Mining Temporally Evolving Graphs

Prasanna Desikan and Jaideep Srivastava
Department of Computer Science
University of Minnesota, Minneapolis, MN 55414, U.S.A
{desikan,srivastava}@cs.umn.edu

Abstract

Web mining has been explored to a vast degree and different techniques have been proposed for a variety of applications that include Web Search, Web Classification, Web Personalization etc. Most research on Web mining has been from a 'data-centric' point of view. The focus has been primarily on developing measures and applications based on data collected from content, structure and usage of Web till a particular time instance. In this project we examine another dimension of Web Mining, namely temporal dimension. Web data has been evolving over time, reflecting the ongoing trends. These changes in data in the temporal dimension reveal new kind of information. This information has not captured the attention of the Web mining research community to a large extent. In this paper, we highlight the significance of studying the evolving nature of the Web graphs. We have classified the approach to such problems at three levels of analysis: *single node*, *sub-graphs* and *whole graphs*. We provide a framework to approach problems of this kind and have identified interesting problems at each level. Our experiments verify the significance of such analysis and also point to future directions in this area. The approach we take is generic and can be applied to other domains, where data can be modeled as graph, such as network intrusion detection or social networks.

Key Words: Web Mining, Evolving Graphs, Temporal Behavior

1 Introduction

Web Mining, defined as the application of data mining techniques to extract information from the World Wide Web, has been classified into three sub-fields: Web Content Mining, Web Structure Mining and Web Usage Mining based on the kind of the data. The evolving structure of interlinked documents, such as the World Wide Web, and their usage over a period of time has evoked new interest to both researchers and industry. These set of documents form a graph, with nodes representing documents and edges representing hyperlinks. Extracting information from the pure structure of such graphs and the usage of these documents, especially with respect to the World Wide Web, has been extensively studied [24]. The significance of the Web graph structure is evident from the success of Google, whose PageRank technology is based on the hyperlink structure of the Web. A survey on Hyperlink Analysis is provided by Desikan et al [7]. Usage aspects of graphs such as the Web and user navigation pat-terns have also received wide attention [20, 23].

Most research has thus focused more recently on mining information from structure and usage of such graphs at a given time instance. This paper focuses on another important dimension of Web Mining - the Temporal Evolution of the Web; as identified by our earlier work [8, 24]. The Web is changing fast over time and so is the user's interaction in the Web suggesting the need to study and develop models for the evolving Web Content, Web Structure and Web Usage. Changes in Web data occur either rapidly or slowly. In more general graphs these changes may also occurs suddenly representing anomalous behavior. A typical example of such graphs would be network flow connections. We have also been applying the techniques in this paper on that domain; where sudden changes may characterize network intrusions. The content of certain web sites, such as news channels, change very dynamically due to new and updated information that comes in frequently. On the contrary, there are other web sites, such as encyclopedias, that are very informative, but whose content does not change very rapidly. Similarly, the change in the web structure is also dependent on the web site in particular. Some web sites may change their structure more rapidly to address the user navigation issues. The structure of the whole Web graph has been found relatively stable [15]. However, analyzing sub-structures that correspond to web sites or web communities will be interesting. It is observed that Web usage graphs, which are constructed from web server logs, will tend to vary more with time; as they reflect the user behavior and needs on daily basis. Due to the nature of different rate of changes in data, the properties that need to be measured and the techniques that need to be developed to capture the changing behavior will need to differ. The techniques developed will also depend on the scope of analysis. The temporal behavior of the Web graph can be analyzed at three levels:

- **Single Node:** Studying the behavior of a single node across different time periods. Node properties, such as total hits, hub and authority scores, indegree and outdegree, may vary across time.
- **Sub-graphs:** Set of nodes that form sub-graph structures that occur in the Web graph. The frequency of their occurrence across time periods and also the sequences of sub-graphs would require different techniques.
- **Whole graph:** The variation of the basic and derived graph properties such as order, size, number of components, maximum indegree, maximum outdegree, average hub score, average authority score; will help in profiling of the behavior of graph.

