CSci 8799: Geographic Databases for Management of Spatial Graphs

Prof. Shashi Shekhar (shekhar@cs.umn.edu)
EE/CS 5-203; Office Hours: M 2-3pm. Telephone: (612) 624-8307

Topics to be Covered:

Application Domains (Transportation, Telephone, Electric/Gas Utilities, Circuits, etc.), Common data types, queries and analysis, Data Models for Graphs/Networks, Languages to Query Graphs/Networks , Query Optimization, Access Methods, Clutering Methods and File Structures, System Architectures and Design.

Textbooks and References:

Collection of papers available as a reader.

Activties:

The main goal of this class is to design a geographic database system for managing Graph/Network data. The system will consist of a set of independent modules, which will work from a set common file-formats. The system will be loosely-coupled to allow independent design/optimization of each component. Following are the two main activities in this class:

1. Analysis of current research literature and application domains to identify the requirements. A short proposal on the component of interest is expected by fifth week. It should summarize the relevant research results and define the requirements. The proposal will be presented in the class for discussion to ensure compatibility with rest of the system.
2. High-level design of a database system for managing graphs/networks. A short design document is expected by the end of the quarter. It should describe its interface with other components (via file formats), psuedo code for external services along with a set of test-cases to define the behaviour. The selection of file/data-structures and algorithms should be justified by comparing those to the alternatives. A simple prototype implementation of the design is desirable but not required.

Collaboration:

Collaboration among participants is expected. After all, we want to come up with a single coherent and consistnet design document for the whole system. It is desirable that participants discuss their design decisions with other participants during and outside the class meetings. It would however be useful for each participant to choose an independent component to ensure diversity of ideas and progress on the system design.

Helpful Comments:

This class is Very Interesting and Useful. We are exploring a current research area which is very important yet fairly open for research. A large number of application domains (e.g. transportation, telephone companies, electricity/gas distribution companies, urban management bodies, etc.) are beginning to build database/GIS to monitor and anlyze their networks. Most of them have to create custom systems since there is no consensus on the data-models, query languages and analysis functions for spatial networks. To get full benefit out of the class you have to work independently and regularly. Read the papers before the meeting and bring comments for discussion. Plan to spend at least 6 hrs a week (a little more during first few weeks till you feel comfortable with spatial graphs/networks) on this class doing projects or reading.

Good Luck, and Welcome to CSci 8799!

Shashi Shekhar

SCHEDULE

WEEK	DATE	TOPICS	DUE
1	W 3/27 1230-2pm	High Performance GIS	
2	W 4/3 930-11am	Terminology	
3	W 4/10 930-11am	Access Methods, Clustering	
4	M 4/15 1230-2pm	Graph Analysis Algorithms	
5	M 4/22 1230-2pm	Data Models, Query Languages	
6	M 4/29 1230-2pm		Proposal Presentations
7	M 5/6 1230-2pm	Query Optimization	
8	M 5/13 1230-2pm	Presentation (Visualization)	
9	M 5/20 1230-2pm	System Architecture	
10	W 5/29 930-11am		Design Presentations
11	M 6/3 1230-2pm		Design Presentations

SUGGESTED TOPICS

1. Applications

Environmental Studies[1], Intelligent Transportation Systems, [2-4],

2. Terminology: GIS, GIS-T, Problems

Geographic Information Systems and Spatial Database[5], Geographic Information Systems for Transportation[6]

3. Data Models, Query Languages

Graph data models and simple queries[7, 8], Query languages for graph traversals[9]

4. Query Optimization

Bulk loading[10], Single Pair Shortest Path[11], Transitive Closure[12].

4. Graph Analysis Algorithms

Network Analysis[13], Graph Partitioning[14, 15].

5. Access Methods, Clustering

Clustering objects[16], Transitive closure and traversal recursion[17-19], Clustering DAG for CAD Databases[20].

6. Data Input and Quality

Road map quality[21]. Remote sensing in GIS[22], Digitizing road maps[23].

7. Presentation (Visualization)

Graphical presentation language[24], Entity Relational model for visual representations[25], visual representations for relational data26].

8. System Architecture

Archive-architecture for global change data[27], extensible database architecture[28, 29], Using OO/extensible database for GIS[30, 31], Interpolation function in SQL[32].

References

1. M. Stonebraker, J. Frew, K. Gardels and Jeff Meredith, The Sequoia 2000 Benchmark, *Proc. of Intl Conference on Management of Data*, ACM, (1993).

2. R. K. Jurgen, The Electronic Motorist, *Spectrum*, IEEE, (March 1995).

3. W. C. Collier and R. J. Weiland, Smart Cars, Smart Highways, *Spectrum*, IEEE, (April 1994). 27-33
4. S. Shekhar and D. R. Liu, Genesis and Advanced Traveler Information Systems (ATIS), *NSF MOBIDATA Workshop on Mobile and Wireless Information Systems*, (November, 1994).
5. R. H. Guting, An Introduction to Spatial Databases, *VLDB Journal* **3**(Fall 1994).
6. A. P. Vonderohe and et. al., Adaption of GIS for Transportation, *NCHRP Report 359*, Transportation Research Board, National Research Council, (1993).
7. R. H. Guting, GraphDB: Modeling and Querying Graphs in Databases, *Proc. of the 20th VLDB conference*, (1994,). 297-308
8. Alejandro Gutierrez et al, Database Graph Views: A Practical Model to Manage Persistent Graphs, *Proc. of the 20th VLDB Conference*, (1994,). 391-402
9. M. Mannino and L.D. Shapiro and Extensions to Query Languages for Graph Traversal Problems, *Trans. on Knowledge and Data Engineering* **2**(3)IEEE, (Sept., 1990).
10. J. L. Wiener and J. F. Naughton, Bulk Loading into an OODB: A Performance Study, *Proc. of the 20th VLDB conference*, (994). 120-131
11. S. Shekhar, A. Kohli and M. Coyle, Path Computation Algorithms for Advanced Traveler Information System, *Proc. of the Ninth Intl Conference on Data Engineering*, IEEE, (April 1993). 31-39
12. Y. Ioannidis and R. Ramakrishnan and L. Winger, Transitive Closure Algorithms Based on Graph Traversal, *Trans. on Database Systems* **18**(3)ACM, (Sept. 1993).
13. Network Analysis: Modeling Network Systems (version 6.1), *ARC/INFO User's Guide*, ESRI, (April 1992).
14. B.W. Kernighan and S. Lin, An Efficient Heuristic Procedure for Partitioning Graphs, *Bell Syst. Tech. J.* **49**(2)(February 1970). 291-307
15. G. Karypis and V. Kumar, A Fast and High Quality Multilevel Scheme for Partitioning Irregular Graphs, *Csci TR 95-035*, (1995). Proc. Intl Conference on Parallel Processing
16. M. M. Tsangaris and Jeffrey.F. Naughton, On the Performance of Object Clustering Techniques, *Proc. of SIGMOD Conference on Management of Data*, ACM, (1992). 144-153
17. P.A. Larson and V. Deshpande, A File Structure Supporting Traversal Recursion, *Proc. of the SIGMOD Conference*, ACM, (1989). 243-252
18. R. Agrawal and Jerry Kiernan, An Access Structure for Generalized Transitive Closure Queries, *Proc. of the Ninth Intl Conference on Data Engineering*, IEEE, (April 1993). 429--438
19. D. R. Liu, S. Shekhar and M. Coyle, An Evaluation of Access Methods for Spatial Networks, *Proceedings of the Second ACM Workshop on Advances in Geographic Information Systems*, ACM, (1994). Also Csci TR 93-78 (To appear in IEEE TKDE)
20. J. Banerjee and W. Kim, S. Kim and J. Garza, Clustering a DAG for CAD Databases, *IEEE Trans. on Software Engineering* **14**(11) pp. 1684-1699 (Nov. 1988).
21. *Truth-in-labeling Standard For Navigation Map Database*, SAE surface vehicle standard J1663 (Aug. 1995).
22. D. M. McKeown, The role of AI in the integration of remotely sensed data with GIS, *Trans. on GeoScience and Remote Sensing* **GE-25**(3)IEEE, (May 1987).
23. A. Zlotnick, Finding Road Seeds in Aerial Images, *Image Understanding Jr.* **57**(2)Academic Press, (March 1993).
24. M. J. Egenhofer, Spatial SQL: A Query and Presentation Language, *Trans. on Know. and Data Eng.* **6**(1)IEEE, (Feb. 1994).
25. M. Ganesh et.al., Visual Data Mining: Framework and Algorithm Development, *Computer Sc. Dept. Tech. Report 96-021*, Univ. of Minnesota, ().
26. J. Mackinlay, Automating the Design of Graphical Presentations of Relational Information, *Trans. on Graphics* **5**(2)ACM, (April 1986). (graphing relational information)
27. M. Stonebraker, Sequoia 2000: A Reflection on the First Three Years, *Computational Science and Engineering*, IEEE, (Winter 1994).
28. M. Stonebraker, The POSTGRES Next Generation Database Management System, *Communication of ACM* **34**(10)(1991).
29. DeWitt et.al., Client Server Paradise, *Proc. VLDB Conf.*, (1994).
30. J. A. Orenstein and F. A. Manola, PROBE Spatial Data Modeling and Query Processing in an Image Database Application, *Trans. on Software Eng.* **14**(5)IEEE, (May 1988).
31. B. C. Ooi, R. Sacks-Davis, and K. J. McDonnel, Extending A DBMS For Geographic Applications, *Proc. Intl. Conf. on Data Eng.*, IEEE, (1989).

32. L. Neugebauer, Extending a Database to Support the Handling of Environmental Measurement Data, *Proc. Symp. on Large Spatial Database*, Springer Verlag LNCS Series, (1989).