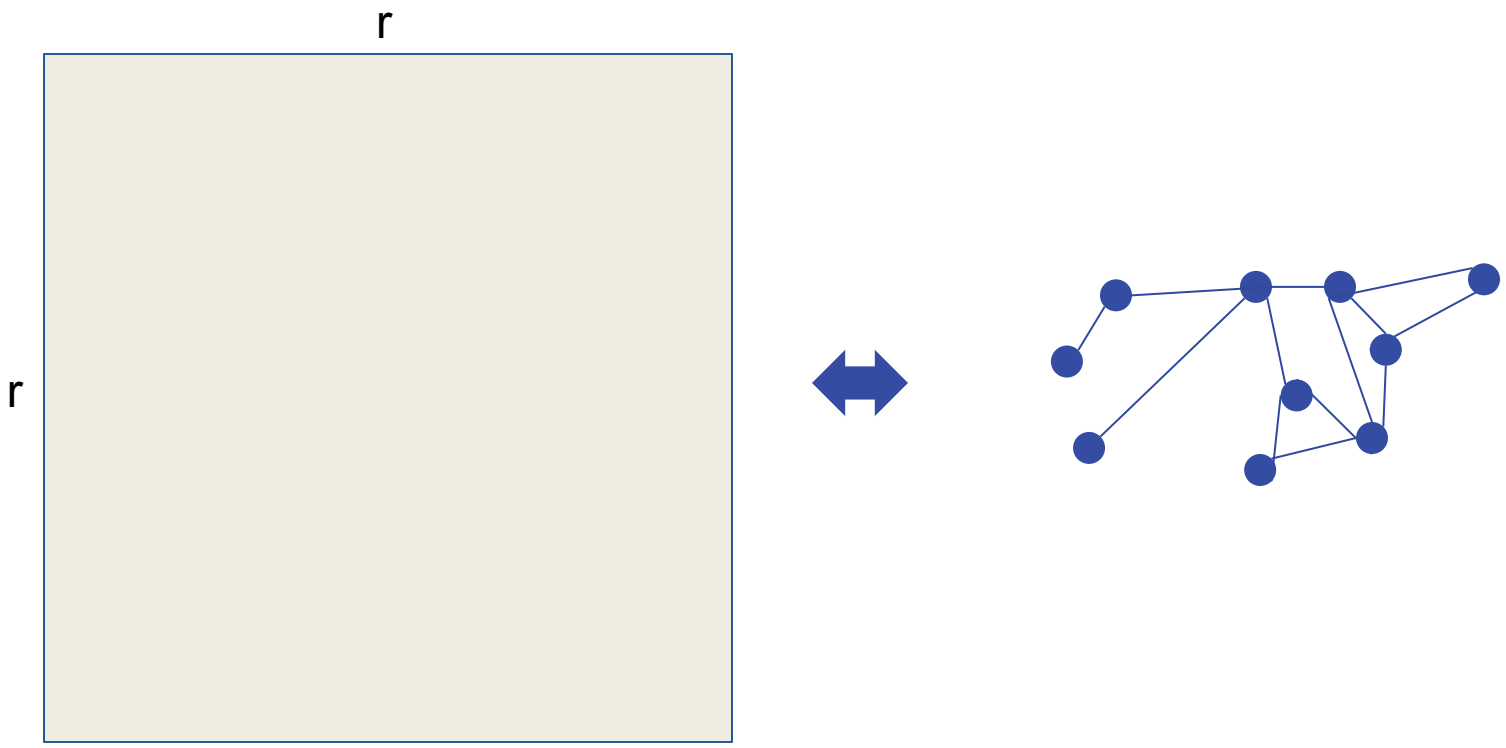
Introduction

Square Matrices



Background

- Square matrices are typically used to represent comparisons among a group of items.
- PageRank
- Cluster analysis
  - Proximity matrices
  - Clustering algorithms: k-means, spectra, hierarchical

