Introducción a las redes, su predicción usando redes neuronales

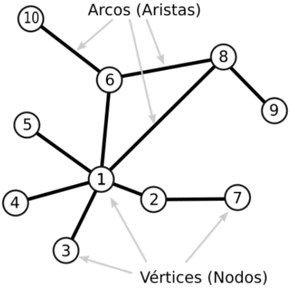
L. Leticia Ramírez Ramírez,
Centro de Investigación en Matemáticas (CIMAT), México

leticia.ramirez@cimat.mx

RIIAA, 2021

Introducción a Redes

Grafo (Gráfica)/Red



- Sociología (Social Network Analysis)
- Mathematicas (Grafos, Graficas)
- Sciencias Computacionales (Grafos, Graficas)
- Estadística, Física (Redes complejas)
- Economia (Redes)
- Bioinformatica (Redes)

Ejemplos

- Redes sociales:
  - Redes de amigos, conocidos
  - Redes de colaboración
  - Redes de llamadas telefonicas
- Redes tecnológicas:
  - Internet
  - Redes telefonicas
  - Redes de trasporte
- Redes biológicas
  - Redes de interacción de proteinas
  - Redes de regulación genética
  - Redes de actividad cerebral
  - Redes sociales animales
- Economía
  - Redes de instituciones, gobiernos, regiones

Grafo (Gráfica)/Red

Definición: Grafo

Un grafo (gráfica) se define como el par $\mathcal{G} = \{V(\mathcal{G}), E(\mathcal{G})\} = \{V, E\}$, donde $V = \{v_1, v_2, \ldots, v_n\}$ es el conjunto de vértices (nodos) y $E = \{e_1, e_2, \ldots, e_m\} \subset V \times V$ es el conjunto de arcos (aristas).

En este caso consideramos que $n < \infty$ pero también se pueden considerar el caso *infinito*.

- El **orden** del grafo (denotado como $|V|$) es su número de vértices (n).
- El **tamaño** del grafo (denotado como $|E|$) es su número de arcos (m)

Grafo (Gráfica)/Red

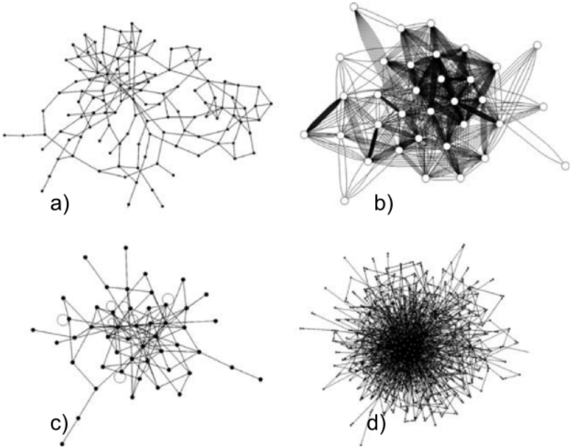
Decimos que

- Dos vértices u y v son **adyacentes** si exite un arco $e \in E$ tal que e=(u,v). En este caso u y v son **incidentes** a e y e es **incidente** a u y v.
- Dos arcos e_1 y e_2 son **adyacentes** si ambos son incedentes al menos a un vértice común.

Definición: Grado

El grado de un vértice es el número de arcos que son incidentes a éste.

Diversos Grafos



a) Circuito electronico, b) Frecuencia en interaccion entre grupos de investigación, c) Interacción de proteína-proteína en el virus de herpes, d) Red metabolica de *Helicobacter pylori*. Fuente: Ernesto Estrada, 2011.

Red direccionada con pesos en arcos

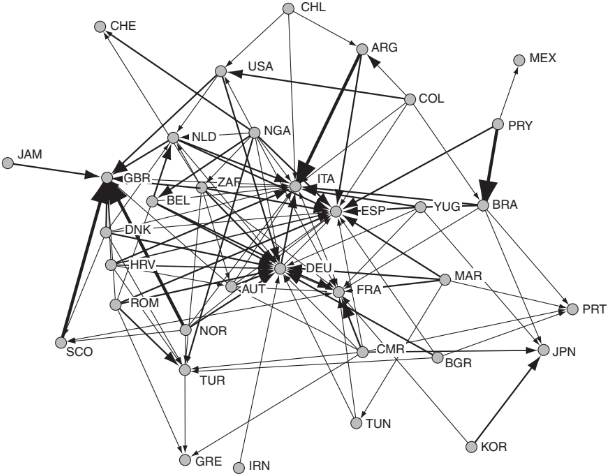


Fig. 1.5
Weighted directed network. Representation of the transfer of football players between countries after the 1998 World Cup in France. Links are drawn, with thickness proportional to the number of players transferred between two countries.

