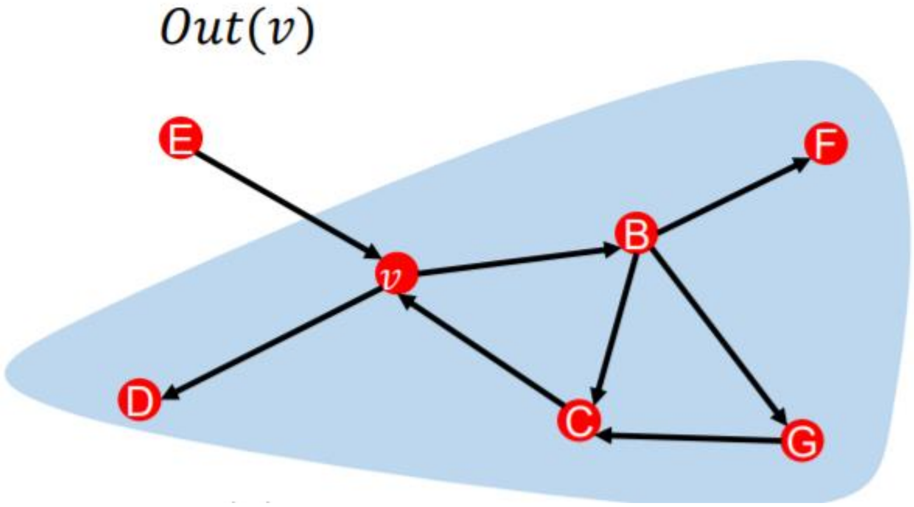
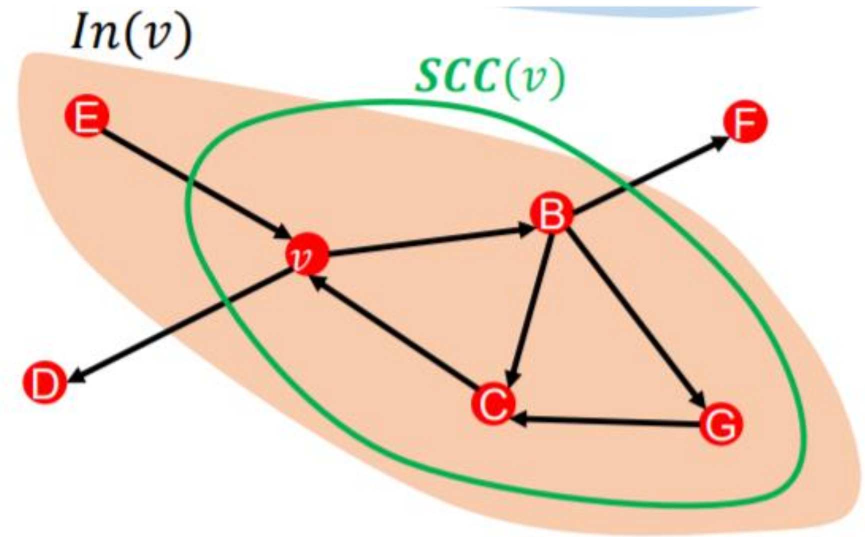
Summary Questions of the lecture

- Define the SCC in a graph and present a computation issue to find a SCC containing a node v.

→ An SCC in a directed graph is the largest possible set of nodes where every pair of nodes in the set can reach each other. The SCC containing a specific node v can be found by intersecting the set of nodes that can reach v and the set of nodes that v can reach.