From Matrix to Ranking

rest_1	9	4	7	9	4	7	9	4	7
rest_2	9	4	7	9	4	7	9	4	7
rest_3	9	4	7	9	4	7	9	4	7
rest_4	9	4	7	9	4	7	9	4	7
rest_5	9	4	7	9	4	7	9	4	7

	rest_1	rest_2	rest_3	rest_4	rest_5
rest_1	1	4	7	9	4
rest_2	9	1	7	9	4
rest_3	9	4	1	9	4
rest_4	9	4	7	1	4
rest_5	9	4	7	9	1

DataScience@SMU

PageRank

PageRank: I

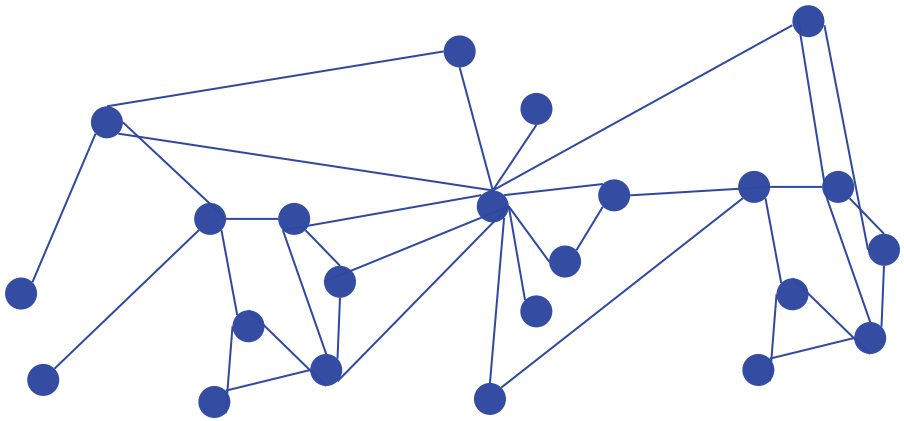
- Node centrality
- Probability distribution of a random walk of infinite length
- Perron-Frobenius establishes single largest eigenvalue and its corresponding eigenvector
- If page *j* → page *i*, set L_{ij} to 1; $c_{i,}$ is a normalizing factor, PageRank for page *i* is:

$$p_i = (1-d) + d\sum_{j=1}^{N}(\frac{L_{ij}}{c_j})p_j$$

- Use eigendecomposition to compute p

PageRank: I

- Random walk analogy: web surfer clicks on links at random choosing to click on an outgoing link with some probability d and (1-d) probability that they jump to a random page (not an outgoing link)



PageRank: II

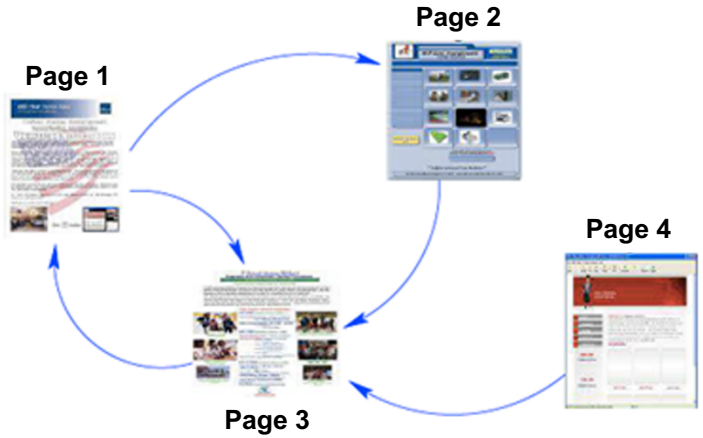


Figure 14.46. *PageRank algorithm: example of a small network*

A small network is shown for illustration in Figure 14.46. The link matrix is

$$\mathrm{L} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad (14.111)$$

and the number of outlinks is $\mathrm{c} = (2,1,1,1)$
The $PageRank$ solution is $\hat{\mathrm{P}} = (1.49, 0.78, 1.58, 0.15)$. Notice that page 4 has no incoming links, and hence gets the minimum $PageRank$ of 0.15.

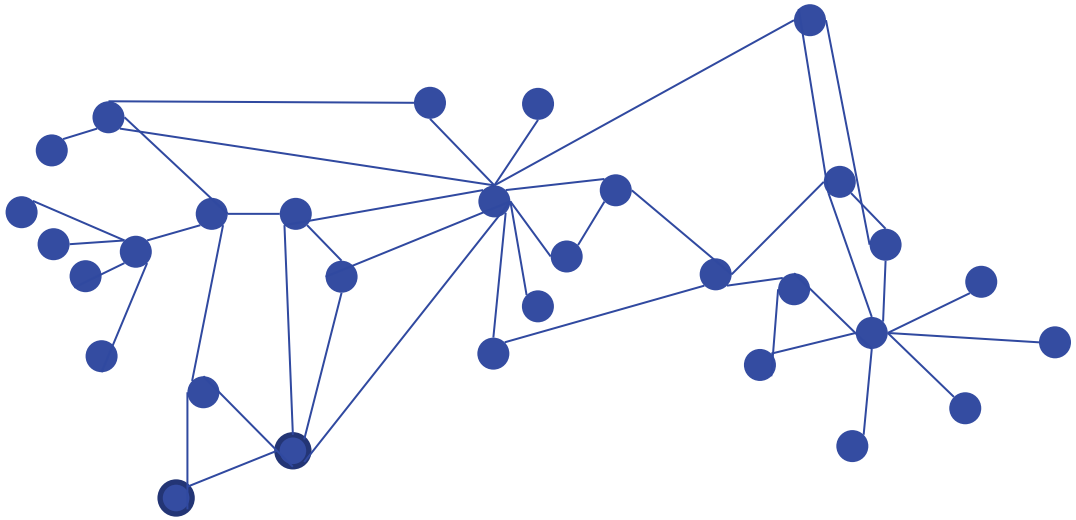
Source: *The Elements of Statistical Learning*, Hastie et al.

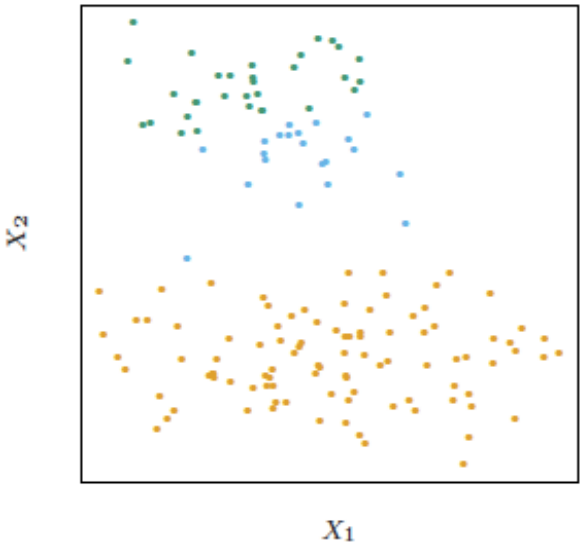
DataScience@SMU

Cluster Analysis

Cluster Analysis

- **Goal:** segmenting/grouping a collection of objects into subsets or clusters, such as those within each cluster, are closely related to one another than to objects in different clusters.
- Measure similarity/dissimilarity between the objects being clustered.
- Additionally, sometime we might want to arrange the clusters themselves into a “hierarchy.”



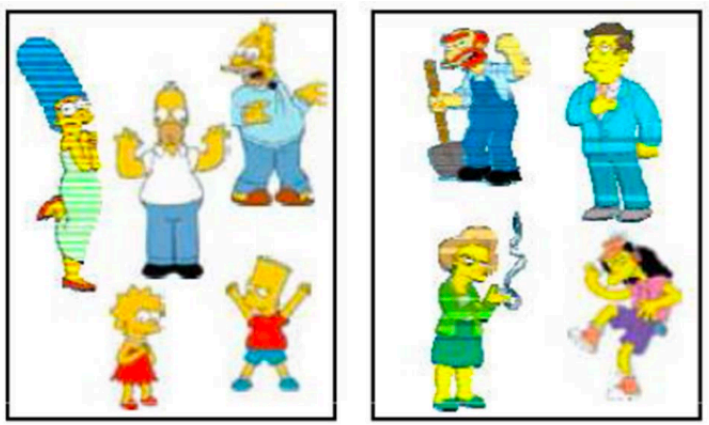


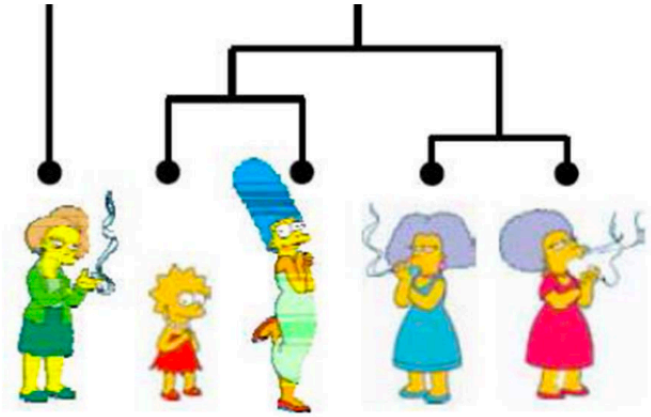
Source: *The Elements of Statistical Learning*, Hastie et al.

Clustering Algorithms

I. Partition (flat) algorithms
   - K-means
   - Mixture of Gaussians
   - Spectral clustering

II. Hierarchical algorithms
   - Agglomerative (bottom up)
   - Top down





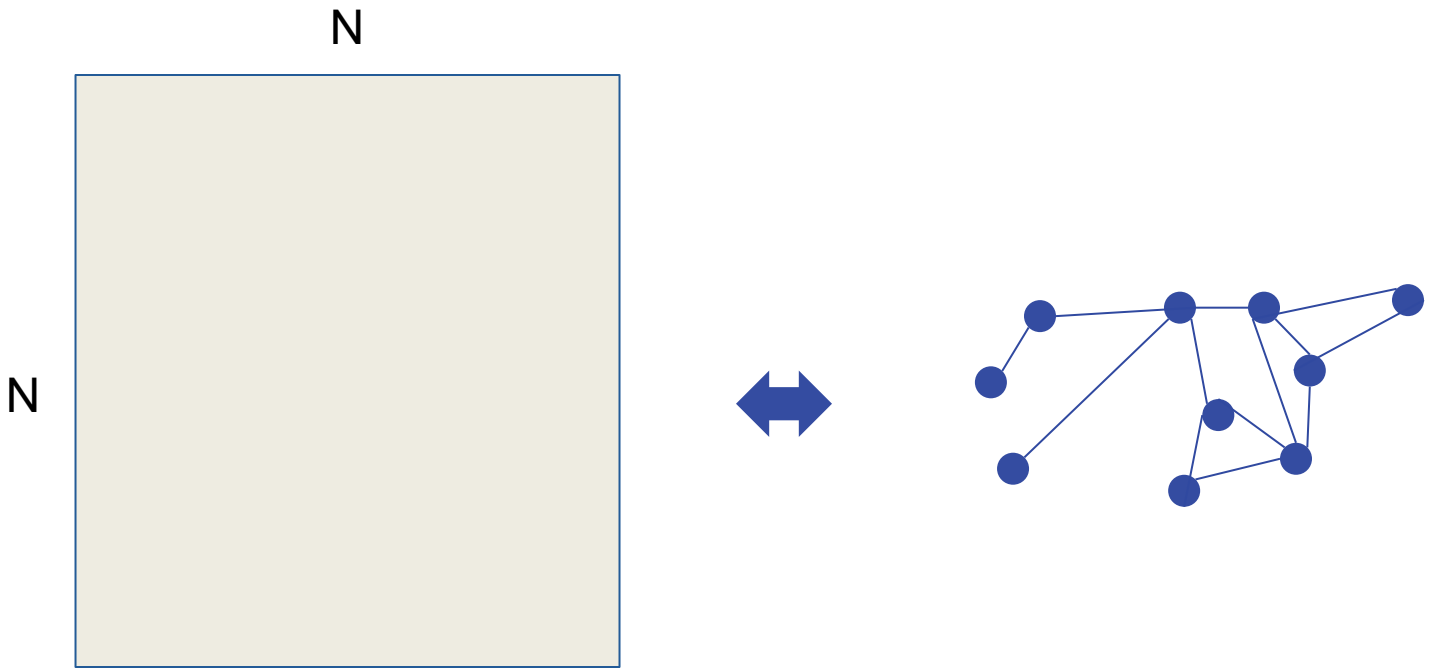
Slide from David Sontag

DataScience@SMU

Proximity Matrices

Proximity Matrices

- The data is represented as pair-wise proximity between data points → *N x N* (i.e., square) matrix **D** in which each element measures the proximity between a pair of data points.
- Typically, clustering algorithms use the dissimilarity between data points (due to the formulation as a minimization problem).

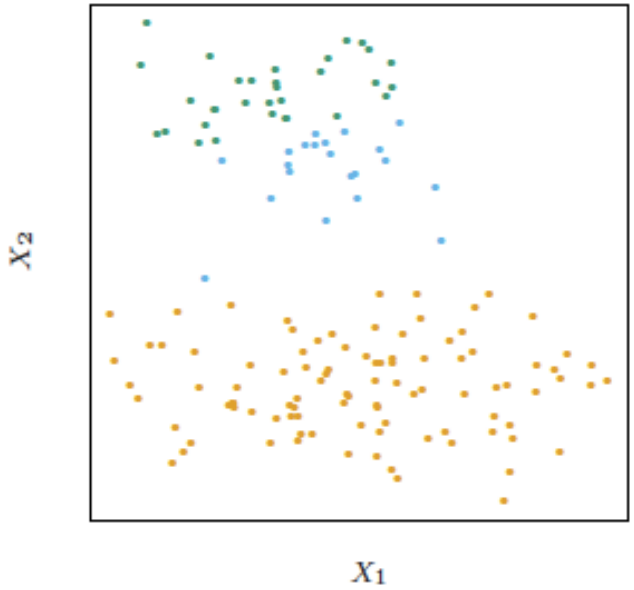


DataScience@SMU

K-Means Analysis

K-Means Clustering

- Start with a guess for each of the three cluster centers.
- Alternate between the following steps until convergence:
  - For each data point, find the closest cluster center (in Euclidean distance).
  - Replace each cluster center with the coordinate-wise average of all data points that are closest to it.

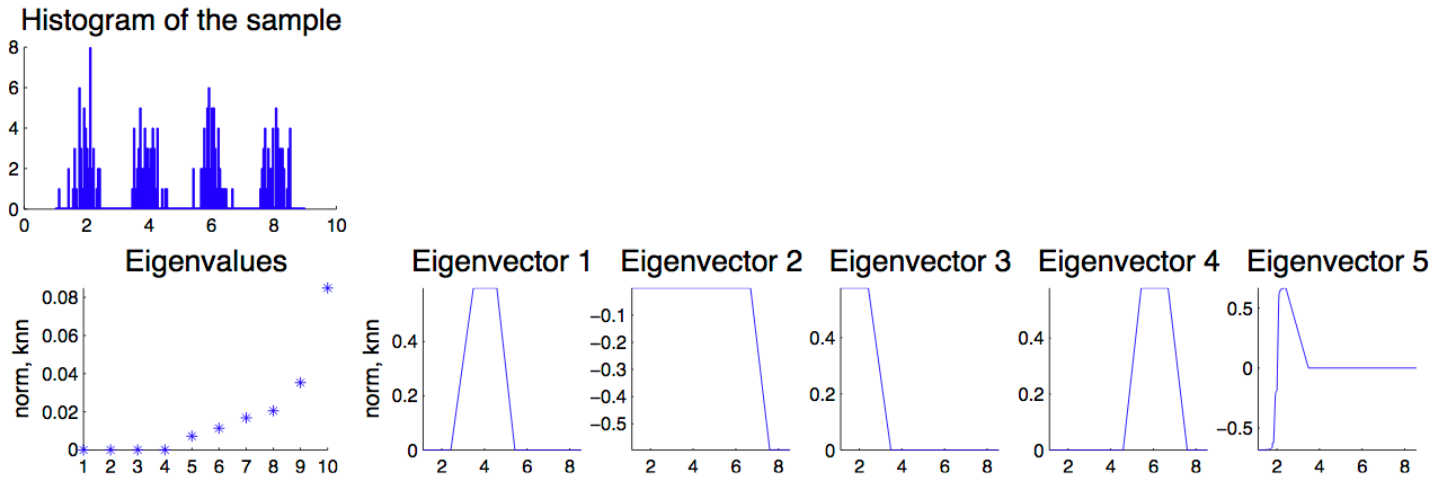


Source: *The Elements of Statistical Learning*, Hastie et al.

DataScience@SMU

Spectral Analysis

Spectral Clustering



- Sample 200 points $x_1, x_2, \ldots . x_{200} \subset R$, forming four clusters (see histogram)
- Define k-nearest graph on those points and take its normalized graph Laplacian, resulting in a 200 x 200 matrix (plot top 10 eigenvalues and top 5 eigenvectors)
- Observations: **1**) four clusters → four zero eigenvalues; **2**) the first four eigenvectors are cluster indicators

If we didn't a priori know there were four clusters, the eigenvalue/vectors would tell us how to group the points into clusters … → main idea of spectral clustering.

Source: *Tutorial on Spectral Clustering*, Ulrike von Luxburg

Spectral Clustering

- Spectral clustering can be sensitive to changes in the similarity graph (e.g., k-nearest neighbor vs. Gaussian distance).
- If data points are at "different scales" (i.e., distances between data points are different in different regions of the space), the similarity graph can be more or less connected depending on the function used.

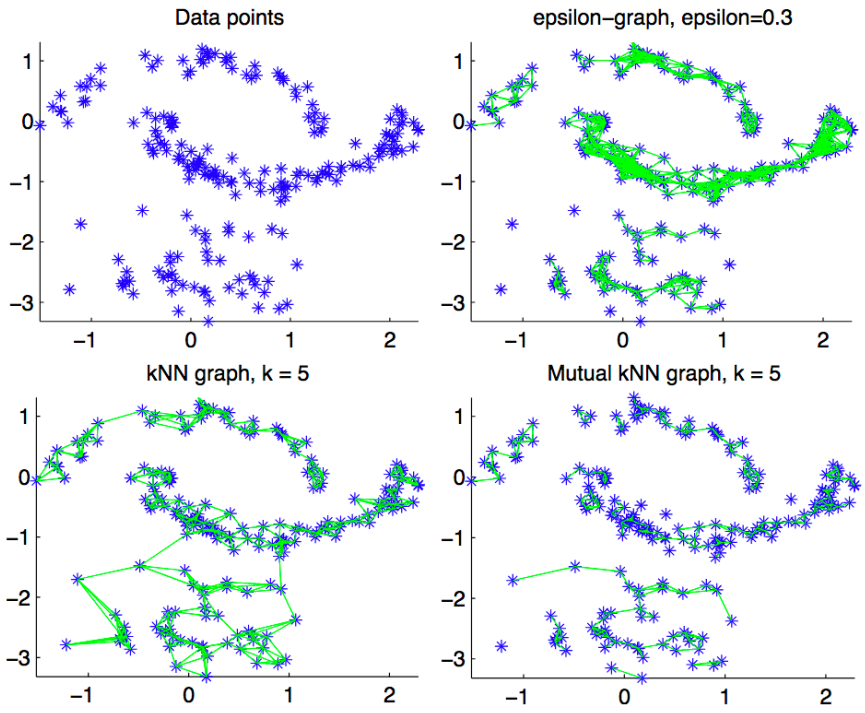


Figure 3: Different similarity graphs, see text for details.

Source: *Tutorial on Spectral Clustering*, Ulrike von Luxburg

DataScience@SMU

Hierarchical Clustering

Hierarchical Clustering

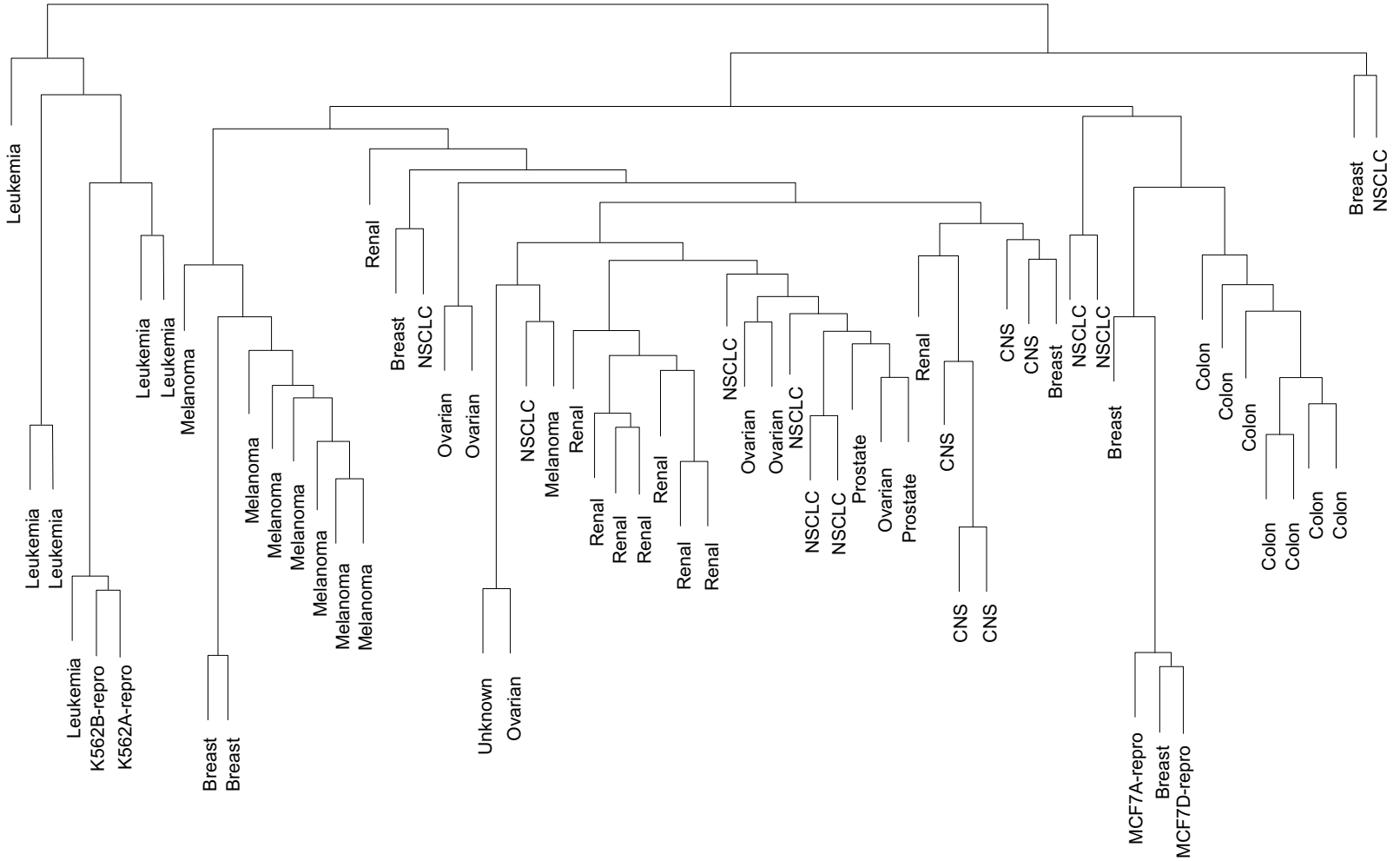


Figure 14.12. *Dendrogram from agglomerative hierarchical clustering with average linkage to the human tumor microarray data.*

Source: *The Elements of Statistical Learning*, Hastie et al.

References

The Elements of Statistical Learning, Hastie et al.

Slides based on work by Hastie et al, Sontag, von Luxburg

DataScience@SMU