Fuente: Ernesto Estrada, 2011.

Grafo (Gráfica)/Red

- Si permitimos múltiples arcos entre dos vertices, tenemos un **grafo múltiple**.
- Si un arco puede tener como vértices incidentes al mismo vértice, este arco se llama **bucle**.
- Si en $\mathcal{G}$, $e = (v_i, v_j)$ es equivalente al arco (v_j, v_i) para todo $e \in E$ entonces el grafo es **no direccionado**.

Definición: Grafo Simple

Un grafo se denomina simple si es un grafo sin múltiples arcos, sin pesos, sin bucles y es no direccionado.

Grado

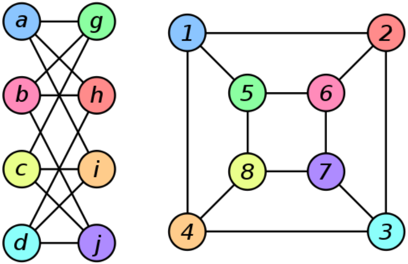
- Cuando el grafo es direccionado, podemos definir el **grado interior** y **grado exterior** de un vértice como el número de arcos que se dirigen y salen del vértice, respectivamente.
- Un vértice con grado 0 se denomina **aislado**.
- Si cada vértice en el grafo tiene grado k ($k \in \mathbb{N}$), el grafo se denomina $k-$regular.
- La **sucesión de grados** de un grafo es una sucesión no creciente de grados de sus vértices.
- Un grafo es **completo** cuando cualquier par de sus vértices son adyacentes. En este caso, el grafo de orden n es $k-$regular con k igual a ...

Isomorfismo en Grafos

Definición 4

Decimos que dos grafos $\mathcal{G}_1$ y $\mathcal{G}_2$ son isomorfos si hay una biyección $f : V(\mathcal{G}_1) \to V(\mathcal{G}_2)$ tal que $(v_i, v_j) \in E(\mathcal{G}_1)$ ssi $(f(v_i), f(v_j)) \in E(\mathcal{G}_2)$.

Ejemplo:



$$
\begin{aligned}
f(a) &= 1\\
f(b) &= 6\\
f(c) &= 8\\
f(d) &= 3\\
f(g) &= 5\\
f(h) &= 2\\
f(i) &= 4\\
f(j) &= 7
\end{aligned}
$$

Observaciones

- Si dos grafos son isomorfos, su sucesión de grados coincide, pero si la suceción de grados de dos grafos son iguales, no necesariamente son isomorfos. De un ejemplo de ésto.
- Dada una sucesión de enteros, ¿Es siempre posible construir un grafo simple?. ¿Existe uno para 2,2,2,2,1?
- Dada una sucesión de enteros, ¿Es siempre posible construir un grafo no direccionado que le corresponda?. ¿Puedes contruir un grafo para la sucesión de grados 6,3,3,2,2?

Matriz de Adyacencias

Definición: Matriz de adyacencias

Para un grafo simple, la matriz de adyacencia de $\mathcal{G}$, se define como la matriz cuadrada A de orden n donde su ij elemento es igual a

$$A_{ij} = \begin{cases} 1 & \text{si } (v_i, v_j) \in E(\mathcal{G}) \\ 0 & \text{eoc} \end{cases}$$

Entonces

- Para una red sin bucles, A ...
- Para una red no direccionada, A ...
- Para una red múltiple, A puede extenderse a ...
- Si la red tiene bucles, A ...
- Si a una red los arcos tienen *pesos*, A puede extenderse a ...

