# A Visualization Model Based on Adjacency Data

- Motivation

  Typically, data are provided in multidimensional format

  - Large table where the rows represent countries and the columns represent socio-economic variables

  Alternatively, data may be provided in in adjacency format

  - Consumers who buy item *a* are likely to buy or consider buying items *b*, *c*, and *d* also
  - Students who apply to college *a* are likely to apply to colleges *b*, *c*, and *d* also

E. Condon, B. Golden, S. Lele, S. Raghavan, E. Wasil, *Decision Support Systems*, *33*, 349-362 (2002)

# A Visualization Model Based on Adjacency Data

- Adjacency

  If the purchase of item $i$ results in the recommendation of item $j$, then item $j$ is *adjacent* to item $i$

  Adjacency data for $n$ alternatives can be summarized in an $n$ x $n$ *adjacency matrix*, $A = (a_{ij})$, where

$$a_{ij} = \begin{cases} 1 & \text{if item } j \text{ is adjacent to item } i \\ 0 & \text{otherwise} \end{cases}$$

  Adjacency is not necessarily *symmetric*

# A Visualization Model Based on Adjacency Data

- Adjacency

Adjacency indicates a notion of similarity

Given adjacency data with respect to *n* items or alternatives, can we display the items in a *two-dimensional map*?

Traditional tools such as *multidimensional scaling* (MDS) and *Sammon maps* work well with data in multidimensional format

# A Visualization Model Based on Adjacency Data

- Adjacency

*Sammon map technique* (1969) places points in 2-space to minimize following objective function

$$\left(\sum_{i<j}(x_{ij}-y_{ij})^2/x_{ij}\right)/\sum_{i<j}x_{ij}$$

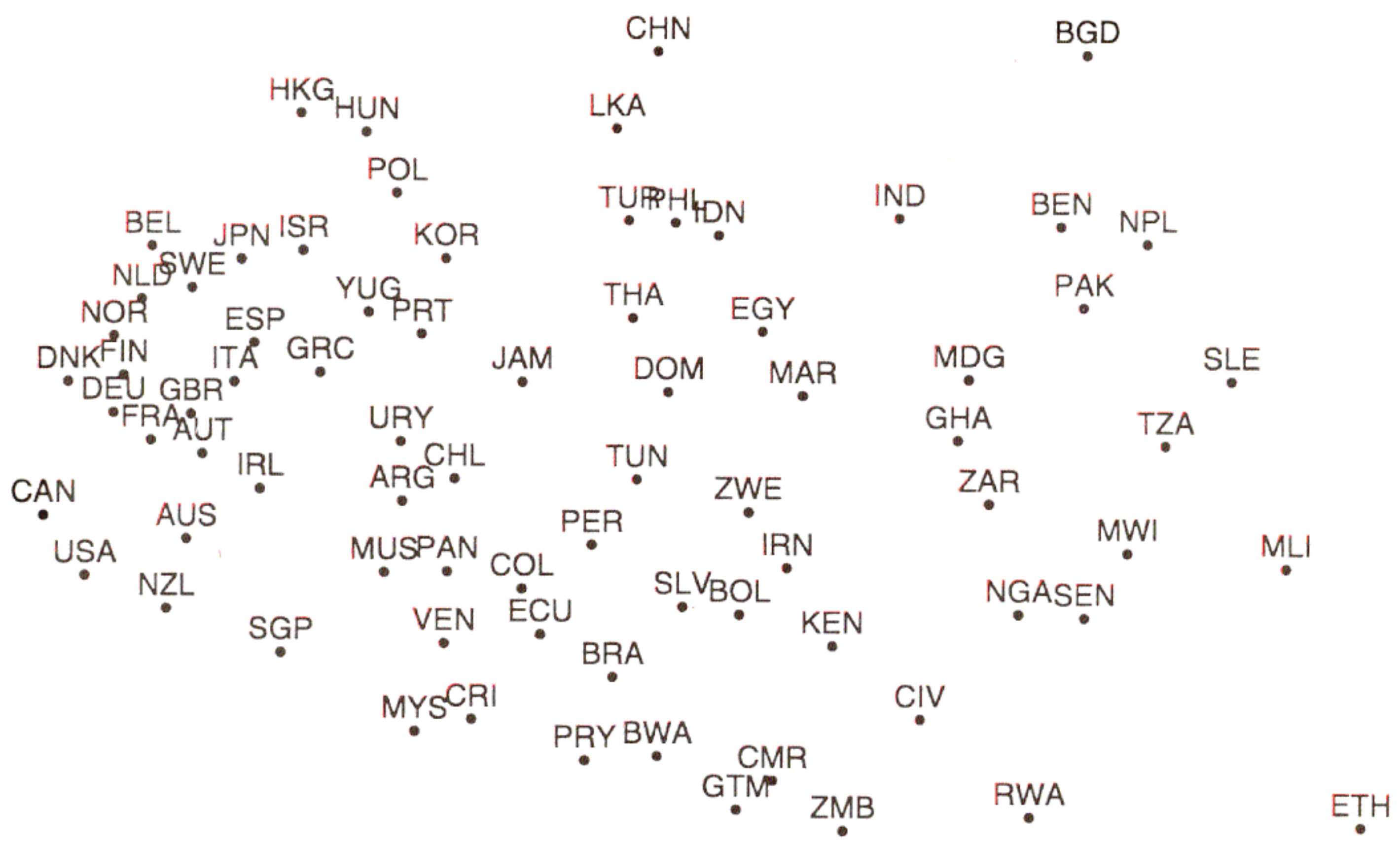
$x_{ij}$ = distance measure between points *i* and *j*

