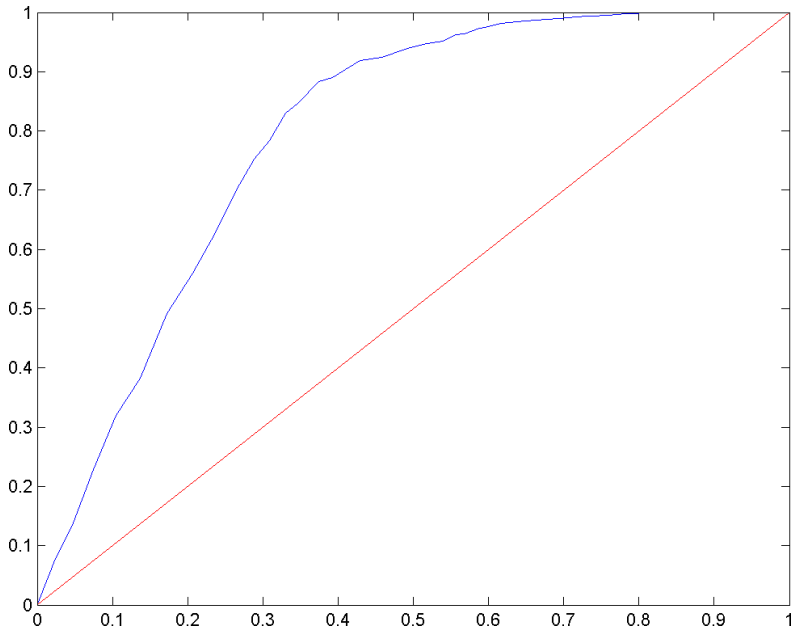
ROC Curves

Miguel Angel Lozano

January 29, 2006

ROC Curves

- Receiver Operating Characteristics curves
- Useful for evaluating the quality of a test
- E.g. a test for matching two points
- Is part of a field called *Signal Dectection Therory*