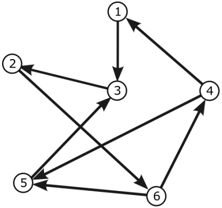
Matriz de Adyacencias: Caso direccionado

Definición: Matriz de adyacencias

Para un grafo direccionado y sin múltiple arcos, la matriz de adyacencia de $\mathcal{G}$, con orden n, se define como la matriz cuadrada A de orden n, donde su ij elemento es igual a

$$A_{ij} = \begin{cases} 1 & \text{si el arco que va de } j \text{ a } i \text{ es elemento de } E(\mathcal{G}) \\ 0 & \text{eoc} \end{cases}$$

Ejemplo:



	1	2	3	4	5	6
1	0	0	**1**	0	0	0
2	0	0	0	0	0	**1**
3	0	**1**	0	0	0	0
4	**1**	0	0	0	**1**	0
5	0	0	**1**	0	0	0
6	0	0	0	**1**	**1**	0

Matriz de Adyacencia, Grados y tamaño

Proposición 1

Sea $\mathcal{G}$ un grafo no direccionado, con orden n y tamaño m. Si k_i el grado del vértice i ($i \in \{1, \ldots, n\}$) entonces

$$k_i = \sum_{i=1}^{n} A_{ij} \qquad \text{y} \qquad 2m = \sum_{i=1}^{n} k_i = \sum_{j=1}^{n} \sum_{i=1}^{n} A_{ij}.$$

Proposición 2

Sea $\mathcal{G}$ un grafo direccionado, con orden n y tamaño m. Si k_i^{in} y k_i^{out} son los grados interior y exterior de $i \in \{1, \ldots, n\} = V(\mathcal{G})$, respectivamente, entonces

$$k_i^{in} = \sum_{j=1}^{n} A_{ij}, \;\; k_j^{out} = \sum_{i=1}^{n} A_{ij} \;\; \text{y} \;\; m = \sum_{i=1}^{n} k_i^{in} = \sum_{j=1}^{n} k_j^{out} = \sum_{j=1}^{n} \sum_{i=1}^{n} A_{ij}$$

Camino (Walk)

Definición 6

Un camino (*walk*) en un grafo $\mathcal{G}$ es una sucesión de vértices $v_1, v_2, \ldots, v_k$ donde $(v_i, v_{i+1}) \in E(\mathcal{G})$, $\forall i = 1, 2, \ldots, k-1$.

Definición 6

La longitud de un camino P es el número de arcos que contiene.

Si la gráfica $\mathcal{G}$ es simple, el número de diferentes caminos de longitud 2, $N_{ij}^{(2)}$, es igual a

$$N_{ij}^{(2)} = \sum_{k=1}^{n} A_{ik} A_{kj} = A^2.$$

Más aun, el número de caminatas de longitud k es igual a

$$N_{ij}^{(k)} = A^k.$$

Camino (Walk)

Definición 7

El *camino* más corto entre dos vértices i y j se define como

$$\min\{k \mid [A^k]_{ij} > 0\}$$

Ejemplo:

- Cuando los arcos tienen pesos, el camino más corto entre dos vértices se considera como el camino que minimiza la suma de los pesos.
- El problema de encontrar el camino más corto tiene diversos algoritmos entre los que destacan
  - Dijkstra
  - Bellman-Ford
  - Búsqueda A*
  - Viterbi

Más sobre conectividad

- Un **camino** es cerrado si $v_0 = v_k$.
- Un **recorrido** (trail) es un camino en el que todos los arcos son distintos.
- Un **camino simple** (path) es un camino en el que todos los vertices y arcos son distintos.
- Un **ciclo** es un recorrido que es cerrado (ie todos los vértices son distintos exceptuando v_0 y v_k)
- Dos vértices i y j **están conectados** si existe un camino de i a j.
- Un **grafo es conectado** si cualquier para de vértices están conectados

Métricas y estructuras topológicas en redes

Métricas y estructuras topológicas en redes

- Diámetro
- Coeficiente de Agrupamiento
- Centralidad
- Asortatividad

Diámetro

Definición: Diámetro

El diámetro de una red es el más largo de los caminos más cortos entre todos los vértices de la red.