$y_{ij}$ = distance between points *i* and *j* in a 2-dimensional plot

Can Sammon maps and MDS work well with adjacency data?

# A Visualization Model Based on Adjacency Data

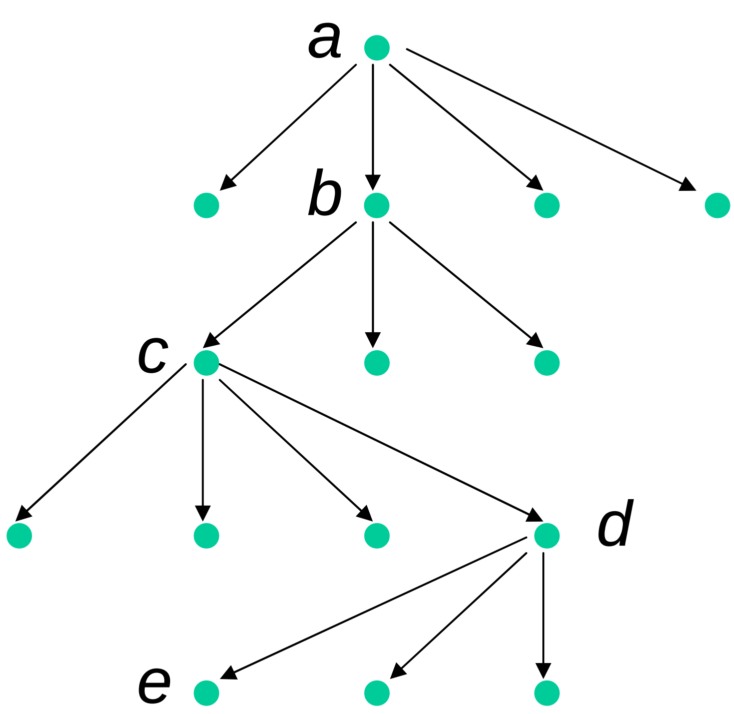
- Sammon Map of World Poverty (World Bank, 1992)



# A Visualization Model Based on Adjacency Data

▶ Obtaining Distances from Adjacency Data

How can we use linkage information to determine distances ?




items adjacent to *a*

items adjacent to *b*

items adjacent to *c*

items adjacent to *d*

# A Visualization Model Based on Adjacency Data

▶ Obtaining Distances from Adjacency Data

1. Start with the $n$ x $n$ 0-1 asymmetric adjacency matrix
2. Convert the adjacency matrix to a directed graph

    Create a node for each item ($n$ nodes)
    Create a directed arc from node $i$ to node $j$ if $a_{ij} = 1$

