secundaria

- Grafo bipartito.
- Una componente gigante (relevante para la difusión de enfermedades, etc.).
- Esta es una estructura similar a la que surge de una red que se forma mediante enlaces i.i.d.

Ejemplos de Redes: Amistades y romances en estudiantes secundaria

- La siguiente red contrasta en varios aspectos con la anterior.
- Se oberva la presencia de homofilia: El 52 % son blancos y el 85 % de sus relaciones de amistad son con blancos. Los hispanos están má integrados pero son muchos menos.

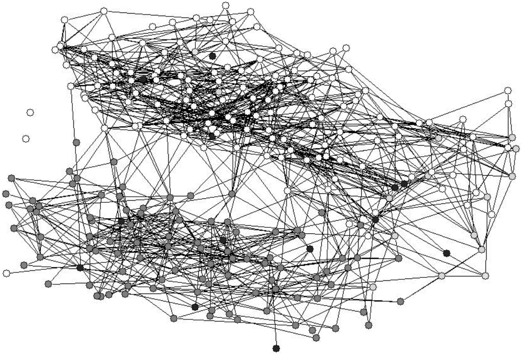


Figure 1.3: "Add Health" Friendships among High School Students Coded by Race: Hispanic=Black, White=White, Black=Grey, Asian and Other = Light Grey

Contenido

Ejemplos de Redes: Formación aleatoria de redes

- Modelo de Erdos - Renyi 1960.
- Sea $N = \{1, ...n\}$ un conjunto de nodos (vertices).
- Suponga que la probabilidad de que se forme un enlace entre i, j es p y que la formación de enlaces es independiente.[2]
- Entonces la probabilidad de que se forme una red con m enlaces es:

$$p^m(1-p)^{\left(\frac{n(n-1)}{2}-m\right)} \tag{2}$$

- Teniendo esta probabilidad se pueden calcular estadísticas descriptivas. El grado de un nodo es el número de nodos con los cuales tiene un enlace.

[2] Formalmente se puede pensar en una variable aleatoria con valores en un espacio de matrices simétricas de ceros y unos.

- La probabilidad de que un nodo i tenga exactamente d enlaces es:

$$comb(n-1,d)p^d(1-p)^{n-1-d} \tag{3}$$

- Cuando n es grande y p es pequeño esta expresión se puede aproximar por una distribución de Poisson (en ocasiones el modelo binomial independiente lleva este nombre - modelo de Poisson).

$$\frac{e^{(n-1)p}((n-1)p)^d}{d!} \tag{4}$$

Ejemplos de Redes: Formación aleatoria de redes

- Las siguientes figuras muestran una red generada usando el modelo binomial ($n = 50$, $p = 0{,}02$, lo cual implica que el valor esperado del grado de un nodo es 1).
- Características sobresalientes de grafos generados de esta forma: la probabilidad de cíclos es baja y existe una componente conexa grande.

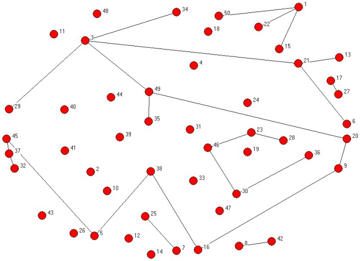


Figure 1.4: A Randomly Generated Network with Probability .02 on each Link

- El siguiente gráfico compara la distribución muestral con la aproximación de Poisson para la distribución binomial.

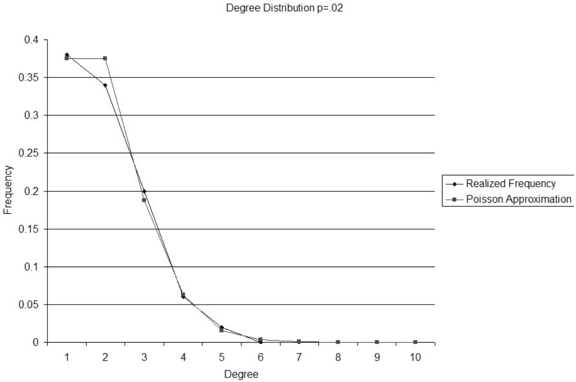


Figure 1.5: Frequency Distribution of a Randomly Generated Network and the Poisson Approximation for a Probability of .02 on each Link

Ejemplos de Redes: Formación aleatoria de redes

- Al aumentar la probabilidad de enlaces disminuyen las componentes conexas (maximales), aumenta la componente dominante y empeora la aproximación de Poisson al grado del grafo.

