

ESTADÍSTICA DE  
NEGOCIOS



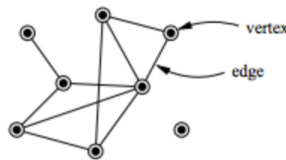
# TAREA SEMANA 2

## MODELOS DE DATOS DISCRETOS

# Tarea semana 2: Modelado de datos discretos

## Grafos

- Un grafo es una colección de nodos, ó vértices (vertex) interconectados por ramas (edge):



- Uno de los modelos más importantes en teoría de grafos y redes complejas es el de un grafo aleatorio.
  - Las grandes estructuras complejas que se encuentran en la naturaleza suelen estar formadas por interacciones individuales no correlacionadas entre un gran número de sus constituyentes, por lo que no sorprende que los grafos aleatorios proporcionen buenos modelos para muchas de ellas.
  - Aunque las estructuras hechas por el hombre son el resultado de un diseño racional (o así lo deseamos), a menudo exhiben suficiente "aleatoriedad", ya sea porque su estructura se desarrolló con el tiempo (como la World-Wide-Web), o porque copian un modelo anterior de red existente (por ejemplo, red aeroportuaria, red eléctrica) o están sujetas a restricciones aleatorias o de otro tipo (por ejemplo, geográficas).
  - Los grafos aleatorios con las características adecuadas pueden servir como modelos útiles para la evolución y el funcionamiento de muchas estructuras artificiales complejas.
  - Si desea conocer más acerca de grafos puede revisar las siguientes lecturas no obligatorias:
    - <https://www.math.tugraz.at/~kang/papers/Kang-Intelligencer.pdf>
    - <https://epubs.siam.org/doi/pdf/10.1137/S003614450342480>

## Problema: Discrete Logistics

- La empresa Discrete Logistics ha realizado un análisis de su cadena logística en una región del país donde opera.
- Ha recopilado la información de 50 centros logísticos y registra el número de otros centros con los que tiene conexión directa (envío y recibo de paquetería).
  - La red de Discrete Logistic puede considerarse una red aleatoria.
  - El número de conexiones que tiene cada centro se conoce como grado del nodo en teoría de grafos.
    - Esto está registrado en la variable Enlaces para cada uno de los 50 centros.

### Archivo de datos:

- [DL\\_data.csv](#)
- Lea el archivo csv.

```
import pandas as pd
import seaborn as sns

data = pd.read_csv("../data/DL_data.csv") # si la estructura es la del repositorio de GitHub

# si está trabajando en local y el archivo de datos y el notebook están en el mismo directorio
# data = pd.read_csv("DL_data.csv")

# También puede leer el archivo desde el repositorio de GitHub.
#data = pd.read_csv("https://raw.githubusercontent.com/marsgr6/EN-online/8a1cee296279c274d8f05f3905fce97c1f7636a2/data/DL_data.csv")

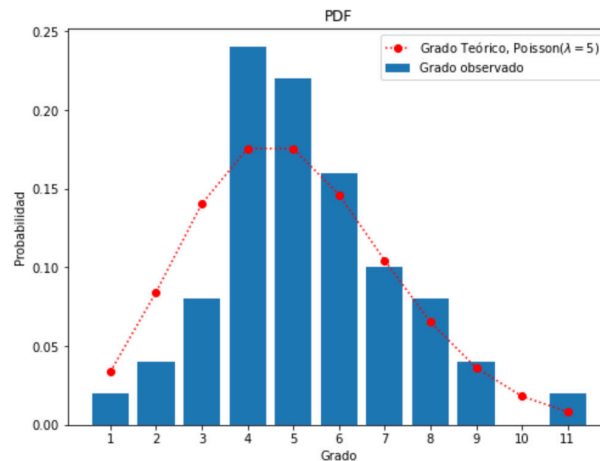
data.head()
```

| Centro | Enlaces |
|--------|---------|
| 0      | 1       |
| 1      | 2       |
| 2      | 3       |
| 3      | 4       |
| 4      | 5       |

- De acuerdo a la teoría de grafos, la distribución del grado de los nodos en una red aleatoria se puede modelar como una variable aleatoria de Poisson  $X$  - Poisson( $\lambda=n \times p$ ), donde  $n$  es el número de nodos y  $p$  la probabilidad de que exista una conexión entre cualquier pareja de nodos.
  - En nuestro estudio se ha determinado que estos valores son  $n=50$ ,  $p=0.1$  para la red de Discrete Logistics.
    - Es decir el grado de la red puede ser modelado con la siguiente distribución  $X$ -Poisson ( $\lambda=5$ ).
- Discrete Logistics está interesado en reconfigurar su red logística y quiere identificar los centros con pocas conexiones que eventualmente pueden ser cerrados.

1. (1 pt) Cuál es la distribución del grado de los centro logísticos.

- Escriba el código Python para replicar el gráfico a continuación.
- El diagrama de barras es la distribución de la variable **Enlaces**. La curva roja representa la distribución X-Poisson( $\lambda=5$ ).



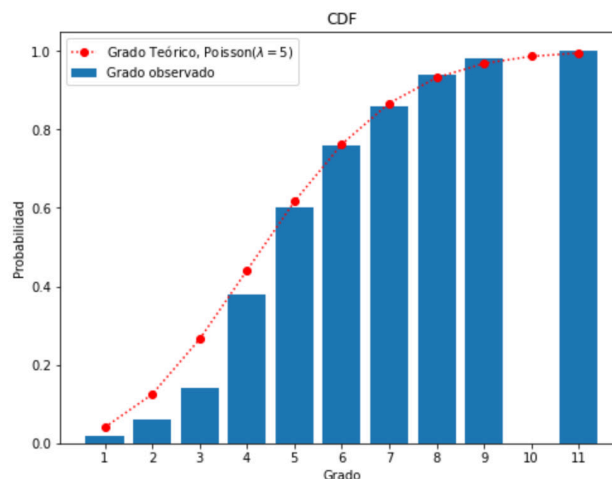
2. (2 pt) Cuál es la probabilidad de que un centro cualquiera tenga 3 o menos conexiones.

- De acuerdo a los valores observados.
- De acuerdo a la distribución de Poisson X-Poisson( $\lambda=5$ ).
- Responda usando la implementación en Python de Poisson.

3. (1 pt) Cuál es la probabilidad de que un nodo tenga entre 4 y 6 conexiones.

- Responda usando la implementación en Python de Poisson.

4. (1 pt) Replique el histograma acumulado y la función de distribución acumulada para la distribución de Poisson:



5. **(2 pt)** Responda usando la CDF:

- Cuál es la probabilidad de que un centro cualquiera tenga 3 o menos conexiones
- Cuál es la probabilidad de que un nodo tenga entre 4 y 6 conexiones.

6. **(1 pt)** Calcule el valor esperado y varianza de la distribución analítica, compare con la media y varianza muestral.

7. **(2 pt)** Discuta qué tan adecuada considera esta distribución para modelar nuestros datos observados.