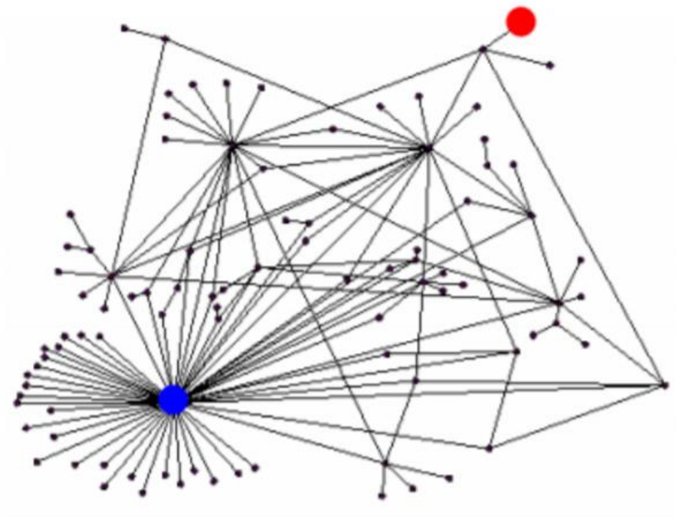




Summary Questions of the lecture

- Discuss the conceptual picture of the Web graph obtained from the result in page 13 of lecture note 14.

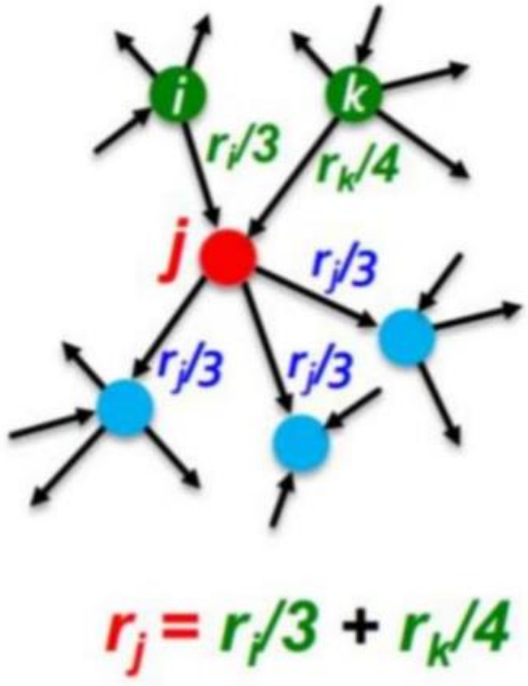
→ In the observation, a randomly chosen node either visits a lot of pages or very few nodes. Then, in average, around half of all pages can reach a randomly chosen page v and can be reached from the page v. Thus, we can say that all web pages are not equally "important".



Summary Questions of the lecture

- Explain the “Flow’ model for PageRank and discuss its validity on why it is worth.

→“Flow’ model is formulated to estimate the importance of a page by voting the incoming links, where links incoming from important pages count more. The importance of each page are equally divided by the number of its outgoing neighbors and propagated to the outgoing neighbors. Then the importance of each page is defined by the sum of all incoming importance from its incoming neighbors. The flow model effectively captures the intuition that links from important pages are worth more than links from those that are not.



Summary Questions of the lecture

- Present a random work interpretation of ‘power iteration’ method of eigenvector formulation for PageRank.

→ The ‘power iteration’ is a method to approximately calculate the rank vector $\boldsymbol{r}$ by repeatedly multiplying the stochastic adjacency matrix M to the previous rank vector $\boldsymbol{r}$. Since each iterarion is equivalent to a web surfer randomly taking an outgoing link in the current page at each time step, the ‘power iteration’ can be interpreted as a random walk having a transition probability of M and $\boldsymbol{r}$ is interpreted as a vector of the probability that the surfer stays at each page.

- Initialize: $\boldsymbol{r}^{(0)} = \left[\frac{1}{N}, \dots, \frac{1}{N}\right]^T$.
- Iterate: $\boldsymbol{r}^{(t+1)} = \boldsymbol{M}\boldsymbol{r}^{(t)}$
- Stop when $\left\|\boldsymbol{r}^{(t+1)} - \boldsymbol{r}^{(t)}\right\|_1 < \epsilon$

$$\begin{bmatrix} r_y \\ r_a \\ r_m \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1 \\ 0 & 1/2 & 0 \end{bmatrix} \begin{bmatrix} r_y \\ r_a \\ r_m \end{bmatrix}$$