

### **C.5.3 Describe the main features of the web graph such as bowtie structure, strongly connected core (SCC), diameter.**

Students must be aware the web has a structure that has emerged from the behaviour of web users

Broder. Major study from 2002 (Numbers are now much higher)

SCC-Strongly Connected Core (56M)  
All pages can connect via directed links.

IN - can connect to but can't be connected by the SCC (eg new sites that have not been 'discovered').

OUT- can be connected by but can't connect to the SCC. (eg corporate site that has internal links which are not linked back to the SCC).

TENDRILS – can't access or be accessed by the SCC.

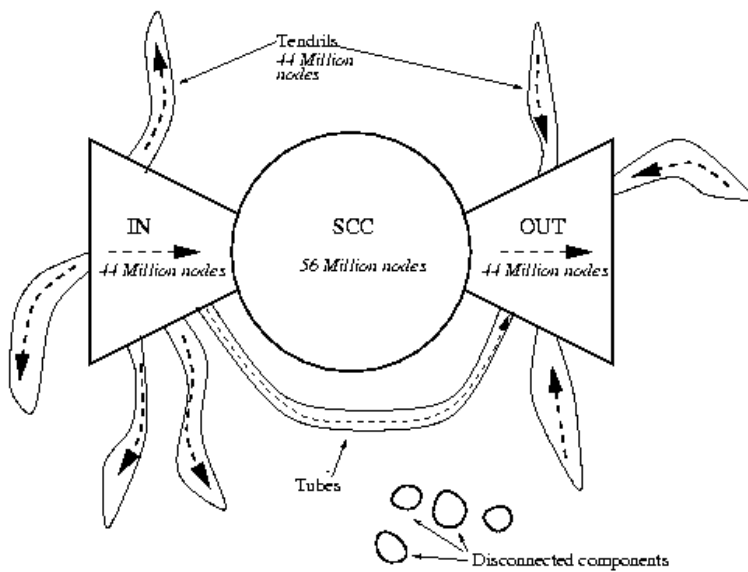
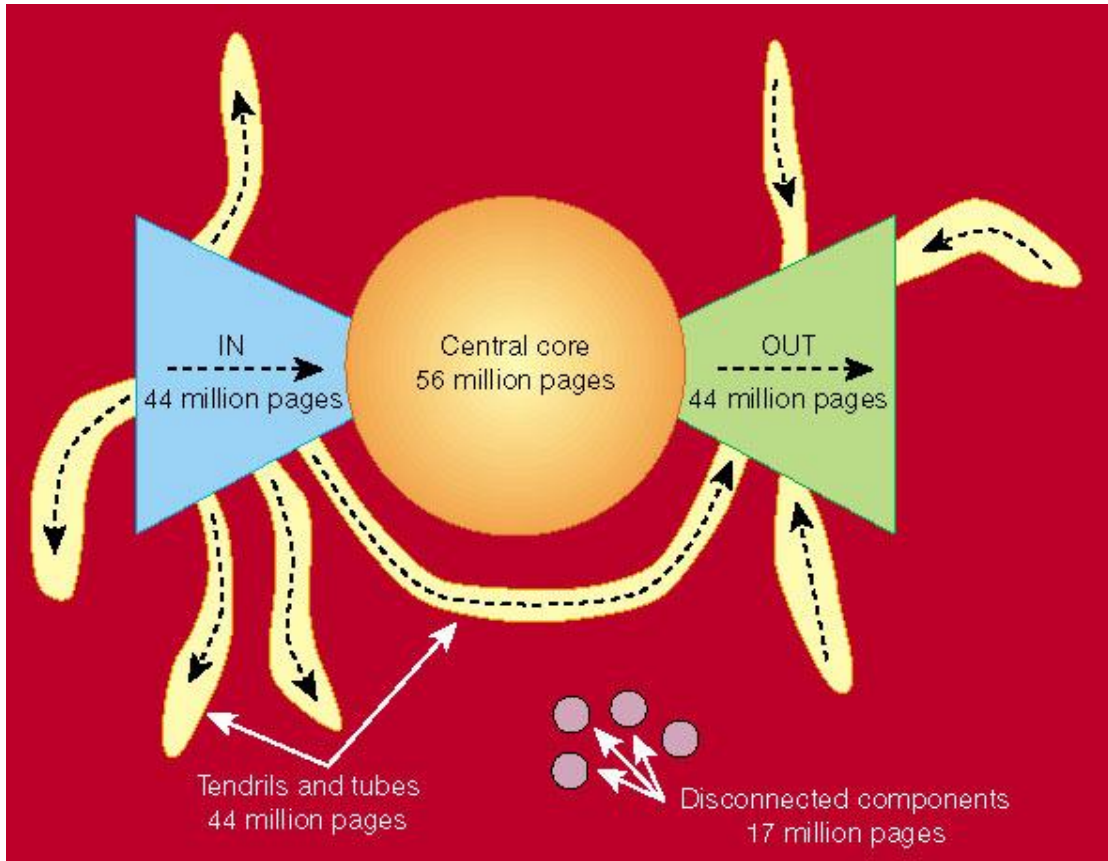