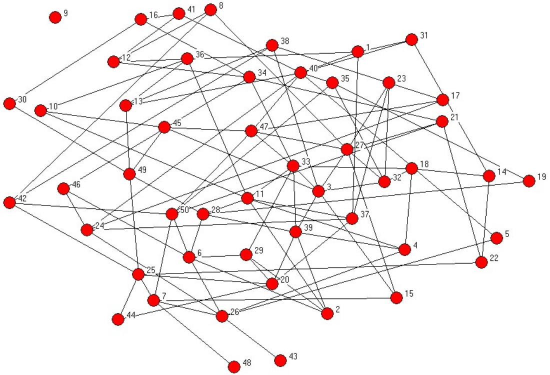


Figure 1.6: A Randomly Generated Network with Probability .08 of each Link

Ejemplos de Redes: Formación aleatoria de redes

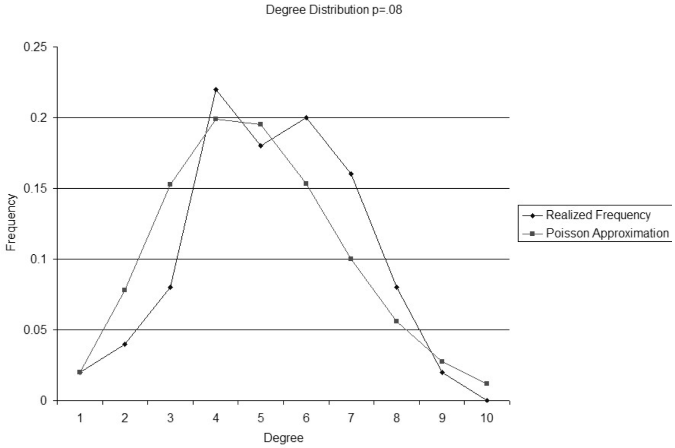


Figure 1.7: Frequency Distribution of a Randomly Generated Network and the Poisson Approximation for a Probability of .08 on each Link

- Jackson y Wolinsky (1996). Intruducen el siguiente modelo de conexiones simétricas.
- Sea (N, g) una red (grafo) donde N es el conjunto jugadores y g es una matriz simétrica de ceros y unos que representa la red ($g_{ij} = 1$ si y solamente si existe un enlace entre el nodo i, j).
- La utilidad $u_i(g)$ que recibe un jugador i de una red g es:

$$u_i(g) = \sum_{j \neq i, i\ j \text{ con un camino que los conecta en } g} \delta^{l_{ij}(g)} - d_i(g)c$$

donde $\delta \in (0, 1)$, $l_{ij}(g)$ es el número de vértices en el camino más corto ente i, j, d_i es el grado de i y c es una constante que refleja el costo para un individuo de mantener un enlace.

- Decimos que una red es eficiente si maximiza la suma de la utilidad de todos los agentes.
- Algunas observaciones: (1) Si el costo es muy bajo (inferior a $\delta - \delta^2$) el grafo óptimo es tener el máximo número de elaces. Si el costo es mayor que ese umbral pero no muy grande, el grafo óptimo es una estrella.
- Teniendo esta caracterización como referente ahora podemos preguntarnos cuál es el resultado de la interacción estratégica de un conjunto de jugadores en una red.

Ejemplos de Redes: Formación estratégica de redes

- Para esto necesitamos un concepto de equilibrio (una teorá positiva del comportamiento).
- Estabilidad por pares es: (1) Ningun agente puede beneficiarse de romper con un enlace en el que está directamente involucrado. (2) Ningun par de agentes puede beneficiarse (almenos uno de ellos estrictamente) de crear un nuevo enlace.

- Cuando el costo es bajo existe una única red estable por pares que es la red que tiene el máximo número de enlaces.
- Cuando el costo es mayor, la estrella es estable por pares pero hay muchas más estructuras de redes que son eficientes. Por ejmplo, cuatro jugadores conectados en un círculo puede ser estable en el rango de costos en el que la estrella es estable y eficiente.
- Esto pone de manifiesto la posibilidad de que en equilibrio las redes no necesariamente son eficientes.
- Con costos aún mayores la estrella puede ser eficiente pero no estable por pares.

Contenido

Grafos

- Un grafo $G = (N, g)$ donde $N = \{1, ...n\}$ es un conjunto de nodos y g es una matriz $n \times n$.
- $g_{ij} = 1$ si i, j estan relacionados por g.
- Convenimos que $g_{ii} = 0$.
- Un grafo es no dirigido (simplemente un grafo) si $g_{ij} = g_{ji}$. Dirigido en caso contrario.
- Gran cantidad de definiciones que describen características de los grafos son obvias: Caminata, camino, ciclo, geodésica, grafo conexo, subgrafo, componente (subgrafo conexo maximal), puente entre componentes (es un vertice que al eliminarlo aumenta el número de componentes).

- Árboles, bosques, circulos, estrellas, red completa (todos lo enlaces posibles están presentes).
- Vecindario de un vertice (no incluye el mismo), grado de un vertice, vecindarios de grado 2, etc.

Grafos

- 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 razon de las probabilides de ciertos grados no depende de la escala.
- Tiene colas (en ambos lados) más gruesas que Poisson.

