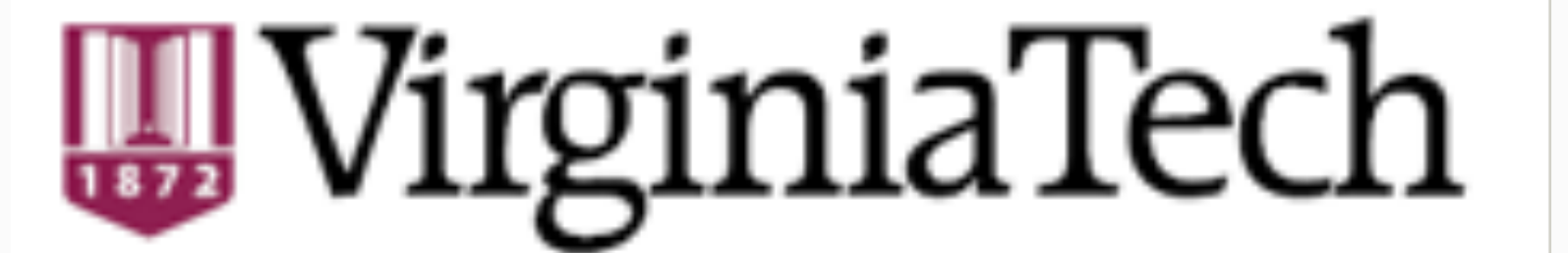# Deduced Social Networks for Educational Portal

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.

Deduced Social Network (DSN)

A Deduced Social Network (DSN) is a Graph with tuple (*Entity*, *Object*, k), where
- *Entity* is the node of the graph,
- *Object* is an attribute linked to *Entity*, where one *Entity* can have multiple *Object*(s), and
- k is a constant or a function that returns the minimum number of *Object*(s) that must be common between two *Entities* to create a connection (i.e., edge) between them.

Example: In a user-based DSN, a connection threshold k=5 for an edge indicates that two users (i.e., *Entity*) have viewed at least *5* common pages (i.e., *Object*). In an object-based DSN, k=3 for an edge would indicate that there are at least 3 users who viewed the two objects (e.g., pages).

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 one site, etc.


Sample entries of AlgoViz log data.

Session ID	Page Title	Internal Path/Page URL	Hostname‡	User ID	Timestamp
ievav83	Lifting the hood of the computer...	node/1413	98.108.111.25	0	1276272047
t5fuuba	biblio/export/ tagged/118/ popup	research.cs.vt.edu/algoviz /biblio	217.23.228.192	0	1276260935
ivuks8s	Has an AV helped you learn a topic in computer science?	research.cs.vt.edu/algoviz /poll/	101.79.145.90	0	1276260943

‡ IPs are masked to protect user identity.

- Data Cleaning: Remove spammers, bots, crawlers, etc.

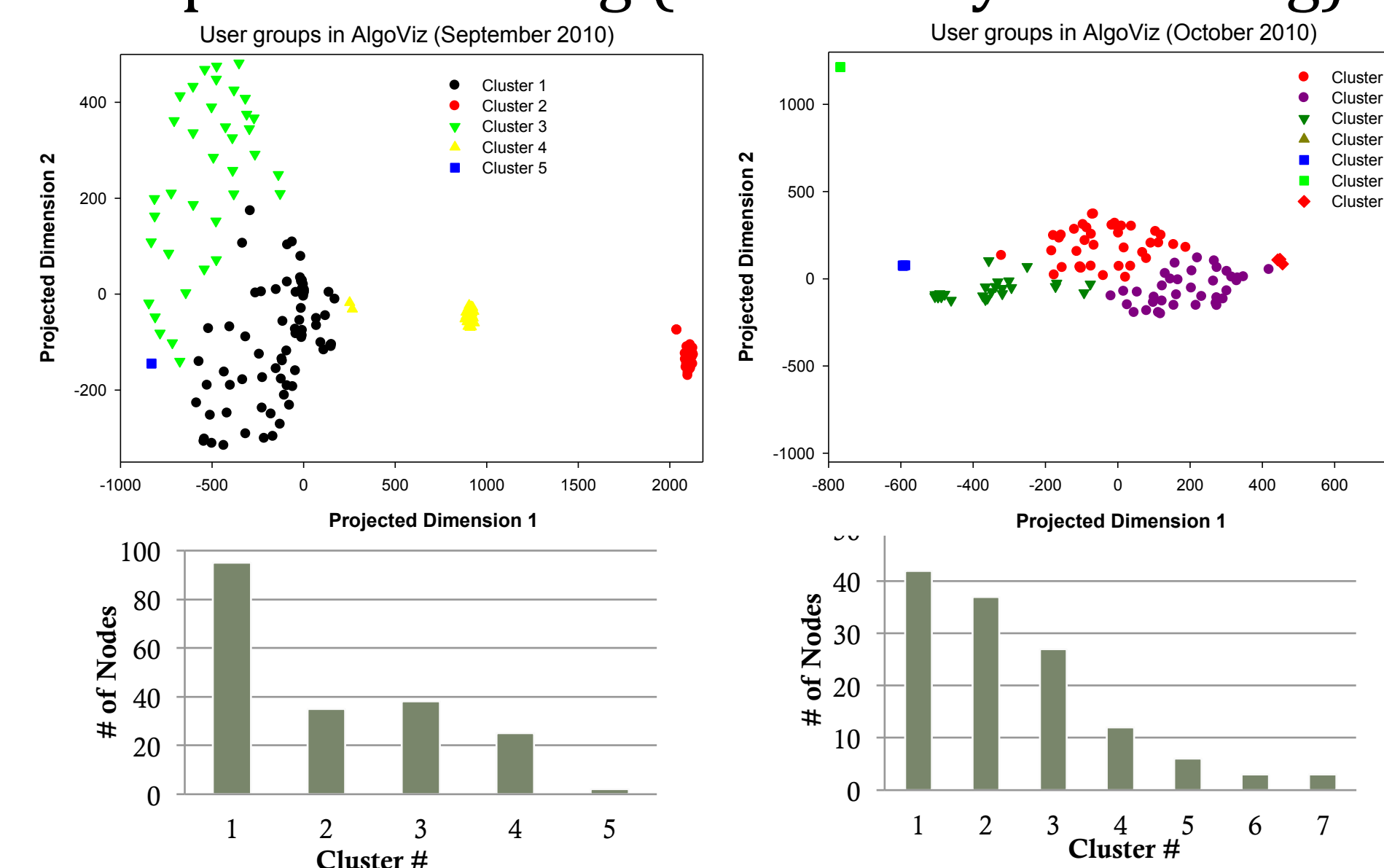### Is there any connection between the users?