El diámetro mide la conectividad de la red en el mismo sentido que la "Teoría de los seis grados de separación" afirma que en el mundo no hay más de cinco intermediarios de conocidos entre cualesquiera dos personas. La teoría fue inicialmente propuesta en 1930 por el escritor húngaro Frigyes Karinthy en un cuento llamado Chains y puesta a prueba experimentalmente por primera vez por el psicologo social estadounidense Stanley Milgram en 1967.

Diámetro

Conderamos una de las redes en **NetworkX**: "el club de karate de Zachary"(Zachary, 1977). Esta red representa las amistades entre miembros de un club de karate estudiado entre 1970 y 1972.

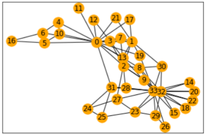
Este club de karate en particular ha sido de interés porque finalmente se dividió en dos clubes diferentes después de un desacuerdo entre el instructor y el presidente del club.

En el estudio original, Zachary usó la estructura de la red para predecir qué miembros se unirían a cuál de los dos clubes con una precisión casi perfecta. Específicamente, usó el algoritmo de corte mínimo. En el artículo de Zachary, el nodo con el ID 0 se identificó como el instructor del club.

Diámetro

In [3]:
```
import networkx as nx
import matplotlib.pyplot as plt
```

In [95]:
```
G = nx.karate_club_graph()
nx.draw_networkx(G,node_color="orange")
```



In [13]:
```
list(G.nodes)[0:10]
```

Out[13]:
```
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
```

In [14]:
```
list(G.edges)[0:10]
```

Out[14]:
```
[(0, 1),
 (0, 2),
 (0, 3),
 (0, 4),
 (0, 5),
 (0, 6)
```

In [96]:
```
nx.diameter(G)
```

Out[96]:
```
5
```

Coeficiente de agrupamiento

Definición: Coeficiente de Agrupamiento

El coeficiente de agrupamiento (clustering coefficient) del vértice i se define como

$$c_i = \begin{cases} \frac{E_i}{k_i(k_i-1)/2} & \text{si } k_i > 1 \\ 0 & \text{eoc} \end{cases}$$

donde E_i es el número de arcos que conectan a vecinos de i.

- Este coeficiente es una medida de la propiedad "los amigos de mis amigos son mis amigos".
- Como el número máximo de conexiones que pueden existir entre los amigos es $k_i(k_i - 1)/2$, el coeficiente es la razón entre las conexiones existentes entre el máximo número posible.
- El **coeficiente de agrupamiento de la red** se define como el promedio de los coeffientes de agrupamiento de todos sus nodos.
- A este coeficiente también se le denomina **Tansitividad de la red**.
- Un coeficiente de agrupamiento cercano a 1 es una indicador de la propiedad de mundo pequeño.

Coeficiente de agrupamiento

In [97]:
```
nx.clustering(G)
```

Out[97]:
```
{0: 0.15,
 1: 0.3333333333333333,
 2: 0.24444444444444444,
 3: 0.6666666666666666,
 4: 0.6666666666666666,
 5: 0.5,
 6: 0.5,
 7: 1.0,
 8: 0.5,
 9: 0,
 10: 0.6666666666666666,
 11: 0,
 12: 1.0,
 13: 0.6,
 14: 1.0,
 15: 1.0,
 16: 1.0,
 17: 1.0,
 18: 1.0,
 19: 0.3333333333333333,
 20: 1.0,
 21: 1.0,
 22: 1.0,
 23: 0.4,
 24: 0.3333333333333333,
 25: 0.3333333333333333
```

In [98]:
```
nx.average_clustering(G)
```

Out[98]:
```
0.5706384782076823
```

Centralidad (de grado)

No existe una única medida de centralidad, pero éstas buscan extraer la importancia que los nodos en la red. Buscan contestar a preguntas como "¿Qué vértice en la red debe ser escogido para asegurar que un mensaje o información se disperse al mayor número de vértices posible? En el caso de modelos epidemiológicos estos individuos son candidatos a ser vacunados o aislados.