Grafos

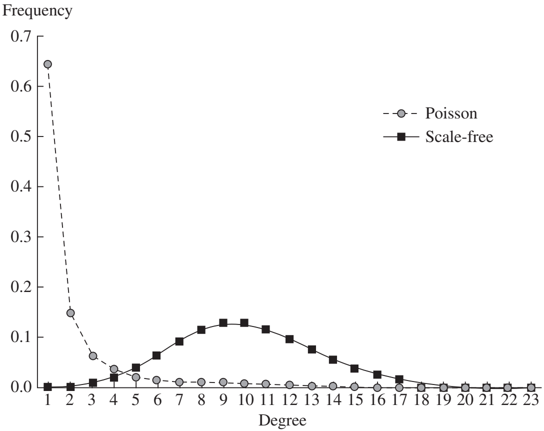


FIGURE 2.8 Comparing a scale-free distribution to a Poisson distribution.

Grafos

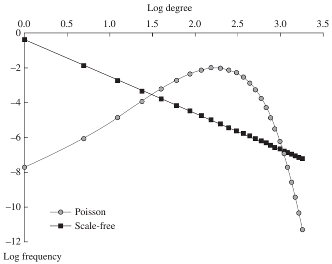


FIGURE 2.9 Comparing a scale-free distribution to a Poisson distribution: log-log plot.

- La distancia entre dos nodos es el número de vertices de una geodésica (cuando no hay un camino entre los nodos es infinito).
- El diametro es la mayor distancia entre cualquier par de nodos. *Girth* es el la distancia del ciclo más corto en la red (infinito si no hay ciclos). La circunferencia es la distancia del ciclo más grande en la red (cero si no existen ciclos).

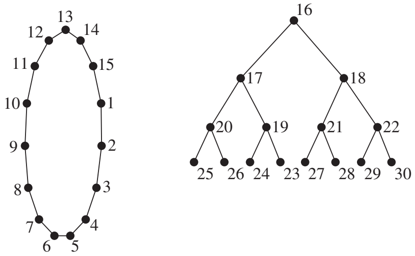


FIGURE 2.10 Circle and tree.

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

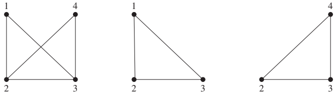


FIGURE 2.11 A network on four nodes and its two cliques.

Grafos: Aglomeración

- Medir el número y tamaño de cliques es valioso pero muy sensible a cambios pequeños en la red.
- La medida más común es: Mira 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:

$$Cl(g) = \frac{\sum_i \{jk \in g : k \neq j; k, j \in N_i(g)\}}{\sum_i \{jk : k \neq j; k, j \in N_i(g)\}} \tag{5}$$

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

Grafos: Centralidad

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

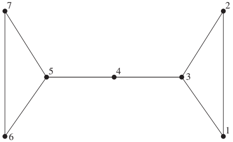


FIGURE 2.13 A central node with low degree centrality.

Grafos: Centralidad

- 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:

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

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

$$P^K(g) = \hat{g} P^K(g) \tag{7}$$

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

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

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

  donde λ es la constante de proporcionalidad.
- Este problema se reduce a encontrar los valores propios y vectores propios de g. La convención es usar el vector propio correspondiete al mayor valor propio (cuando es positivo).

- 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

Contenido

- 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.
- Una interpretación de los grafos bipartitos es que representan las potenciales relaciones que podrían darse entre los nodos. El problema típico se puede plantear como: Si A, B es 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.
- La propiedad de ser indpendiente se preserva cuando se pasa a subgrupos. Sin embargo, para subgrafos propios simepre existe un conjunto independiente (maximal) que no es independiente (maximal) en la extensión.

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.

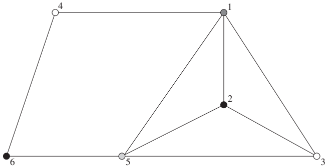


FIGURE 2.16 A planar network on six nodes with chromatic number 4.

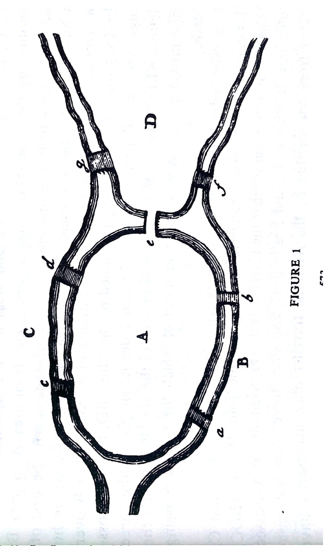
Toros Eulerianos y Ciclos Hamiltonianos

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

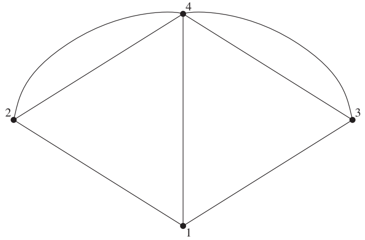
Theorem

Una red conexa tiene un un circuito si y sólo si el grado de cada nodo es par.

Toros Eulerianos y Ciclos Hamiltonianos



Toros Eulerianos y Ciclos Hamiltonianos



Toros 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.