- Generate DSN between users based on various criteria such as pageviews, ratings, reviews, etc.
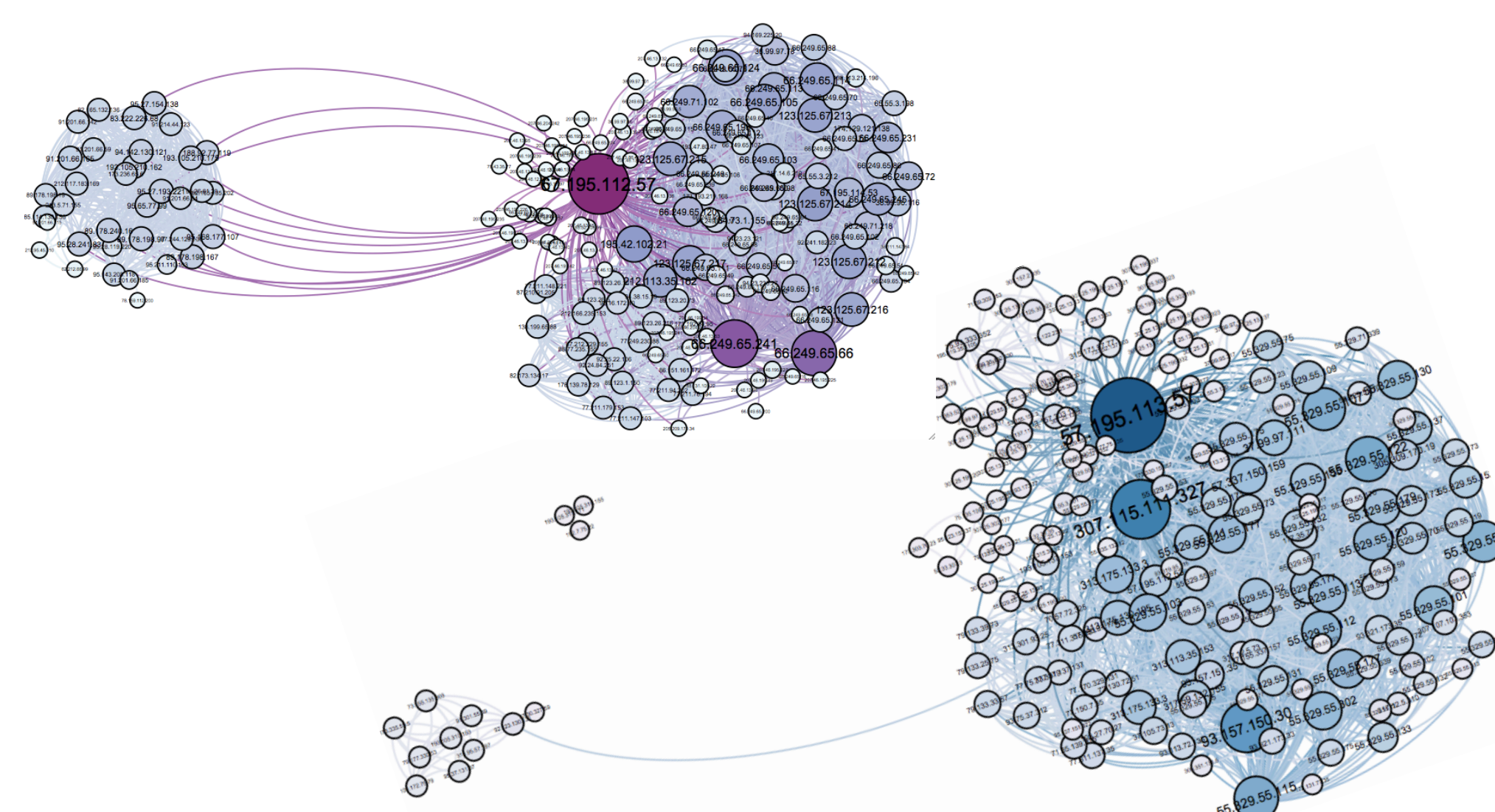
AlgoViz DSN for Sept-10 (top) and Oct-10 (bottom) (k = 10)