3. Compute distance measures

    Each arc has a length of 1
    Compute the all-pairs shortest path distance matrix $D$
    Distance from node $i$ to node $j$ is $d_{ij}$

# A Visualization Model Based on Adjacency Data

▶ Obtaining Distances from Adjacency Data

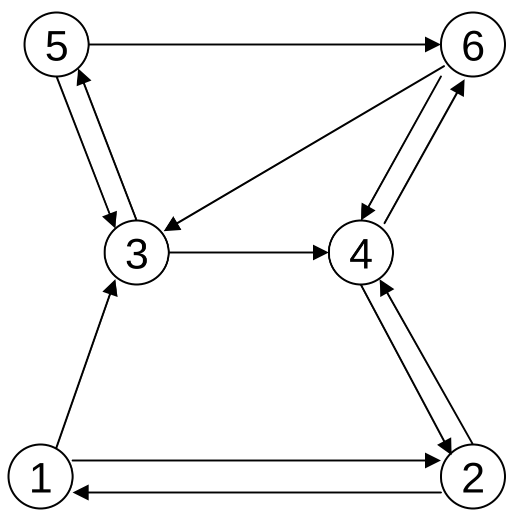
4. Modify the distance matrix *D*, to obtain a final distance matrix *X*

    Symmetry
    Disconnected components

Example 1

A =

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 1 | 0 | 0 | 0 |
| 2 | 1 | 0 | 0 | 1 | 0 | 0 |
| 3 | 0 | 0 | 0 | 1 | 1 | 0 |
| 4 | 0 | 1 | 0 | 0 | 0 | 1 |
| 5 | 0 | 0 | 1 | 0 | 0 | 1 |
| 6 | 0 | 0 | 1 | 1 | 0 | 0 |



# A Visualization Model Based on Adjacency Data

▶ Obtaining Distances from Adjacency Data

Find shortest paths between all pairs of nodes to obtain *D*

Average $d_{ij}$ and $d_{ji}$ to arrive at a symmetric distance matrix *X*

$$D = \begin{array}{c|cccccc} & 1 & 2 & 3 & 4 & 5 & 6 \\ \hline 1 & 0 & 1 & 1 & 2 & 2 & 3 \\ 2 & 1 & 0 & 2 & 1 & 3 & 2 \\ 3 & 3 & 2 & 0 & 1 & 1 & 2 \\ 4 & 2 & 1 & 2 & 0 & 3 & 1 \\ 5 & 4 & 3 & 1 & 2 & 0 & 1 \\ 6 & 3 & 2 & 1 & 1 & 2 & 0 \end{array} \qquad X = \begin{array}{c|cccccc} & 1 & 2 & 3 & 4 & 5 & 6 \\ \hline 1 & 0 & 1 & 2 & 2 & 3 & 3 \\ 2 & 1 & 0 & 2 & 1 & 3 & 2 \\ 3 & 2 & 2 & 0 & 1.5 & 1 & 1.5 \\ 4 & 2 & 1 & 1.5 & 0 & 2.5 & 1 \\ 5 & 3 & 3 & 1 & 2.5 & 0 & 1.5 \\ 6 & 3 & 2 & 1.5 & 1 & 1.5 & 0 \end{array}$$

# A Visualization Model Based on Adjacency Data

- Obtaining Distances from Adjacency Data

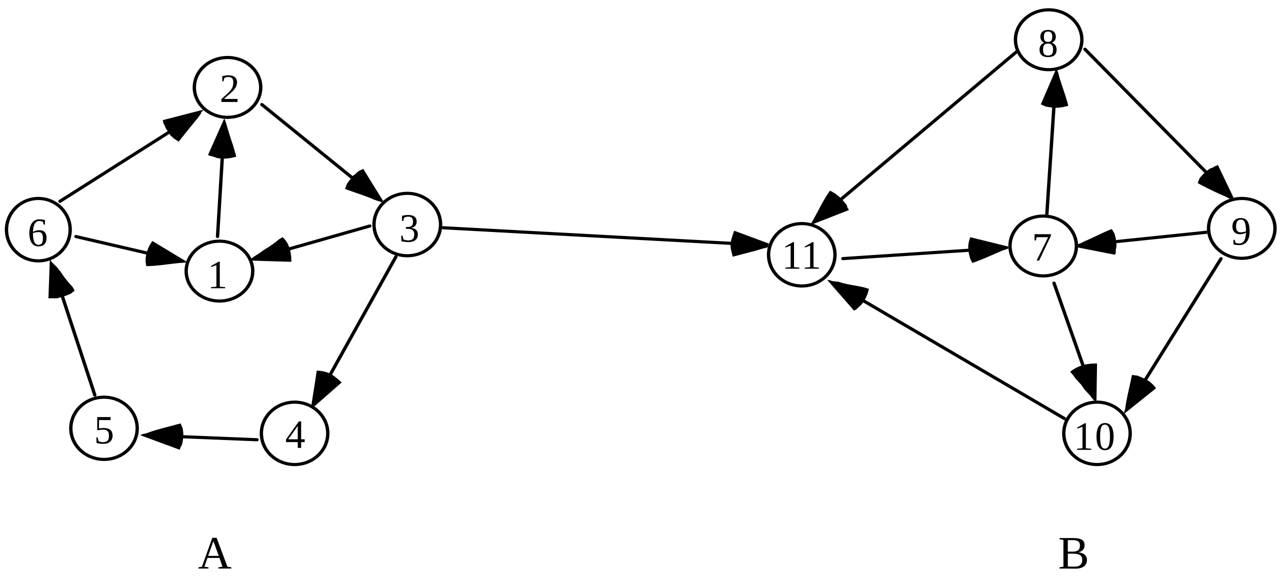
  Example 2

  A and B are *strongly connected* components

  Graph below is *weakly connected*

  Paths from A to B, but none from B to A

  MDS and Sammon maps require *finite* distances



# A Visualization Model Based on Adjacency Data

- Ensuring Finite and Symmetric Distances

  Basic idea: Simply replace all infinite distances with a large finite value, say R

  If R is *too large*

  - Points within each strongly connected component will be pushed together in the map
  - Within-component relationships will be difficult to see

  If R is *too small*

  - Distinct components (e.g., A and B) may *blend together* in the map

# A Visualization Model Based on Adjacency Data

- Ensuring Finite and Symmetric Distances

  R must be chosen carefully

  This leads to a finite distance matrix *D*

  Next, we obtain the final distance matrix *X,* where

  $$x_{ij} = x_{ji} = (d_{ij} + d_{ji})/2$$

  *X* becomes input to a Sammon map or MDS procedure

# A Visualization Model Based on Adjacency Data

- Application: College Selection

  *Data source*: *The Fiske Guide to Colleges*, 2000 edition

  - Contains information on 300 colleges
  - Approximately 750 pages
  - Loaded with statistics and ratings
  - For each school, its biggest overlaps are listed

  *Overlaps*: “the colleges and universities to which its applicants are also applying in greatest numbers and which thus represent its major competitors”

# A Visualization Model Based on Adjacency Data

- Overlaps and the Adjacency Matrix

  Penn’s overlaps are Harvard, Princeton, Yale, Cornell, and Brown

  Harvard’s overlaps are Princeton, Yale, Stanford, M.I.T., and Brown

  Note the lack of symmetry

  - Harvard is adjacent to Penn, but not vice versa

# A Visualization Model Based on Adjacency Data

- Overlaps and the Adjacency Matrix

| School | Brown | Cornell U. | Harvard | MIT | Penn | Princeton | Stanford | Yale |
|---|---|---|---|---|---|---|---|---|
| Brown | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 |
| Cornell U. | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 |
| Harvard | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 |
| MIT | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 |
| Penn | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 |
| Princeton | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |
| Stanford | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 |
| Yale | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |

# A Visualization Model Based on Adjacency Data

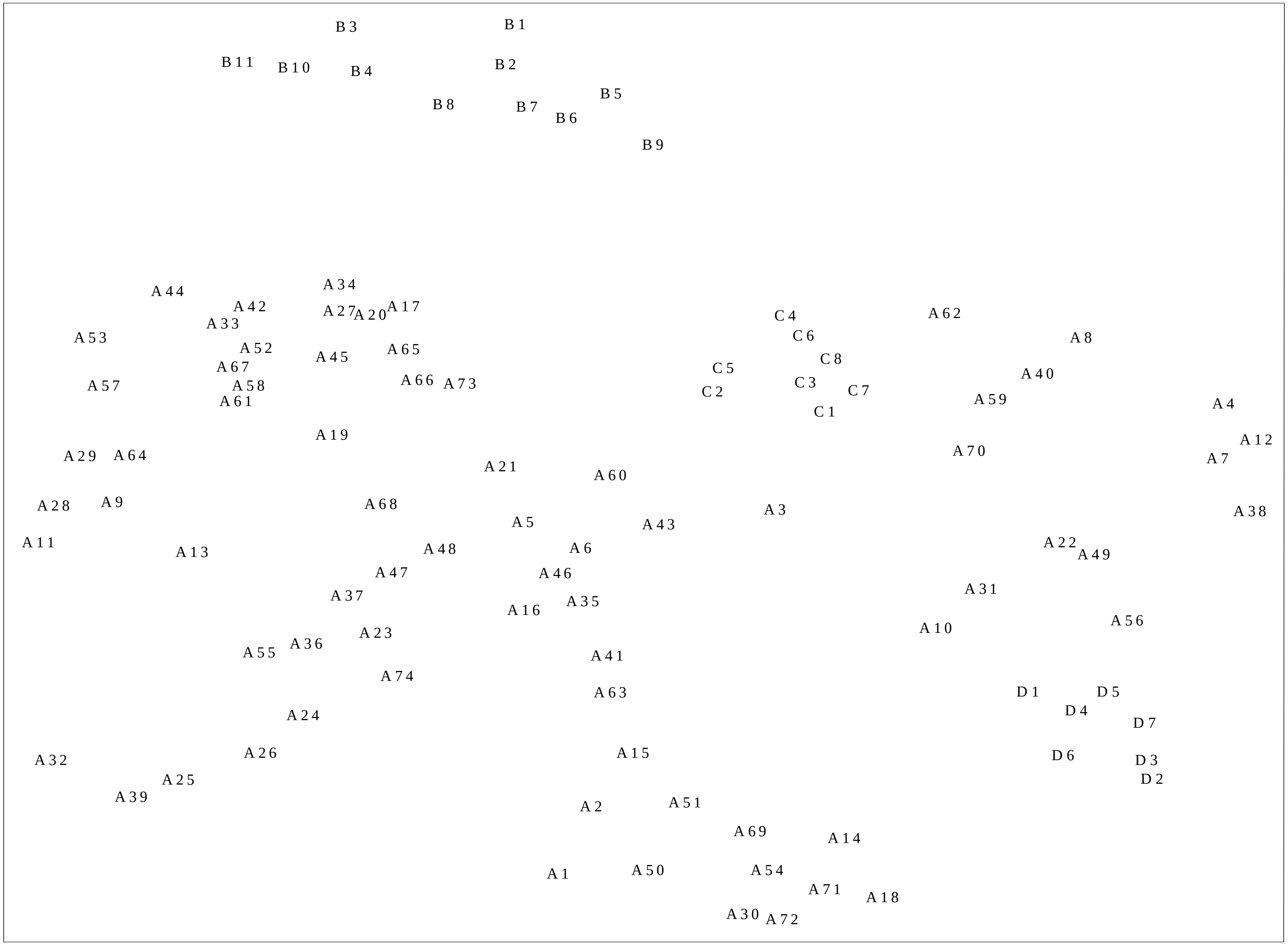
- Application: 100 Colleges and Universities

  Start with 300 colleges and the associated adjacency matrix

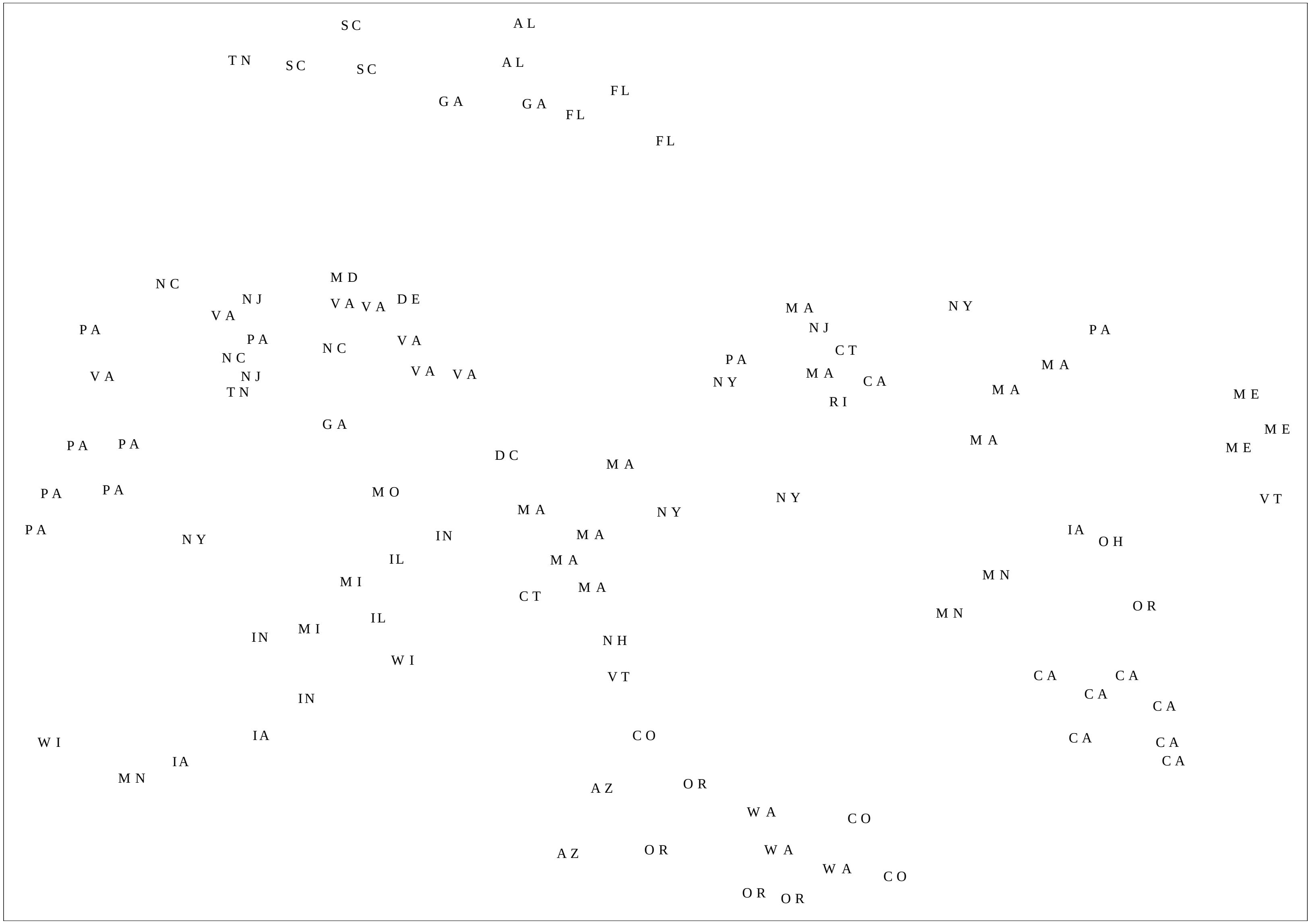
  From the directed graph, several strongly connected components emerge

  We focus on the four largest components

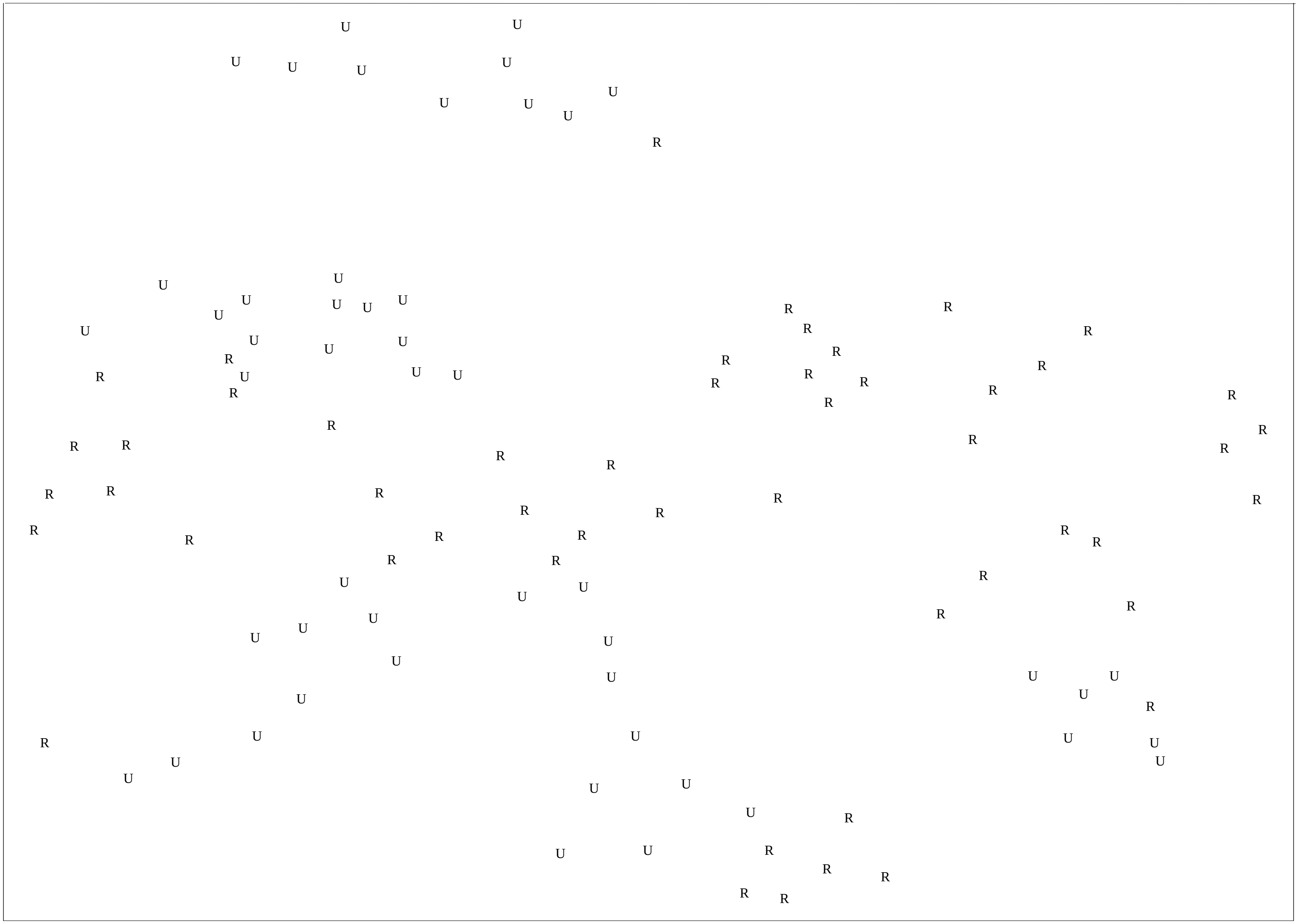
    Component A has 74 schools
    Component B has 11 southern colleges
    Component C has 8 mainly Ivy League schools
    Component D has 7 California universities



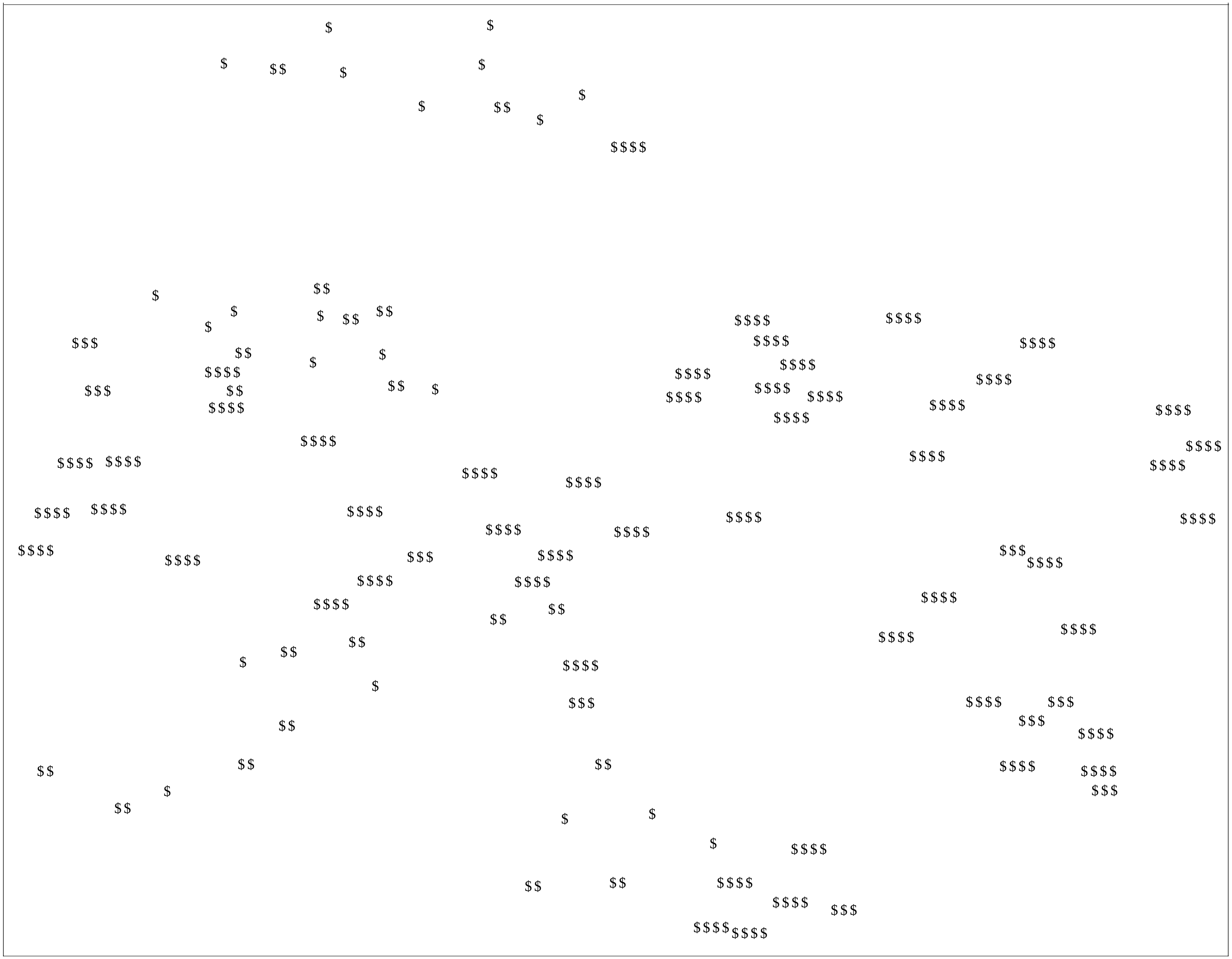

**Sammon Map with Each School Labeled by its Component Identifier**




**Sammon Map with Each School Labeled by its Geographical Location**



**Sammon Map with Each School Labeled by its Designation (Public (U) or Private (R))**



**Sammon Map with Each School Labeled by its Cost**

# A Visualization Model Based on Adjacency Data

- Benefits of Visualization

  Adjacency (overlap) data provides *local* information only

  For example, which schools are Maryland’s overlaps?

  With visualization, *global* information is more easily conveyed

  For example, which schools are similar to Maryland?

# A Visualization Model Based on Adjacency Data

- Benefits of Visualization

  Within group (strongly connected component) and between group relationships are displayed at same time

  A variety of what-if questions can be asked and answered using maps

  Based on this concept, a Web-based DSS for college selection is easy to envision

# A Visualization Model Based on Adjacency Data

- Conclusions

  The approach represents a nice application of shortest paths to data visualization

  The resulting maps convey more information than is immediately available in *The Fiske Guide*

  Visualization encourages what-if analysis of the data

  Can be applied in other settings (e.g., Web-based recommender systems)