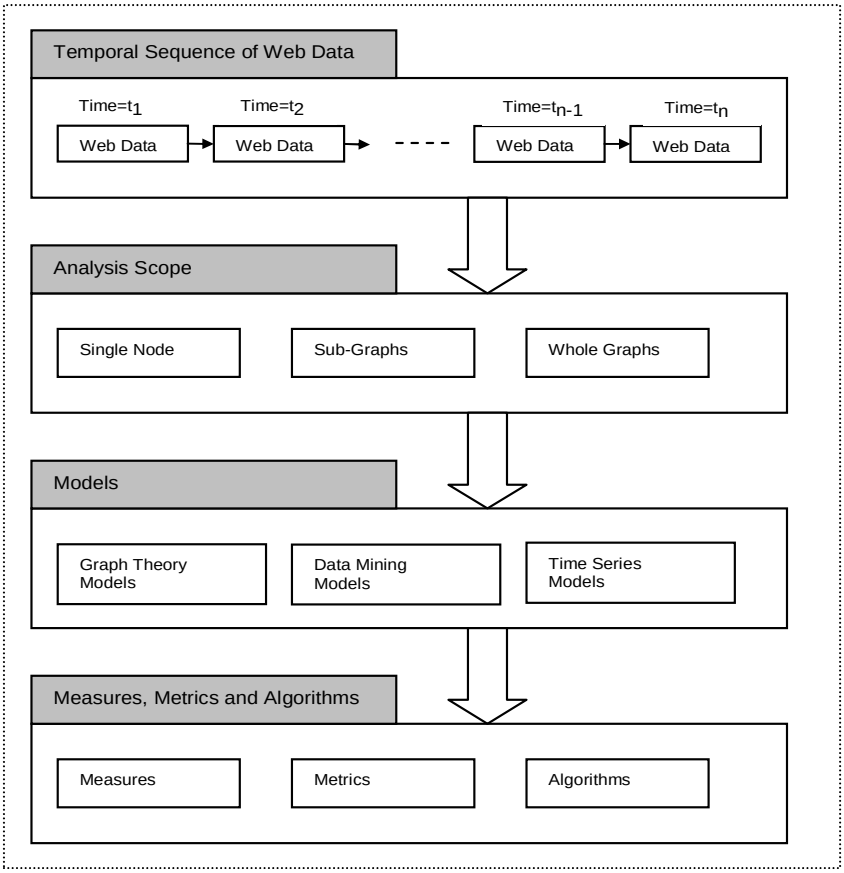


Figure1: Overall Approach for Temporal Mining of Web Graphs

Figure 1 describes an overall approach that needs to be taken for mining temporal behavior groups. The need to study the Temporal Evolution of the Web, understand the change in the user behavior and interaction in the World Wide Web has motivated us to analyze the Web Usage data. We use the user logs obtained from the Web server to construct Web usage graphs and study their evolution over time.

The rest of this document is organized as follows: In Section 2 we talk about related work in this area and in the following section, we put forward various kinds of analysis that need to be carried out. We identify key issues and provide framework to address such problems. Section 4 discusses our experiments performed and results. In section 5 we summarize our approach and provide pointers to future work.

2 Related Work

Recent works in Web mining has focused on Web usage and Web structure data. Web usage data has captured attention due to its nature of bringing user's perspective of the Web as opposed to creator's perspective. Understanding user profiles and user navigation patterns for better adaptive web sites and predicting user access patterns has evoked interest to the research and the business community. Methods for pre-processing the user log data and to separate web page references into those made for navigational purposes and those made for content purposes have been developed [6]. User navigation patterns have evoked much interest and have been studied by

various other researchers [2, 5, 13, 20]. Srivastava et al [23] discuss the techniques to pre-process the usage and content data, discover patterns from them and filter out the non-relevant and uninteresting patterns discovered.

Usage statistics has been applied to hyperlink structure for better link prediction in field of adaptive web sites. The concept of adaptive web sites was proposed by Perkowitz and Etzioni [19]. Since then, Markov models have been used extensively to predict user behavior [19, 21, 22, 25]. Information foraging theory concepts have also been successfully applied to the Web domain by Chi et al [4] to incorporate user behavior into the existing content and link structure. Clustering of user sessions and navigation paths with different underlying models is used to capture similarity in the user behavior [3, 11].

Research in Web structure mining has focused primarily on hyperlink analysis and has found its utility in a variety of applications [7]. Determining the quality of a page has been the primary focus and various measures and metrics have been developed for the same. PageRank[18] is a metric, developed by Google founders, for ranking hypertext documents. The key idea is that a page has high rank if it is pointed to by many highly ranked pages. So the rank of a page depends upon the ranks of the pages pointing to it. Intuitively, the approach can be viewed as a stochastic analysis of a random walk on the Web graph. It measures the probability that a person arrives at a page either by traversing a link or by other means such as typing a URL or following bookmarks.

The other popular metric is hub and authority scores. From a graph theoretic point of view, hubs and authorities can be interpreted as 'fans' and 'centers' in a bipartite core of a Web graph. The hub and authority scores computed for each Web page indicate the extent to which the Web page serves as a 'hub' pointing to good 'authority' pages or the extent to which the Web page serves as an 'authority' on a topic pointed to by good hubs. The hub and authority scores for a page are not based on a formula for a single page, but are computed for a set of pages related to a topic using an iterative procedure, namely HITS algorithm [14].

