



SNS COLLEGE OF ENGINEERING

Kurumbapalayam (Po), Coimbatore – 641 107



An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A' Grade
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

DEPARTMENT OF COMPUTER SCIENCE AND TECHNOLOGY

COURSE NAME : 19CS731- Social Network Analysis

III YEAR /V SEMESTER

Unit 1- Introduction

Topic 8: WEB BASED NETWORKS



1. There are two features of web pages that are considered as the basis of extracting social relations: links and co-occurrences.
2. The linking structure of the Web is considered as proxy for real world relationships as links are chosen by the author of the page and connect to other information sources that are considered authoritative and relevant enough to be mentioned.
3. The biggest drawback of this approach is that such direct links between personal pages are very sparse: due to the increasing size of the Web searching has taken over browsing as the primary mode of navigation on the Web.



1. Co-occurrences of names in web pages can also be taken as evidence of relationships and are a more frequent phenomenon
2. On the other hand, extracting relationships based on co-occurrence of the names of individuals or institutions requires web mining as names are typically embedded in the natural text of web page.
3. Web mining is the application of text mining to the content of web pages. The techniques employed here are statistical methods possibly combined with an analysis of the contents of web pages



1. Note that this corresponds to a very shallow parsing of the web page as indirect references are not counted this way (e.g. the term “the president of the United States” will not be associated with George Bush even if he was mentioned as the president elsewhere in the text.)

2. A disadvantage of the Jaccard -coefficient is that it penalizes ties between an individual.

whose name often occurs on the Web and less popular individuals.

3. For this reason we use an asymmetric variant of the coefficient. In particular, we divide the number of pages for the individual with the number of pages for both names and take it as evidence of a directed tie if this number reaches a certain threshold.



REFERENCES

1. Dion Goh and Schubert Foo, Social information retrieval systems :Emerging technologies and applications for searching the web effectively, IGI Global snippet, 2008

THANK YOU