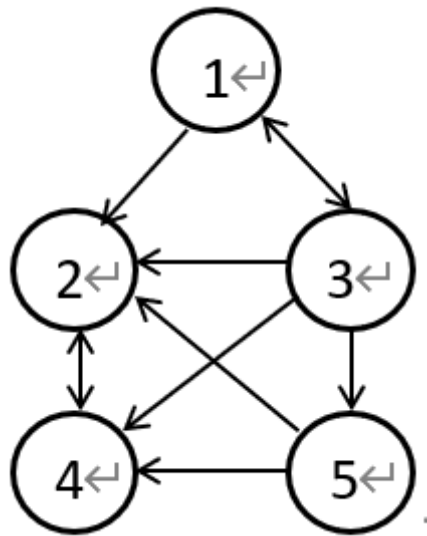


2022夏_第四次作业

4.1. 给出文档“no need to light a night light on a light night”的4-shingle集合

4.2. 采集器中，为何同时有URL去重和内容去重？试讨论二者关系及各自所起的作用。

4.3. 对于如下web图



a). 取 $\alpha=0.3$ ，求其转移概率矩阵。

b). 求按ppt35页的公式迭代两次之后的hub值向量和authority值向量。