Definición: Centralidad de grado

La centralidad de grado (degree centrality) del vértice i considera que un nodo es importante si tiene muchos vecinos. Bajo esta idea, la centralidad está definida por su grado.
Si la red es direccionada hay dos versiones de este grado: la centralidad hacia dentro y hacia afuera. Si el nodo es más importante por cuantos se ligan a éste, el grado hacia adentro es el que se considera.

Centralidad (de grado)

In [99]:
```
G.degree
```

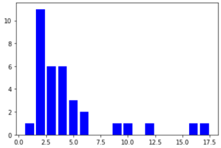
Out[99]:
```
DegreeView({0: 16, 1: 9, 2: 10, 3: 6, 4: 3, 5: 4, 6: 4, 7: 4, 8: 5, 9: 2, 10: 3, 11: 1, 12: 2, 13:
5, 14: 2, 15: 2, 16: 2, 17: 2, 18: 2, 19: 3, 20: 2, 21: 2, 22: 2, 23: 5, 24: 3, 25: 3, 26: 2, 27:
4, 28: 3, 29: 4, 30: 4, 31: 6, 32: 12, 33: 17})
```

In [106]:
```
max(G.degree)
```

Out[106]:
```
(33, 17)
```

In [105]:
```
degree_sequence=sorted([d for n,d in G.degree()], reverse=True)
import collections
degreeCount=collections.Counter(degree_sequence)
deg, cnt = zip(*degreeCount.items())
plt.bar(deg, cnt, width=0.80, color='b')
```

Out[105]:
```
<BarContainer object of 11 artists>
```



Centralidad (Eigencentralidad)

Definición: Eigencentralidad

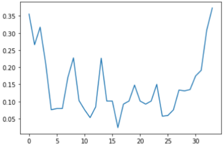
La eigencentralidad del nodo i se define como la $i-$ésima entrada del eigenvector asociado con el mayor de los eigenvalores de A.

- La centralidad de eigenvector (eigenvector centrality) es más sofisticada que la de grado ya que depende no sólo de su grado sino de que tan bien estén, a su vez, conectados sus vecinos, y los vecinos de sus vecinos, y así sucesivamente.
- Una justificación a este índice se puede construir a partir del teorema de Perron-Frobenious aplicado a la matriz $B = A + I$ ya que B es simétrica, con eigenvalores reales y eigenvalores iguales a los de A más uno.
- La forma en que Google califica a las páginas web según su importancia se basa en esta idea.

Centralidad (Eigencentralidad)

In [114]:
```
ec=nx.eigenvector_centrality(G, max_iter=10000)
lists = sorted(ec.items()) # sorted by key, return a list of tuples
x, y = zip(*lists) # unpack a list of pairs into two tuples
plt.plot(x, y)
```

Out[114]: [<matplotlib.lines.Line2D at 0x7fa52443f6d0>]



In [108]:
```
max(nx.eigenvector_centrality(G, max_iter=10000))
```

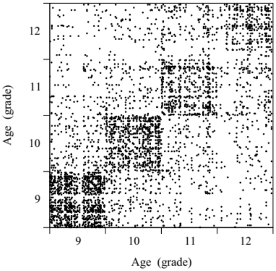
Out[108]: 33

Asortatividad

Definición: Asortatividad

La asortatividad es la preferencia de los nodos de una red por unirse a otros que le son similares en alguna característica (En una red poblacional, según raza o religión, por ejemplo).

Ejemplo:

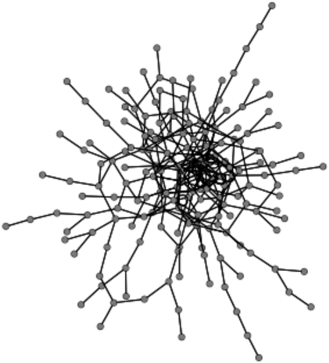


Fuente: Newman, 2010.

Cuando la asortatividad está en función de los grados de los nodos, ésta puede calcularse como una función de las correlaciones entre los nodos que tienen grado similar.

Asortatividad

Ejemplos de redes asortativas (a) y antisortativas (b).



(a)

(b)

Fuente: Newman, 2010.

Asortatividad

```
In [116]: nx.degree_assortativity_coefficient(G)

Out[116]: -0.47561309768461457
```