

Reto | Análisis de empleados en una red de correo electrónico*Entregable final del reto**Documento editable por el participante***Nombre del participante:****Indicaciones:**

- Descargar el archivo **email-enron-only.mtx**, desde los archivos adjuntos de la plataforma.
- Abre los archivos descargados para conocer su estructura y familiarizarte con la información contenida en ellos.
- Genera una libreta en **Google Colab** para resolver el reto con *scripts* de Python.
- Crear un *dataframe* en Pandas para el archivo.
- Construye una red no dirigida a partir del *dataframe*. Utiliza NetworkX para tener una visualización adecuada de la red, puedes manipular el color y tamaño de nodos y arcos.
- **Descarga** tu libreta de **Google colab** con extensión .py y nómbrala respetando el siguiente lineamiento: **SC66_Reto_APELLIDOS_NOMBREdelparticipante.ipynb** (sin espacios)
- **Dale** permisos al archivo para que podamos evaluarlo (Opción: "Anyone with the link" en inglés, o "Cualquier persona que tenga el vínculo" en español).

Paso 1: Construcción de una red

Construye una red no dirigida a partir de *dataframe*. Utiliza NetworkX para tener una visualización adecuada de la red, puedes manipular el color y tamaño de nodos y arcos.

A) A partir de la red y usando NetworkX, responde las siguientes preguntas:

¿Cuál es el nodo con el grado máximo?	Nodo 105
¿Cuál es el nodo con el grado mínimo?	Nodo 98
¿Cuál es el diámetro de la red?	8
¿Cuál es la distancia promedio más corta?	2.967

¿Es una red bipartita?	No
------------------------	----

Paso 2: Identificación de la red

La red que estás analizando es conexas. Un grafo conexo puede dividirse en un grafo desconexo al remover ciertos nodos o arcos o ambos. La habilidad de la red para funcionar incluso después de remover nodos o arcos se llama robustez o resiliencia. Con ayuda de NetworkX responde lo siguiente:

B) Con ayuda de NetworkX responde lo siguiente:

¿Cuántos y cuáles nodos hay que remover para que la red sea desconexa?	1 nodo, el 85
¿Cuántos y cuáles arcos hay que remover para que la red sea desconexa?	1 Arco, el que se forma entre los nodos 112 y 80
Con base en lo anterior, ¿la red es resistente ante fallas, cambios o errores que pudieran ocurrir?	En general la red no dirigida es tiene buena resiliencia a fallas o errores. Sin embargo, al tener nodos con centralidades altas, la red es vulnerable, lo cual hace que su resiliencia disminuya en ese caso específico.

Una vez encontrados los nodos y arcos del paso anterior, identificalos con un color diferente en la visualización de la red.

Paso 3: Calcula y visualiza las centralidades de la red

Las centralidades miden la importancia de los nodos basados en diferentes criterios. Con base en diferentes definiciones de importancia de un nodo, calcula la centralidad de grado, de intermediación y de cercanía. Muestra con un color diferente en la red a los 10 nodos más importantes para cada centralidad.

Paso 4: Crea una red dirigida

Ahora sí importa quién manda el correo y quién lo recibe. En esta nueva red calcula el PageRank de los nodos, y para visualizarlos en la red

C) Anota tus conclusiones:



Total de empleados, total de interacciones por correo electrónico.	143 Empleados, 623 Interacciones
¿Quién tiene el máximo grado, quién el mínimo, y qué significa?	Máximo grado Nodo 105, mínimo grado el nodo 98. Significa que el nodo 105 es aquel que la suma de correos entrantes y salientes es mayor que la de los demás, mientras que el nodo 98 es aquel que tiene un menor número de interacciones con los demás nodos en términos de cuántos emails manda y recibe.
¿Cuál es el diámetro de la red y qué significa?	La red tiene un diámetro de 8, y significa que es la mayor distancia entre dos nodos unidos por arcos, o la mayor distancia entre dos vértices de la red.
¿Cuál es el promedio de los caminos más cortos y qué significa?	2.967, es el promedio de las rutas más cortas entre todos los pares posibles conectados.
¿La red es robusta?, ¿cómo reacciona el sistema antes de fallas o ataques?	Sí, para el caso de fallas o errores, no tanto para el caso de ataques.
Tomando en cuenta las tres medidas de centralidad, ¿cuáles son los nodos más importantes?	17, 10, 11 y 105.
Según el PageRank, ¿quién es el nodo más importante? Y con base a qué se determina su importancia.	Nodo 17, es el que es más probable que reciba la mayor cantidad de correos, es decir, el destino más común de los correos enviados.