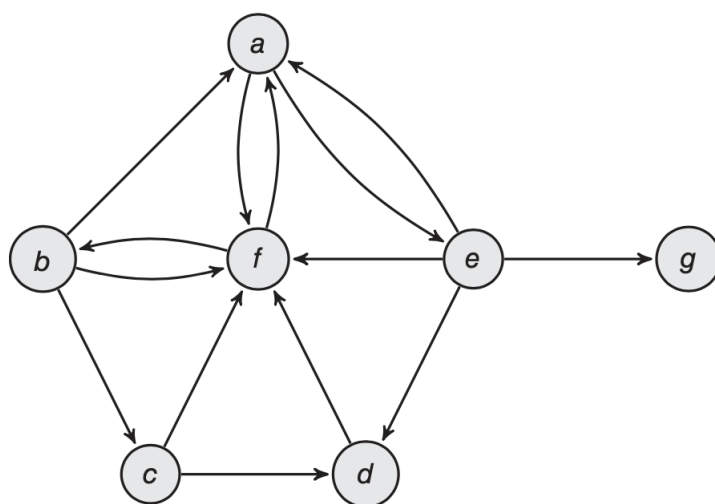


Redacta un documento .tex con formato de artículo en el que se exponga y se resuelva el siguiente problema. Para los cálculos, usar SAGE. Entregad en una carpeta el fichero .tex con todos los ficheros adjuntos que se requieran para su compilación, así como los que resultan de esta, que debe producir un pdf (que también se entrega).

### Problema sobre Page Rank y Familias Florentinas

El algoritmo de búsqueda de PageRank de Google se basa en el modelo de navegación aleatoria, que es una caminata aleatoria en el gráfico de la web. Para este gráfico, cada vértice representa una página de Internet. Una arista dirigida conecta  $i$  con  $j$  si hay un enlace de hipertexto de la página  $i$  a la página  $j$ . Cuando el usuario está en la página  $i$ , se mueve a una nueva página eligiendo entre los enlaces disponibles en  $i$  de manera uniforme y aleatoria.

La siguiente figura muestra una red simplificada con siete páginas.



La red está descrita por la matriz:

$$N = \begin{matrix} & \begin{matrix} a & b & c & d & e & f & g \end{matrix} \\ \begin{matrix} a \\ b \\ c \\ d \\ e \\ f \\ g \end{matrix} & \begin{pmatrix} 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 \\ 1/3 & 0 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1/4 & 0 & 0 & 1/4 & 0 & 1/4 & 1/4 \\ 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{matrix}.$$

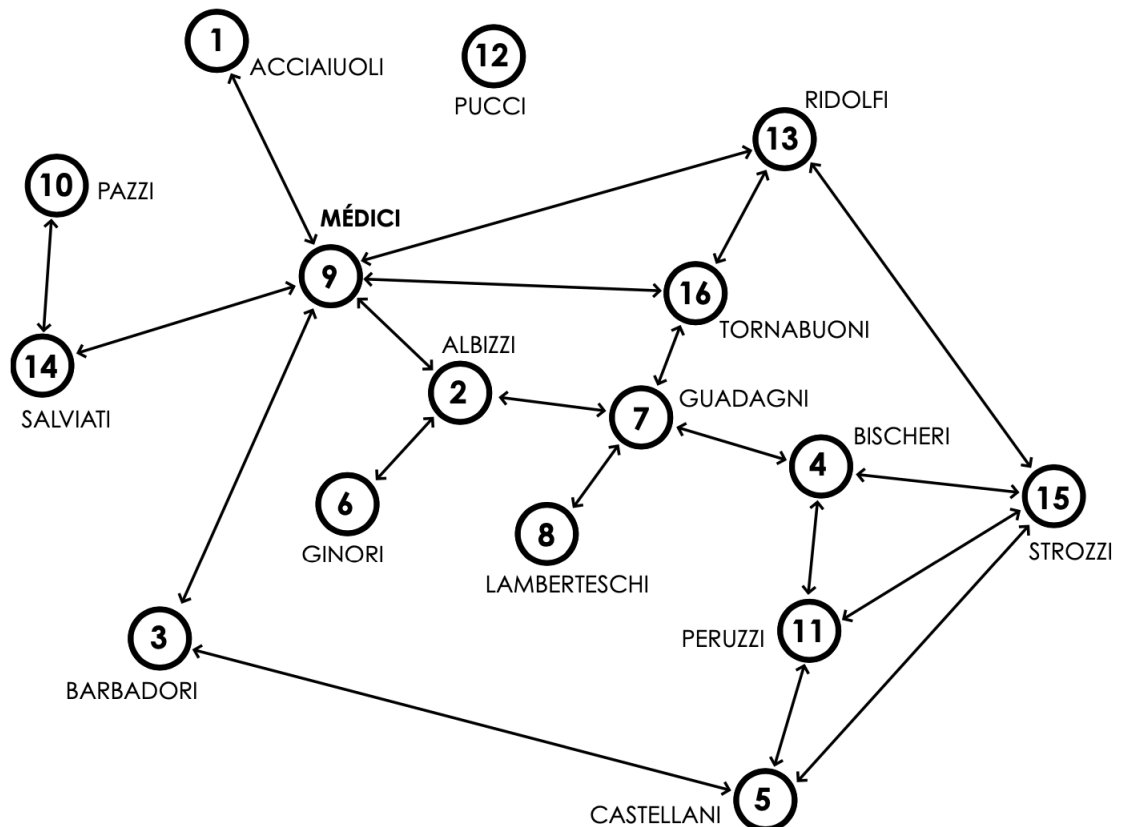
Tenga en cuenta que  $N$  no es una matriz estocástica, ya que la página  $g$  no tiene enlaces de salida, de modo que la fila asociada a  $g$  está formada completamente por ceros (y no suma 1). Para asegurarse de que el camino aleatorio llegue a todas las páginas de la red, el algoritmo debe tener en cuenta (i) las páginas que no tienen enlaces de salida, llamadas nodos colgantes, y (ii) los grupos de páginas que pueden hacer que el camino se atasque en un subgrafo. En la red del ejemplo,  $g$  es un nodo colgante. Suponga que la red consta de  $k$  páginas. En el algoritmo de PageRank, la solución para los nodos colgantes es asumir que cuando el usuario aleatorio llega a una página de este tipo, salta a una nueva

página en la red de manera uniforme y aleatoria. Se obtiene una nueva matriz  $Q$  donde cada fila de  $N$  correspondiente a un nodo colgante se cambia por una en la que todas las entradas son  $1/k$ . La nueva matriz  $Q$  es una matriz estocástica. Para el problema de atascarse potencialmente en pequeños subgrafos de la red, la solución propuesta en el artículo original por Brin y Page (1998) fue fijar un factor de amortiguamiento  $0 < p < 1$  para modificar la matriz  $Q$ . En su modelo, desde una página dada, el internauta aleatorio, con probabilidad  $1 - p$ , decide no seguir ningún enlace en la página y, en cambio, navegar a una nueva página aleatoria en la red. Por otro lado, con probabilidad  $p$ , sigue los enlaces de la página como de costumbre. Esto define la matriz de transición de PageRank

$$P = pQ + (1 - p)A,$$

donde  $A$  es una matriz  $k \times k$  cuyas entradas son todas  $1/k$ . El factor de amortiguación utilizado por Google se estableció originalmente en  $p = 0,85$ . Con el factor de amortiguación, la matriz de PageRank  $P$  es estocástica y el camino aleatorio resultante es aperiódico e irreducible. El PageRank de una página en la red es la probabilidad estacionaria de esa página.

1. Hacer los cálculos correspondientes con el grafo que se ha mostrado más arriba.
2. Un ejemplo clásico de red social de tamaño pequeño es el de las familias más influyentes en la política, la economía y la vida social de la Florencia del siglo XV. Los historiadores John Padgett y Christopher Ansell estudiaron esta red social en 1993, basándose en datos tomados de archivos históricos (ver el artículo [2], que se ha adjuntado en Recursos del aula virtual). El grafo que representa dicha red social, que no es dirigido, y se forma dibujando una arista entre dos familias cada vez que un matrimonio las vincula, es el siguiente:



A simple vista se observa que la familia Médici, con el vértice 9, ocupa una posición central, pues ostenta el grado máximo. Los autores del estudio conjeturaron que los Médici alcanzaron una posición de dominio en la sociedad florentina precisamente forzando su alianza con el resto de familias a través de múltiples matrimonios.

- Aplica el algoritmo PageRank a esta red social para confirmar que, en efecto, la familia Médici era la más relevante en esta pequeña red social.
- Repite el apartado anterior, tomando ahora  $p = 0,8ab$  y  $p = 0,6ab$ , donde  $ab$  son las dos últimas cifras de tu DNI. ¿Varía mucho el resultado? ¿Por qué?
- Explica el resultado que has obtenido. ¿Qué se puede decir de las otras familias florentinas a partir de los datos obtenidos?

## Referencias

- [1] J. M. ALMIRA, *Le reti sociali & La teoria dei grafi*, Vol. 3 de ‘La Matematica che trasforma il mondo’, RBA, 2020.
- [2] JOHN F. PADGETT; CHRISTOPHER K. ANSELL, Robust Action and the Rise of the Medici, 1400-1434, *The American Journal of Sociology* Vol. 98, No. 6. (May, 1993) 1259-1319.