Oztekin et al [17], proposed Usage Aware PageRank. Their modified PageRank metric incorporates usage information. Weights are assigned to a link based on number of traversals on that link, and thus modifying the probability that a user traverses a particular link. Also the probability to arrive at a page directly is computed using the usage statistics. The Internet Archive [26] is one of the key data sources for studying the change in the web structure and content of different web sites. There are also tools available such as the AT&T Internet Difference Engine that detects the change in the existing html documents. However, these tools do not capture the temporal nature of changing documents.

3 Analysis of temporal behavior of graphs

The dynamic nature of Web data has aroused interest to mine temporal patterns of the Web data. The study of change in behavior of Web content, Web structure and Web usage over time and their effects on each other would help in understanding better, the way Web is evolving and necessary steps that can be taken to make it a better source of information. The time dimension brings into perspective concepts such as 'recently popular' versus 'historically significant'. As discussed in Section 1, the analysis of the temporal behavior can be classified at three levels that are discussed below in detail.

3.1 Single Node Analysis

The behavior of Web data can be modeled at the level of a single node. For each single node that is labeled, properties based on its content, structure and usage can be computed. Over a period of time, content of a page (represented as a node) can change, signifying the change in the topic addressed by the Web page. Also, structural behavior of a node, over a time period, can be analyzed by studying the variation of properties that are based on link structure, such as indegree, outdegree, authority score, hub score or PageRank score. Such kind of behavior will also serve as a very useful feedback to the web site designer. Finally, study of usage data of a single node across a time period, will reflect the popularity of a node during a given time period. The temporal dimension will help to reflect current trends and help in prediction of popular topics. A vector of a property, for set of nodes, as a function of time, is described in Equation 1:

$$SN(t)_{n,p} = [s_1(t) \quad s_2(t) \quad s_3(t)..... \quad s_n(t)] \qquad (1)$$

Where $s_i(t)$ represents the score of the i^{th} node at time t; n is the number of nodes in the graph and $SN(t)_{n,p}$ represents the vector of scores of a particular property, p, of each node at a given time instance t. Here, n is the number of nodes (assuming each node has a unique label that it retains across the time period) for the graph built

over the time period. If a node doesn't exist at any time instance it would be assigned a score of zero for that particular time instance. This is a naïve model at this stage, but captures most significant properties we are interested in. These models could however be extended.

3.2 Sub-graph Analysis

At the next hierarchical level, changing sub-graph patterns evoke interest. These sub-graphs may represent different communities and abstract concepts that evolve over time. The idea of mining frequent sub-graphs has been applied with a large graph as an input or a set of small graphs as input [16]. However, with addition of a temporal dimension, we look at an evolving graph, which may have different set of sub-graphs at different time instances. Figure 2 illustrates an example of an evolving graph, and the sequential patterns that can be mined. In the example, it is seen that if a subgraph pattern, g1, occurs during a time interval, the probability that a subgraph, g2, will occur in the next time period is higher than any other sequence of subgraphs over adjacent time periods.

It can be seen that mining of sequential patterns of sub-graphs might provide useful information in profiling the change behavior. Sequence mining may also help in predicting an upcoming trend or predict an abnormal behavior in the Web structure. These changes may occur due to change in the strategy of an organization or a launch of new product by an e-commerce site.

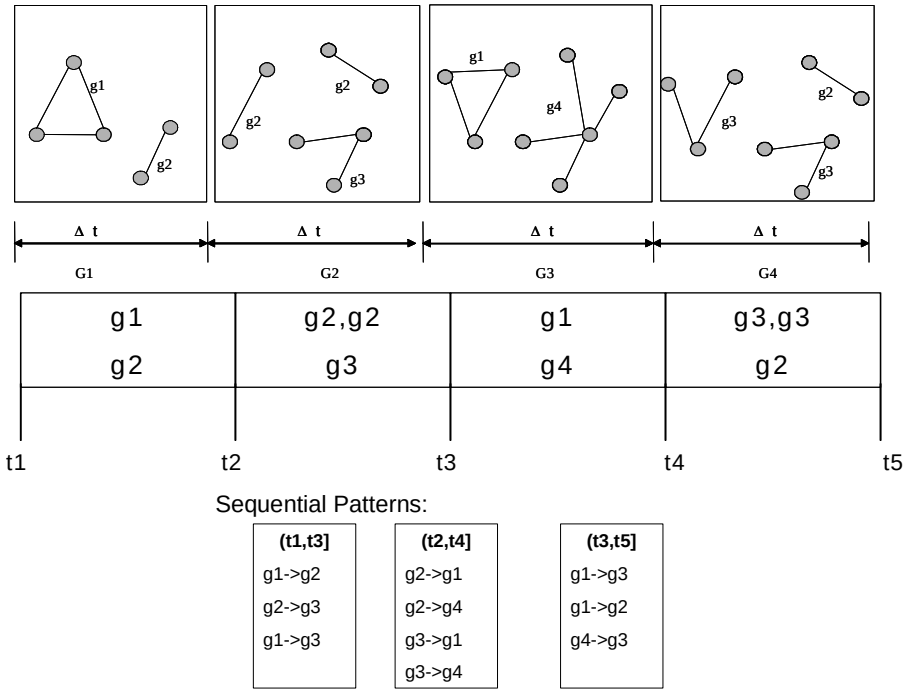


Fig.2. Sequential Pattern Mining of Subgraphs.

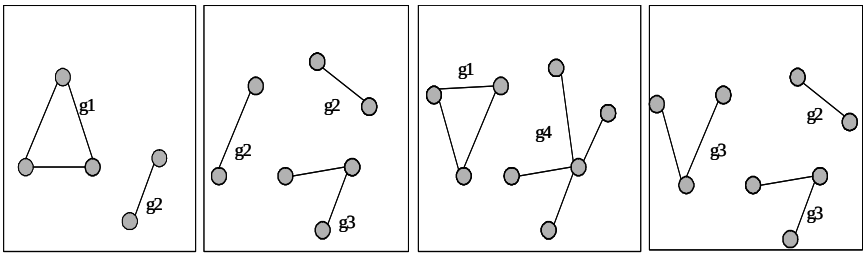
While such changes provide interesting conceptual information, mining such patterns poses interesting challenges to the research community to develop efficient algorithms, due to size of the graphs and the number of such graphs that are needed to model the temporal behavior. Formally the set of such sequences can be captured using Equation 2.

$$G_{t,t+\Delta t}(\Delta t) = \left\{ g_i \rightarrow g_j \middle| g_i \in FSG(G_t); g_j \in FSG(G_{t+\Delta t}) \right\} \quad (1)$$

where $G_{t,t+\Delta t}(\Delta t)$ represents the set of sequences of subgraphs in a Δt time period; g_j represents a subgraph and $FSG(G_t)$ represents all the frequent subgraphs of the graph, G_t, where G_t represents the graph at a time instance t. Equation 2 provides the basis of information to mine the interesting subgraph sequences.

3.3 Whole Graph Analysis

While analysis of single nodes and sub-graphs tend to give specific information, analysis at the level of a whole graph will reveal higher level concepts. For each graph at a given time instance, a vector of feature space consisting of basic properties and derived properties can be built. Choosing the appropriate components of such a vector and its variation in the time dimension is an interesting area of research. Figure 3 illustrates the concept of the graph evolving and how the different graph properties change with time. Modeling such a vector space and analyzing it over a time period poses interesting challenges.



G 1	G 2	G 3	G 4
|v|=n(G 1)=5	|v|=n(G 2)=7	|v|=n(G 3)=8	|v|=n(G 4)=8
|E| =e(G)=4	|E| =e(G 2)=4	|E| =e(G)=7	|E| =e(G 4)=5
# of Cpts= 2	# of Cpts= 3	# of Cpts= 2	# of Cpts= 3

Fig. 3. Modeling features of graphs that evolve over a time period. (Note: the set of properties shown is just an illustration, it is not exhaustive)

The research directions involve in identifying key properties that model behavior of change of Web structure; and any correlation between such properties. The feature space can be divided into two different sets of properties for a whole graph:

- **Basic Properties:** These are basic properties of any graph such as order, size, density, number of components. These can measured using basic graph algorithms.
- **Derived Properties**: Properties of graphs such as average or maximum hub score, average or maximum authority score.

Other interesting issues might be how the distribution of such scores varies over a period of time. Such variation in distributions may indicate the variation of kind of graph such as scale-free graphs versus random graphs. The vector space of features as a function of time is formulated in Equation 3.

$$GF(t) = [BF_1(t) \quad BF_2(t) \quad ... \quad BF_m(t) \quad DF_1(t) \quad DF_2(t) \quad ... \quad DF_n(t)] \qquad (2)$$

where $GF(t)$ represents the feature space of basic features $(BF_i(t))$ and derived features $(DF_i(t))$ at a time instance t. Analyzing this feature space over a time period will reveal information about the temporal behavior of whole graph.

4 Experiments and Results

The data used in our experimentation is the Web data obtained from the Computer Science website at the University. The structure was obtained from a crawl and web server logs for the corresponding period were collected for the

first part of our experimentation. We used this data as Web server logs are available for the website and it makes a fair ground for comparison of the web structure based metrics versus the web usage based metrics. Our experiments were performed at two different levels of analysis – single node level and the whole graph level. Mining sequential sub-graph patterns are the focus of our present research, since those require much storage and efficient algorithms. Our present analysis was focused on change in Web usage data, as the change in such usage graphs is more evident over a short period of time.

4.1 Single Node Analysis of Web Usage Graph

The data considered was from April through June 2002. We clustered the Web pages access patterns over the three month period; with the granularity of day to count the number of accesses at time instance. Clustering results are shown in Figure 4(a). Some interesting clusters were observed. Sample results from these clusters are presented in Figure 4(b). One of the clusters, labeled 1 in Figure 4, belongs to the set of pages that were accessed a lot during a very short period of time. Most of them are some kind of wedding photos that were accessed a lot, suggesting some kind of a 'wedding' event that took place during that time. Similar behavior was observed in another cluster, labeled 2 in Figure 4, and the set of pages belonged to talk slides of some event that took place during that period. The third cluster, labeled 3 in Figure 4 was the most interesting. It had mostly pages related to 'Data Mining' notes. These set of pages had high access during the first period of time, possibly the spring term and then their access died out; indicating the end of semester. Interestingly enough, there was no Data Mining course offered during that term suggesting someone was studying 'data mining' during that semester, as a part of another course or due to other interests. Our observations suggest that clustering Web page access patterns over time help in identifying a 'concept' that is 'popular' during a time period.

We also defined a naïve metric Page Usage Popularity gives more weight to 'recent' history. The metric was computed by simply weighing the number of hits to a page in the last one-third of the time period more than the first two-thirds of the time period. The results are presented in Appendix A. It was observed that PageRanks (Figure 5(a)) tend to give more importance to structure; hence pages that are heavily linked may be ranked higher though not used. Usage Aware PageRank (Figure 5(b)) combines usage statistics (cumulative for the time period considered) with link information giving importance to both the creator and the actual user of a web page. Page Usage Popularity (Figure 5(c)) helps in ranking 'obsolete' items lower and boosting up the topics that are more 'popular' during that time period. Thus it signifies the importance of bringing the temporal dimension as opposed to analysis of a static graph.

4.2 Whole Graph Analysis of Web Usage Graph

Our next set of experiments involved analyzing the basic and derived features of Web usage graphs over a three month period. We have plotted these features against time period granularity of single day. Our initial results suggested some seasonal patterns that were not observed for the new dataset. From these plots shown in Appendix B, there were some interesting observations. The first set of plots, in Figure 6(a), reflects variation of basic properties of the graph, such as order, size and the number of components. While order and size reveal the number of pages traversed and links traversed, the number of components will reflect how connected the graph is. We eliminated robot behavior collected in web server logs. Any robot like behavior would be decrease the number of components sharply. In the network log data, the drop in number of components also helped us detect machines sending spam mails to other machines in the network. Those results are not presented due to lack of space and relevance to this context. The second set of plots, in Figure 6(b), reflects behavior of de-rived properties with time. Due to the normalization of hub, authority and PageRank scores; we can just judge that Web pages that were most relevant for that given time instance. The variation of Indegree and Outdegree scores reflect number of connections as opposed to the pure relevance score. It is seen that that the maximum outdegree score of the graph seems to have an increasing trend.

These results suggest that the variation of different properties of these graphs need to be studied. The difference in the variation of related metrics, such as hub scores versus outdegree scores; suggest that the distribution of these scores change and studying the change in the distribution of the scores would be interesting. It would reflect the nature of evolving graphs.

5 Conclusions

We have presented in this paper the significance of introducing the dimension of time into Web mining research as identified by some of our earlier works. We have classified the scope of analysis into three levels namely; single node analysis, sub-graph analysis and whole graph analysis. At a single node level analysis, we observed, the usage based ranking metrics, boost up the ranks of the pages that are used as opposed to the pure hyperlink based metrics that rank pages that are used rarely high. In particular, we notice that time based metrics , such as Page Usage Popularity, can be effectively used to boost ranks of recently popular pages to those that are more obsolete. At the level of the whole graph, we notice the trend and seasonal components in the behavior of the basic properties. The issues that need to be addressed at each level are discussed in detail.

The two primary tasks for research are to identify the appropriate set of properties that define the changing behavior at each level and to develop efficient algorithms that scale to the large data sets. We have also applied these generic techniques to other graph data such as Netflow connections, to help profile normal and anomalous behavior. We thus see a clear need to pursue research in the area of evolving graphs which hold significance in application domains like Web, Network Intrusion or Social Networks. Another engineering issue would be developing a grammar for querying interesting graph patterns.

Acknowledgements

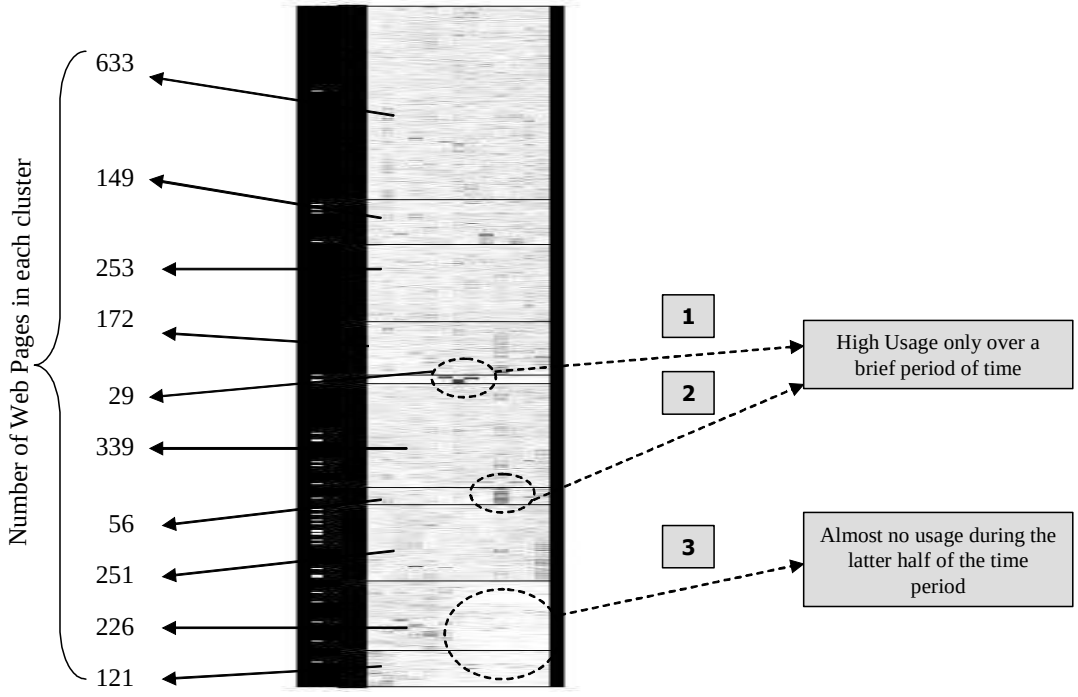
We would like to thank Prof. Vipin Kumar and Data Mining Research group at the Department of Computer Science for providing valuable suggestions. Uygar Oztekin provided the ranking results using PageRank and Usage Aware PageRank metrics. This work was partially supported by Army High Performance Computing Research Center contract number DAAD19-01-2-0014. The content of the work does not necessarily reflect the position or policy of the government and no official endorsement should be inferred. Access to computing facilities was provided by the AHPCRC and the Minnesota Supercomputing Institute.

References

1. S.Acharyya and J.Ghosh, “A Maximum Entropy Framework for Link Analysis on Directed Graphs”, in LinkKDD2003, pp 3-13, Washington DC, USA, 2003
2. A. Buchner, M. Baumagarten, S. Anand, M.Mulvenna, and J.Hughes. Navigation pattern discovery from internet data. In Proc. of WEBKDD’99 , Workshop on Web Usage Analysis and User Profiling, Aug 1999.
3. I Cadez, D. Heckerman, C. Meek, P. Smyth, S. White, “’Visualization of Navigation Patterns on a Web Site Using Model Based Clustering”, Proceedings of the KDD 2000
4. E.H. Chi, P. Pirolli, K. Chen, J. Pitkow. Using Information Scent to Model User Information Needs and Actions on the Web. In Proc. of ACM CHI 2001 Conference on Human Factors in Computing Systems, pp. 490--497. ACM Press, April 2001. Seattle, WA.
5. M. S. Chen, J.S. Park, and P.S. Yu. Data Mining for path traversal patterns in a web environment. In 16th International Conference on Distributed Computing Systems, 1996.
6. R. Cooley, B. Mobasher, and J.Srivastava. “Data Preparation for mining world wide web browsing patterns”. Knowledge and Information systems, 1(!) 1999.
7. P. Desikan, J. Srivastava, V. Kumar, P.-N. Tan, “Hyperlink Analysis – Techniques & Applications”, Army High Performance Computing Center Technical Report, 2002.
8. P. Desikan, J. Srivastava, “Temporal Behavior of Web Usage”, AHPCRC technical report, August 2003
9. C. Ding, H. Zha, X. He, P. Husbands and H.D. Simon, “Link Analysis: Hubs and Authorities on the World Wide Web” May 2001. LBNL Tech Report 47847
10. F. Douglis, T. Ball, Y-F. Chen, E. Koutsofios, “The AT&T Internet Difference Engine: Tracking and Viewing Changes on the Web”, World Wide Web, January 1998, pp. 27-44.
11. O. Etzioni, “The World Wide Web: Quagmire or Gold Mine”, in Communications of the ACM, 39(11):65=68,1996
12. J.Z Huang, M. Ng, W.K Ching, J. Ng, and D. Cheung, “A Cube model and cluster analysis for Web Access Sessions”, In Proc. of WEBKDD’01, CA, USA, August 2001.
13. X.Jin, Y.Zhou and B. Mobasher, “Web Usage Mining Based on Probabilistic Latent Semantic Analysis”. In Proceedings of KDD'04, Seattle, August 2004
14. J.M. Kleinberg, “Authoritative Sources in Hyperlinked Environment”, 9th Annual ACM-SIAM Symposium on Discrete Algorithms, pages 668-667, 1998
15. J. Klienberg et al. “The web as a graph: measurement models & methods”.Proc ICCC 1999.

16. M. Kuramochi and G. Karypis “Finding Frequent Patterns in a Large Sparse Graphs”, SIAM Data Mining Conference, 2004
17. B.U. Oztekin, L. Ertoz and V. Kumar, “Usage Aware PageRank”, World Wide Web Conference, 2003.
18. L. Page, S. Brin, R. Motwani and T. Winograd “The PageRank Citation Ranking: Bringing Order to the Web” Stanford Digital Library Technologies, January 1998.
19. M. Perkowitz and O. Etzioni, “Adaptive Web sites: an AI challenge”. IJCAI97
20. D. Pierrakos, G. Paliouras, C. Papatheodorou, and C. Spyropoulos, ‘Web usage mining as a tool for personalization: A survey’. User Modeling and User-Adapted Interaction, 2003.
21. P. Pirolli, J. E. Pitkow, “Distribution of Surfer’s Path Through the World Wide Web: Empirical Characterization.” World Wide Web 1:1-17, 1999.
22. R.R. Sarukkai, “Link Prediction and Path Analysis using Markov Chains”, In the Proc. of the 9th World Wide Web Conference , 1999.
23. J. Srivastava, R. Cooley, M. Deshpande and P-N. Tan. “Web Usage Mining: Discovery and Applications of usage patterns from Web Data”, SIGKDD Explorations, 2000.
24. J. Srivastava, P. Desikan and V. Kumar, “Web Mining – Concepts, Applications and Research Directions”, NGDM, MIT/AAAI Press (to be released in 2004)
25. J. Zhu, J. Hong, and J.G. Hughes, Using Markov Chains for Link Prediction in Adaptive Web Sites. In Proc. of ACM SIGWEB Hypertext 2002.
26. Internet Archive, http://www.archive.org/

Appendix A



(a)

Wedding Pictures from 'ctlu' that were accessed high for few days

1

www-users.cs.umn.edu/~user1/Wedding/speech.html
www-users.cs.umn.edu/~user2/user3/re0.html
www-users.cs.umn.edu/~user2/user3/wap.html
www-users.cs.umn.edu/~user1/Wedding/photo2.html
www-users.cs.umn.edu/~user1/Wedding/wedding.html

Webpages of a talk – accessed high for few days in and around the talk date

2

www.cs.umn.edu/event/spin/talks/presenter-jan-00/sld016.htm
www.cs.umn.edu/crisys/spin/talks/presenter-jan-00/sld017.htm
www.cs.umn.edu/Research/airvl/its/
www.cs.umn.edu/event/spin/talks/presenter-jan-00/sld055.htm
www.cs.umn.edu/event/spin/talks/presenter-jan-00/sld056.htm

Class Web pages that were not frequently accessed after the end of semester

3

www-users.cs.umn.edu/~user4/hpdmtut/sld110.htm (Data Mining Tutorial slides)
www-users.cs.umn.edu/~user4/hpdmtut/sld113.htm
www-users.cs.umn.edu/~user4/hpdmtut/sld032.htm
.
www.cs.umn.edu/classes/Spring-2002/csci5707/
www.cs.umn.edu/classes/Spring2002/csci5103/www_files/contents.htm

(b)

Figure 4 : (a)Web pages clustered according to their usage over time.
(b) Sample results of interesting clusters.

1. 0.01257895 www.cs.umn.edu/cisco/data/doc/cintrnet/ita.htm
2. 0.01246077 www.cs.umn.edu/cisco/home/home.htm
3. 0.01062634 www.cs.umn.edu/cisco/data/lib/copyrght.htm
4. 0.01062634 www.cs.umn.edu/cisco/data/doc/help.htm
5. 0.01062634 www.cs.umn.edu/cisco/home/search.htm
8. 0.00207191 www.cs.umn.edu/help/software/java/JGL-2.0.2-for-JDK-1.1/doc/api/packages.html
9. 0.00191835 www.cs.umn.edu/help/software/java/JGL-2.0.2-for-JDK-1.0/doc/api/packages.html
10. 0.00176296 www.cs.umn.edu/help/software/java/JGL-2.0.2-for-JDK-

e.g These Web pages do not figure in usage based rankings

(a)

1.	0.000723	www.cs.umn.edu/
2.	0.000145	www-users.cs.umn.edu/~mein/blender/
3.	0.000081	www-users.cs.umn.edu/~oper/linuc/allen/
4.	0.000053	www-users.cs.umn.edu/~xyz/hpdmtut/
5.	0.000051	www-users.cs.umn.edu/~1234/wiihist/
6.	0.000050	www-users.cs.umn.edu/grad.html
7.	0.000050	www.cs.umn.edu/faculty/faculty.html
8.	0.000046	www-users.cs.umn.edu/~oper/allen/200393/
9.	0.000044	www-users.cs.umn.edu/~self/links.html
10.	0.000042	www.cs.umn.edu/courses.html

(b)

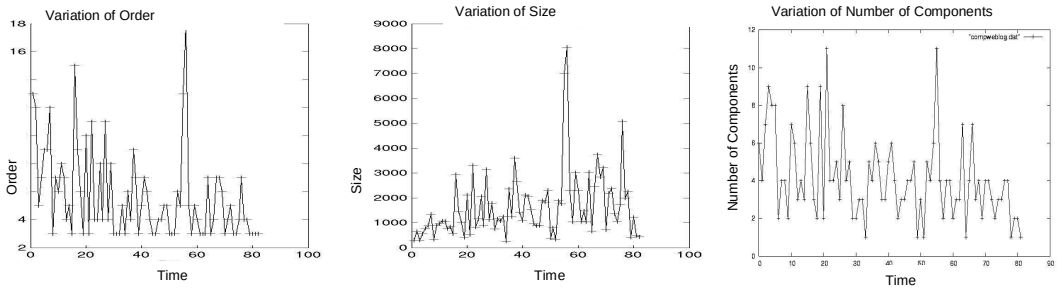
Low–Ranked because course web-pages were not accessed in the month of June, which was considered recent and hence more weight given to pages accessed during that period.

1. 0.01116774 www.cs.umn.edu/
2. 0.00099434 www-users.cs.umn.edu/~abc/blender/
3. 0.00067178 www-users.cs.umn.edu/~xyz/popsoft/
4. 0.00066555 www.cs.umn.edu/Research/GroupLens/
5. 0.00065136 www-users.cs.umn.edu/
6. 0.00064202 www-users.cs.umn.edu/~prof/class/ *(37 acc. to Page Usage Popularity)*
7. 0.00061524 www.cs.umn.edu/grad/
8. 0.00056791 www-users.cs.umn.edu/~uname/os/
9. 0.00050809 www-users.cs.umn.edu/~prof2/courses/CNUM/ *(72 acc. To Page Usage Popularity)*

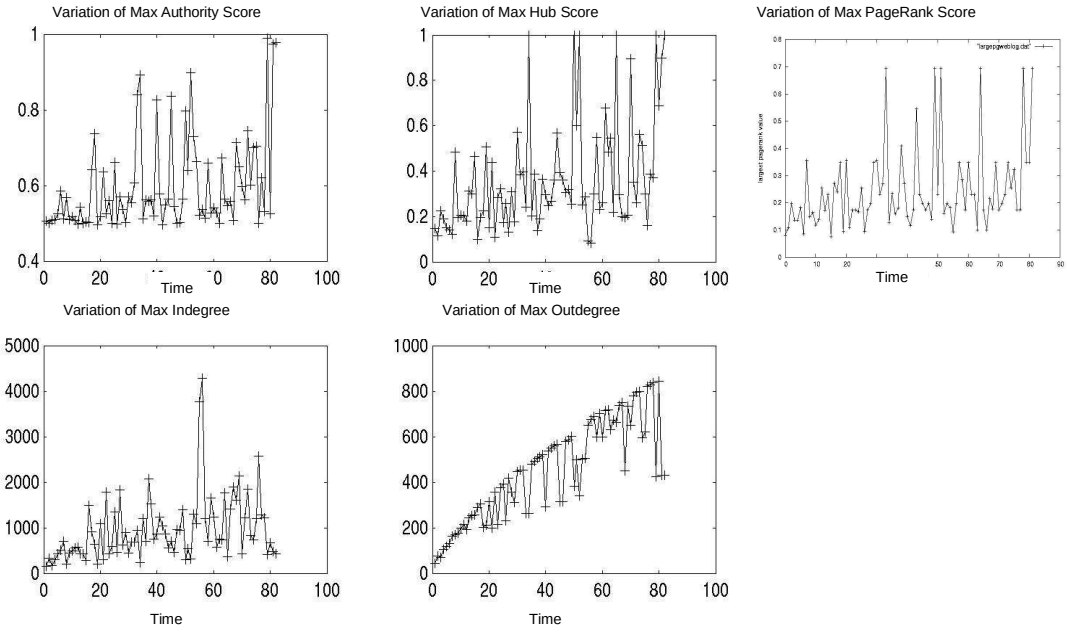
(c)

Figure 5: (a) PageRank Scores for CS Website
(b) Page Usage Popularity for CS Website
(c) Usage Aware PageRank Scores for CS Website

Appendix B



(a)



(b)

Figure 6: (a) Variation of Basic Properties - Order, Size and Number of Components
(b) Variation of Derived Properties