Within the network, are there groups of users?

- Graph Partitioning (Modularity clustering)

Clusters and cluster- sizes of Sept-10 DSN (left) and Oct-10 DSN (right) (k = 10)

What are the interests of the groups?

- Topic modeling
  - Latent Dirichlet allocation (LDA)

Topic ID	Words in Topic
1	biblio export xml bibtex rtf set
2	biblio author bibtex export function algorithm
3	algorithms data author trees demo computer
4	visualization sort algorithm structure tree animation
5	biblio java sorting programming learning sorts

Clust.	Top Topic (1)	Contrib.	Top Topic (2)	Contrib.	Top Topic (3)	Contrib.
1	3	0.667	5	0.25		
2	3	0.312	1	0.212	5	0.209
3	1	0.539	2	0.193	4	0.119
4	3	0.75	5	0.156		
5	2	0.254	1	0.216	4	0.213

| Topic ID	Words in Topic
1	forum list binary education path biblio
2	trees learning automata animations visual systems
3	biblio export rtf linked functional teach
4	algorithm visualization animation computer data java
5	author algorithms program animal software programming
6	biblio export bibtex xml author node
7	author view sort tree limit structure

Clust.	Top Topic (1)	Contrib.	Top Topic (2)	Contrib.	Top Topic (3)	Contrib.
1	3	0.987				
2	1	0.828	2	0.138		
3	6	0.421	5	0.145	3	0.12
4	1	0.815				
5	5	0.253	4	0.217	6	0.137
6	3	0.326	6	0.161	4	0.144
7	6	0.306	3	0.221	7	0.119

Topics and cluster-wise topic distribution for Sept-10 (top) and Oct-10 (bottom) DSN

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.
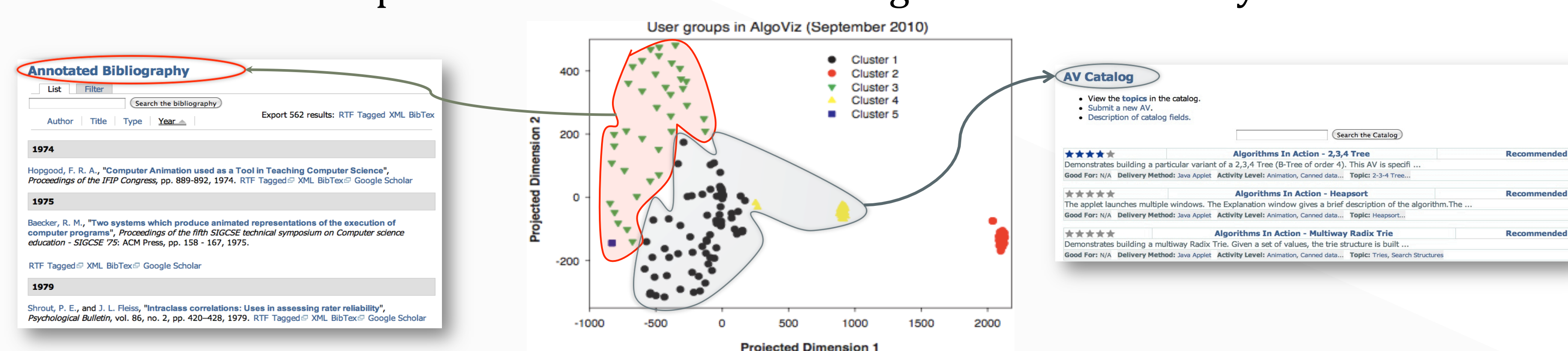
Applications

- Refine Ranking and Recommendations
  - 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, ..., c_n 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.

Catalog Entry (CE) Ranking

$$score(CE_i) = x + y + ... + z$$

where, $x = \begin{cases} 20 & \text{if the CE has 'Yes' in the 'Works' field} \\ 0 & \text{otherwise} \end{cases}$,

$y = \begin{cases} 6 & \text{if the CE is 'Recommended'} \\ 0 & \text{otherwise} \end{cases}$, and

$z = \begin{cases} 3 & \text{if the CE was present at least } m \text{ times, in a cluster dominated by Catalog Entry content type} \\ 0 & \text{otherwise} \end{cases}$.

The recommendation block (left) in AlgoViz for catalog entries based on cluster results that are dominated by this content type