Crawler:

http://www.cis.uni-muenchen.de/~yeong/Kurse/ss09/WebDataMining/kap8_rev.pdf

Link Analysis:

http://cse.iitkgp.ac.in/~animeshm/analysis.pdf (Slides with heading Eigenvector centrality, Katz centrality, PageRank, Interpreting Web Surfing, Transition Matrix, Random Walk, Steady State Calculations, Hubs and Authorities, Co-citation Index, Bibliographic Coupling) + a nice simulation of PageRank here.

Trawling the web for cyber-communities:

Section 2.2 and 4.1 of:
http://www.facweb.iitkgp.ernet.in/~niloy/COURSE/Spring2006/CNT/Resource/vertex-copying-2.pdf

1. Note that as I already told you in class, the authors could successfully carry out enumerations of $C_{i,j}$ with i=3 to 9 and j=3 to 20.
2. One more thing which probably needs a little more clarification: the in-degree of node u need to be exactly 4 for u to be a part of $C_{4,4}$. Note that this is the strictest possible case. If the in-degree of u is 5 or more, while the neighborhood intersection of the nodes pointing to u is just 4, then also u can qualify as a part of $C_{4,4}$ given a $K_{4,4}$ exists.