Deduced Social Networks for Educational Portal

Monika Akbar and Clifford A. Shaffer
Department of Computer Science, Virginia Tech
{amonika, shaffer}@vt.edu

Abstract
Educational portals such as Algoviz.org contain rich information resources. A key concern is directing users to specific resources that are of interest to them. While AlgoViz has significant traffic, we cannot count on active user participation in the form of explicit ratings of individual resources. Lacking active user data (e.g., user ratings on resources), we instead use log data to deduce user trends. We describe our techniques for clustering users based on the log data. We show how cluster analysis can be used to improve searching and browsing within AlgoViz. Our approach has the potential to be useful for a wide range of educational resource portals.

Generating and Analyzing DSN in AlgoViz

Getting the Data
- Store user activities.
- Various logging options: Server log, system logs.
- Sites such as Google Analytics provide more advanced metrics like visits, pageviews, bounce rate, time on site, etc.
- Data Cleaning: Remove spammers, bots, crawlers, etc.

Is there any connection between the users?
- Generate DSN between users based on various criteria such as pageviews, ratings, reviews, etc.

Within the network, are there groups of users?
- Graph Partitioning (Modularity clustering)

What are the interests of the groups?
- Topic modeling
  - Latent Dirichlet allocation (LDA)

Findings
- DSNs can identify connections between anonymous users who are otherwise disconnected.
- DSNs can identify the existence of groups with specific interest.
- DSNs can be used to improve search on sites with high volumes of anonymous traffic.

Applications
- Catalog Entry (CE) Ranking
  - We use a custom ranking function that places different weights on AlgoViz-specific fields of an Algorithm Visualization catalog entry.
  - Clusters representing a specific content type are used to add weight to content of that type. Top contents \( c_1, c_2, c_3, \ldots, c_p \) of cluster \( x \) that is dominated by a content type of \( y \) (e.g., forum, page, catalog entry, etc.), receive certain points.
  - Search results and browsing list are ordered by ranking score.
  - Top contents of clusters dominated by a specific content type are used to create recommendations.

This work has been supported by NSF DUE-0836940, DUE-0937863, DUE-0840719.