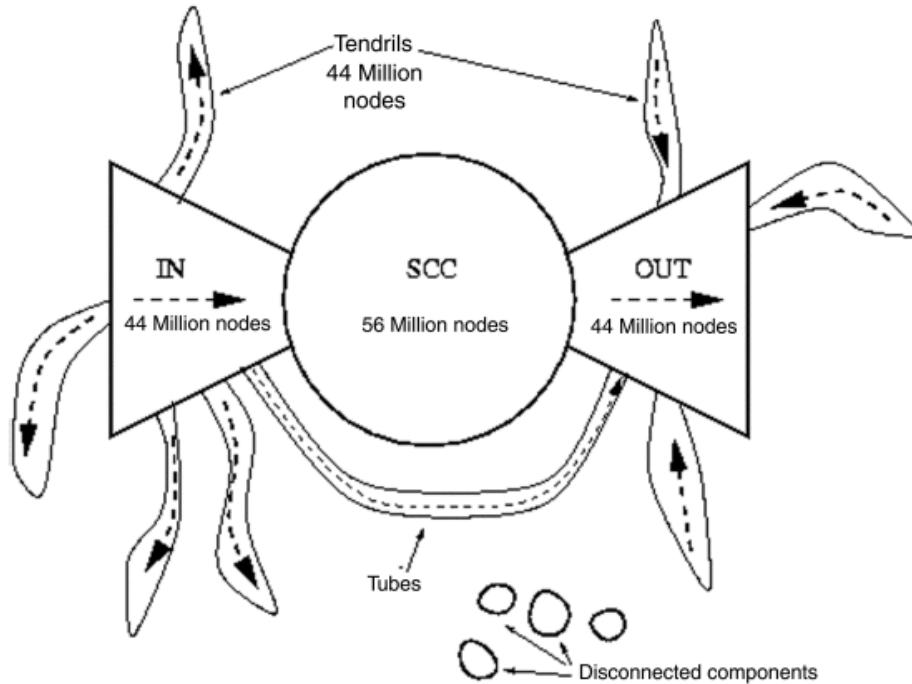


Web Structure

Intro to Web Science
10 points

Name _____

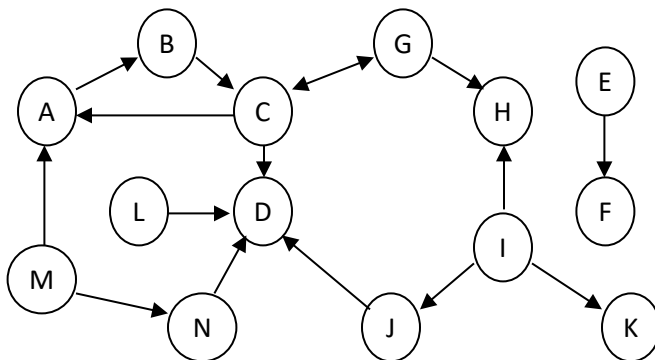
In 2000, Andrei Broder and other researchers from Alta Vista, IBM, and Compaq wrote a seminal paper called *Graph Structure in the Web*. From a large crawl of 200 million pages and 1.5 billion links, they showed the Web had a bow-tie structure:



Components:

1. **IN:** Pages with no in-links or in-links from IN pages and out-links to pages in IN, SCC, Tendrils, or OUT (via a Tube).
2. **SCC:** Pages with in-links from IN or SCC and out-links to OUT or SCC. There exists some path of links from every page in SCC to every other page in SCC.
3. **OUT:** Pages with no out-links or out-links to other pages in OUT, and all in-links come from OUT, SCC, Tendrils, or Tubes.
4. **Tendrils:** Pages that can only be reached from IN or have only out-links to OUT.
5. **Disconnected:** Pages that have no in-links from any other components and no out-links to other components. These pages may be linked to each other.

Suppose the following graph structure represented all of the pages and links in a new experiment. In which components would each of the pages be placed?



IN:
SCC:
OUT:
Tendrils:
Disconnected: