Learning Analytics:
A Hands-On Conceptual Introduction

Ruben R. Puentedura, Ph.D.
KNIME
The Basics
Anscombe's Quartet
Finding Groups: k-Means Clustering
Finding Groups: Hierarchical Clustering
Finding Groups: Multidimensional Scaling

\[ \Delta := \begin{pmatrix} \delta_{1,1} & \delta_{1,2} & \cdots & \delta_{1,I} \\ \delta_{2,1} & \delta_{2,2} & \cdots & \delta_{2,I} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{I,1} & \delta_{I,2} & \cdots & \delta_{I,I} \end{pmatrix} \]
Classifying Members: k-Nearest Neighbors
Classifying Members: Decision Tree

- Orange Color?
  - Yellow Color?
    - Lime
    - Lemon
  - Round?
    - Small?
      - Tangerine
      - Clementine
    - Orange
Classifying Members: Naive Bayes

\[
p(C|F_1, \ldots, F_n) = \frac{p(C) p(F_1, \ldots, F_n | C)}{p(F_1, \ldots, F_n)}
\]
Classifying Members: Support-Vector Machines
Modeling: Linear Regression
Classifying and Modeling: Logistic Regression

![Logistic Regression Curve](image)
Gephi
Degree
Paths and Diameter

• Average Path Length: average shortest path between all pairs of nodes

• Network Diameter: longest shortest path between two nodes
Centrality and Eccentricity

- Betweenness Centrality: how often does a node show up on shortest paths between nodes?

- Closeness Centrality: average distance from a given node to all other nodes in the network

- Eccentricity: maximum entry in the shortest path matrix for any given point
Graph Density
Modularity & Connectedness

- Modularity: measure of quality of division of network into modules (communities)
  - High modularity: corresponds to dense connections within modules, sparse connections between modules.
  - Modularity for a given division:
    \[
    \text{fraction of edges that fall within groups} - \text{fraction of edges if dist. at random}
    \]
- Connected component: all nodes in the component are reachable by all other nodes in the same component
  - Strong: direction matters
  - Weak: direction does not matter
Clustering Coefficient

- Measures how close its neighbors are to being a clique (i.e., a complete graph)

- Ratio of number of triangles on a node (3 edges & vertices) to number of triples on a node (3 vertices, 2 edges)
Eigenvector Centrality

- Similar to PageRank

- Connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes

- A node is central to the extent that it is connected to other nodes that are central