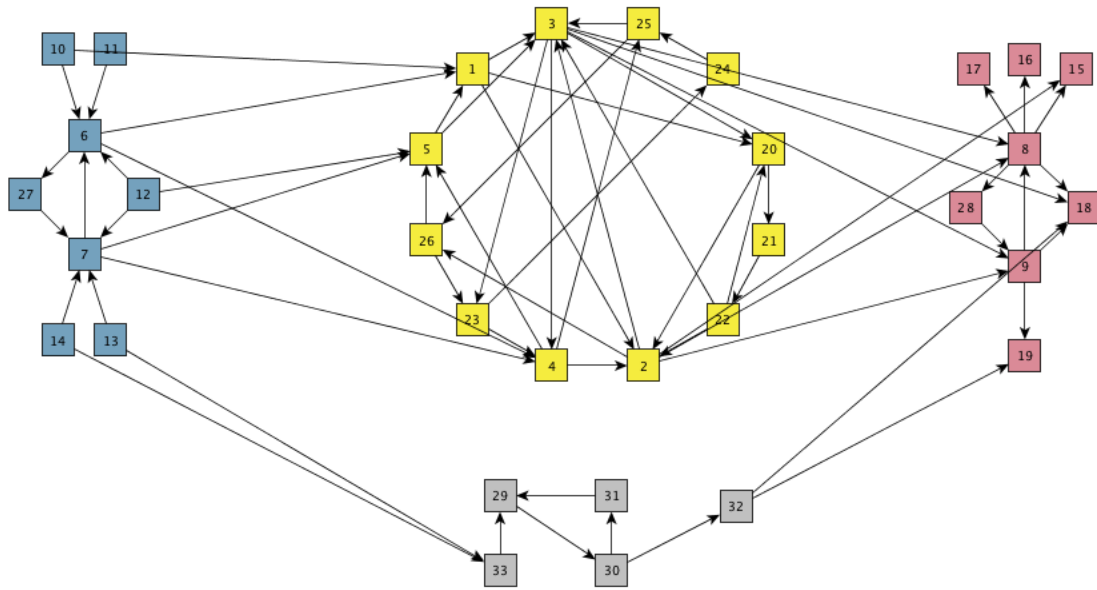


Figure 9: Connectivity of the web: one can pass from any node of IN through SCC to any node of OUT. Hanging off IN and OUT are TENDRILS containing nodes that are reachable from portions of IN, or that can reach portions of OUT, without passage through SCC. It is possible for a TENDRIL hanging off from IN to be hooked into a TENDRIL leading into OUT, forming a TUBE -- a passage from a portion of IN to a portion of OUT without touching SCC



<http://www.nature.com/nature/journal/v405/n6783/full/405113a0.html>

Further reading

[http://cs.wellesley.edu/~pmetaxas/Why\\_Is\\_the\\_Shape\\_of\\_the\\_Web\\_a\\_Bowtie.pdf](http://cs.wellesley.edu/~pmetaxas/Why_Is_the_Shape_of_the_Web_a_Bowtie.pdf)

<http://www9.org/w9cdrom/160/160.html>