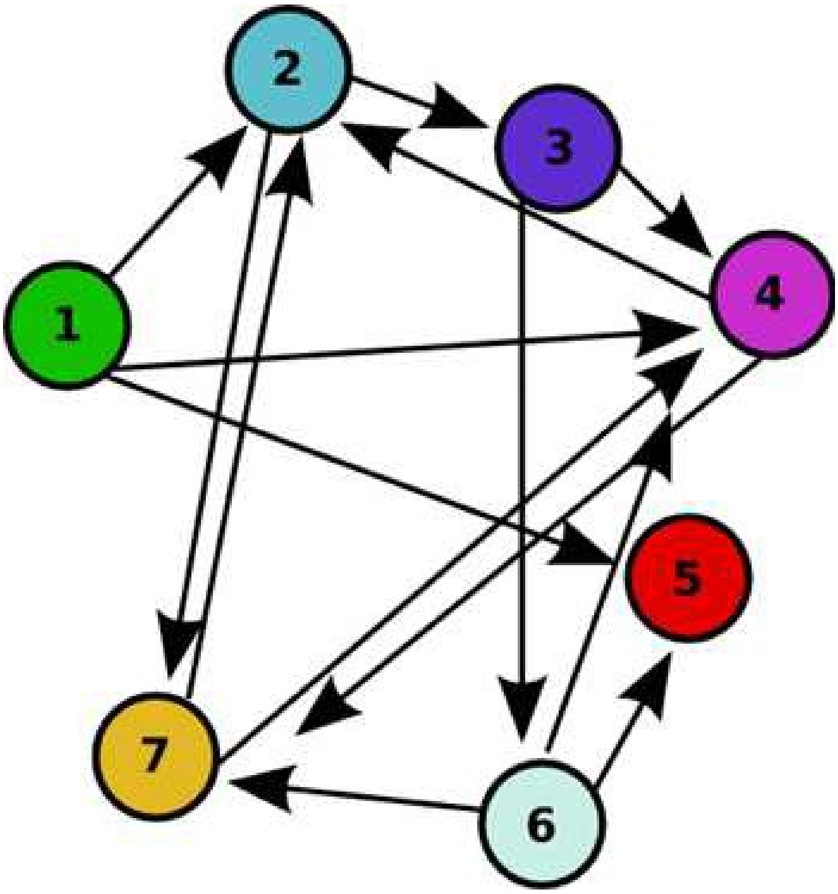
Calcular e interpretar el pagerank de un grafo 7x7

Sea el siguiente grafo 7x7.



Se desea calcular e interpretar su pagerank, utilizando la nomenclatura y definiciones de las trasparencias del curso (AulaVirtual).

Realizar los siguientes apartados.
La codificación deberá ser lo más 'elegante' y eficiente posible, y fácilmente 'escalable' (se debe poder aplicar a cualquier matriz A, con valores muy grandes de la dimensión de A).
En cada apartado se deberá entregar: el código utilizado, volcado de los datos, las gráficas, contestar a las preguntas e incluir comentarios, según corresponda en cada caso.

A. Calcular la matriz G de pagerank del grafo

- Definir la matriz de conectividad C del grafo 7x7 con el comando *sparse* de Matlab.

```
Cs=sparse(j,i,1,n,n); % Crea la matriz dispersa de tamaño nxn tal que S(j(k),i(k))=1. Los vectores i,j
del mismo tamaño
C=full(Cs)  %  Crea la matriz completa
```

- Utilizar el comando *sum* para calcular el vector Nj que contiene el número de links de salida de cada nodo. ¿Tienen salida todos los nodos?. En su caso, ¿Qué nodos no tienen salida?.

- A partir del vector Nj calcular el vector dj: dj(i)=1 si Nj(i)=0, dj(i)=0 en otro caso.

```
i=find(Nj==0);
```

- Calcular la matriz S. Comprobar que la matriz S es estocástica por columnas.
- Calcular la matriz G, con el parámetro alfa=0.85. Comprobar que la matriz G es estocástica por columnas.

B. Codificar la rutina del método de la potencia. Reescalar una matriz

- Codificar la rutina [lambda,x,iter]=potencia(A,tol,nmax,x0) que implementa el siguiente método de la método de la potencia.

$$\begin{cases} x1=\text{ones(n,1);}\ \ \text{\% vector arranque} \\ \text{for k= 1, 2,....} \\ \quad x = x1; \quad x = x/norm(x); \\ \quad x1 = A * x; \quad \lambda = x' * x1; \\ \text{end} \end{cases}$$

Las variables de entrada son: la matriz A, la tolerancia tol, el número de iteraciones nmax y el vector de arranque x0.
Las variables de salida son: lambda el autovalor dominante de A, su autovector asociado x y el número de iteraciones iter.
Los criterios de parada son alcanzar el número de iteraciones, o bien, que $\left|\lambda^{k-1} - \lambda^{k}\right| < tol$.

Incluir el código (que asigna las variables tol, x0 y nmax en caso de que el usuario ejecute la rutina con una sola variable de entrada)

```
if nargin == 1
    tol = 1e-12;     % Tolerancia
    x0=ones(n,1);    % Vector de arranque
    nmax=500;        % Nº máx iteraciones
end
```

Reescalar una matriz

- Definir una matriz aleatoria de dimensión 10x10 con el comando A=rand(10).
- Reescalar la matriz A: dividir cada columna de A por la suma de elementos de esa columa. Llamar B a la matriz reescalada. Comprobar que la matriz B es estocástica por columnas.

Ejecutar el método de la potencia

- Ejecutar la rutina *potencia* para calcular el autovalor y autovector dominantes de la matriz B.
- Calcular las precisiones obtenidas: $\left\|Bx - x\right\|, \left|\lambda - 1\right|$.

Teorema de Perron-Frobenius

- Comprobar si la matriz B verifica el teorema de Perron-Frobenius:
  - o Las condiciones de entrada. Enunciarlas y comprobar que B las verifica.
  - o Las condiciones de salida. Enunciarlas y comprobar que B las verifica. Utilizar el comando *eig(B).*

C. Calcular el pagerank del grafo

- Se denomina *pagerank* al vector proporcional al autovector dominante que verifica que la suma de sus elementos es 1.
- Modificar la rutina de la potencia para obtener como variable de salida el pagerank de la matriz G

[lambda,pagerank,iter]=potencia(G,tol,nmax,x0)

- Calcular las precisiones obtenidas: $x = pagerank,\ \|Gx - x\|, |\lambda - 1|$.
- Dar el comando *bar(pagerank)* para visualizar los resultados. ¿En qué orden mostrará Google las páginas del grafo?. Podeis utilizar la notación $P_1 > P_2 > P_3 > \ldots$
- Calcular el pagerank de la matriz S. Mostrar el orden de las páginas. ¿Coincide con el orden obtenido en el apartado anterior?.