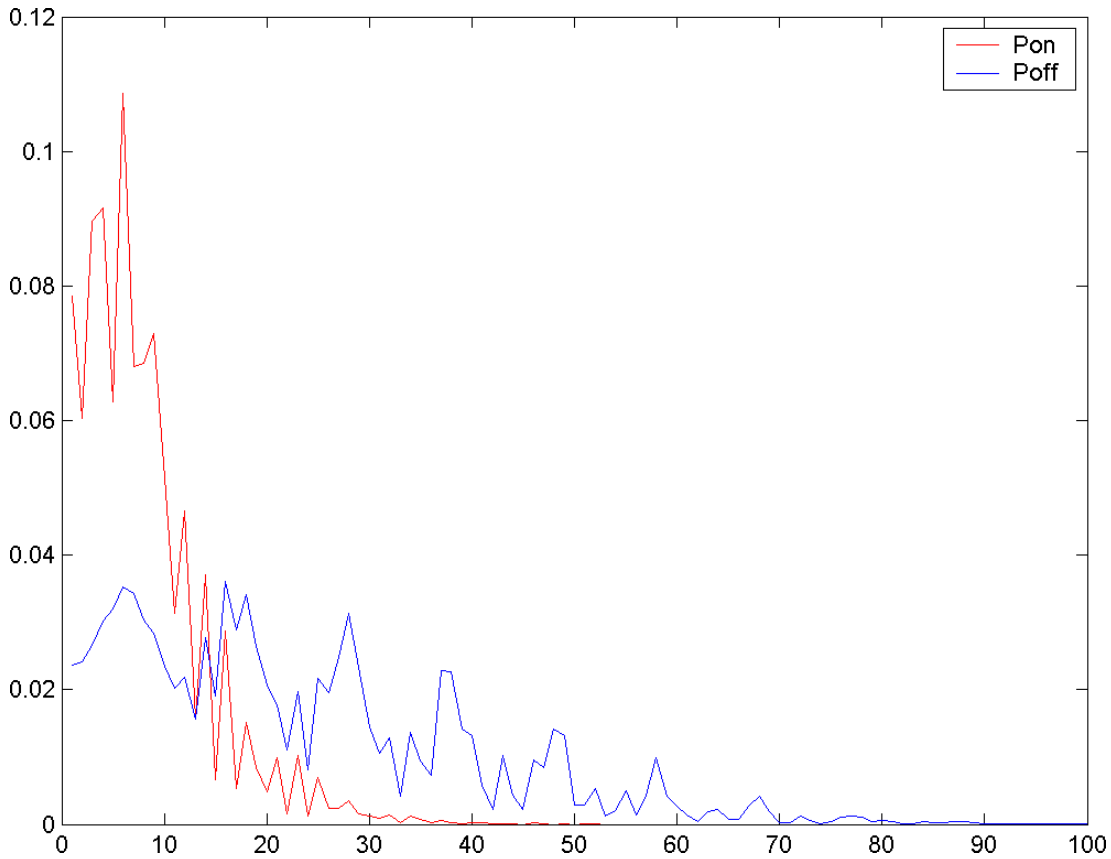


History

- Developed during World War II for the analysis of radar images.
- Radar operators had to decide whether a blip on the screen represented an enemy target, a friendly ship, or just noise.
- Signal detection theory measures the ability of radar receiver operators to decide whether a blip on the screen represents an enemy, a friend, or noise. Their ability to do so was called the Receiver Operating Characteristics.
- It was not until the 1970's that signal detection theory was recognized as useful for interpreting medical test results.

Probability density functions

- Given a continuous value X returned by our test, we can define the following PDFs:
  - $P(X|on)$
  - $P(X|off)$

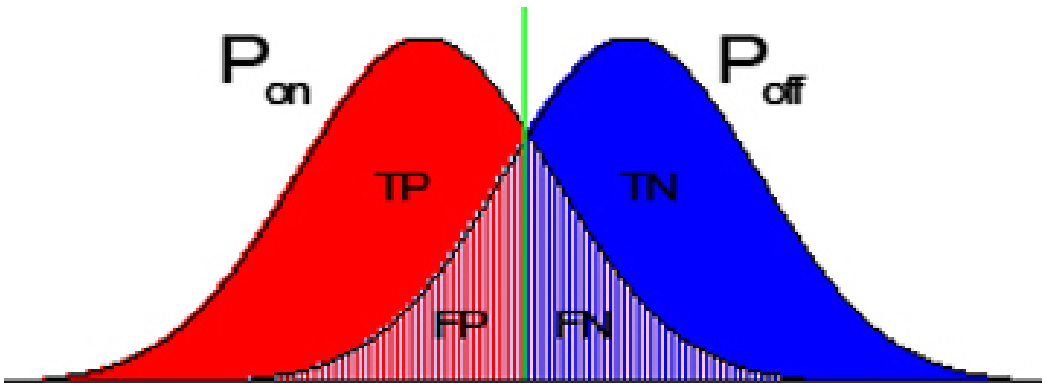


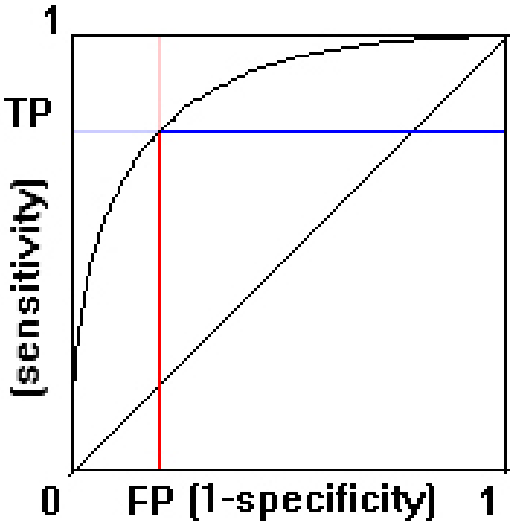
Sensitivity and Specificity

- **TP** - True positive fraction (*sensitivity*)
- **FP** - False positive fraction
- **TN** - True negative fraction (*specificity*)
- **FN** - False negative fraction

$$TP = 1 - FN \tag{1}$$

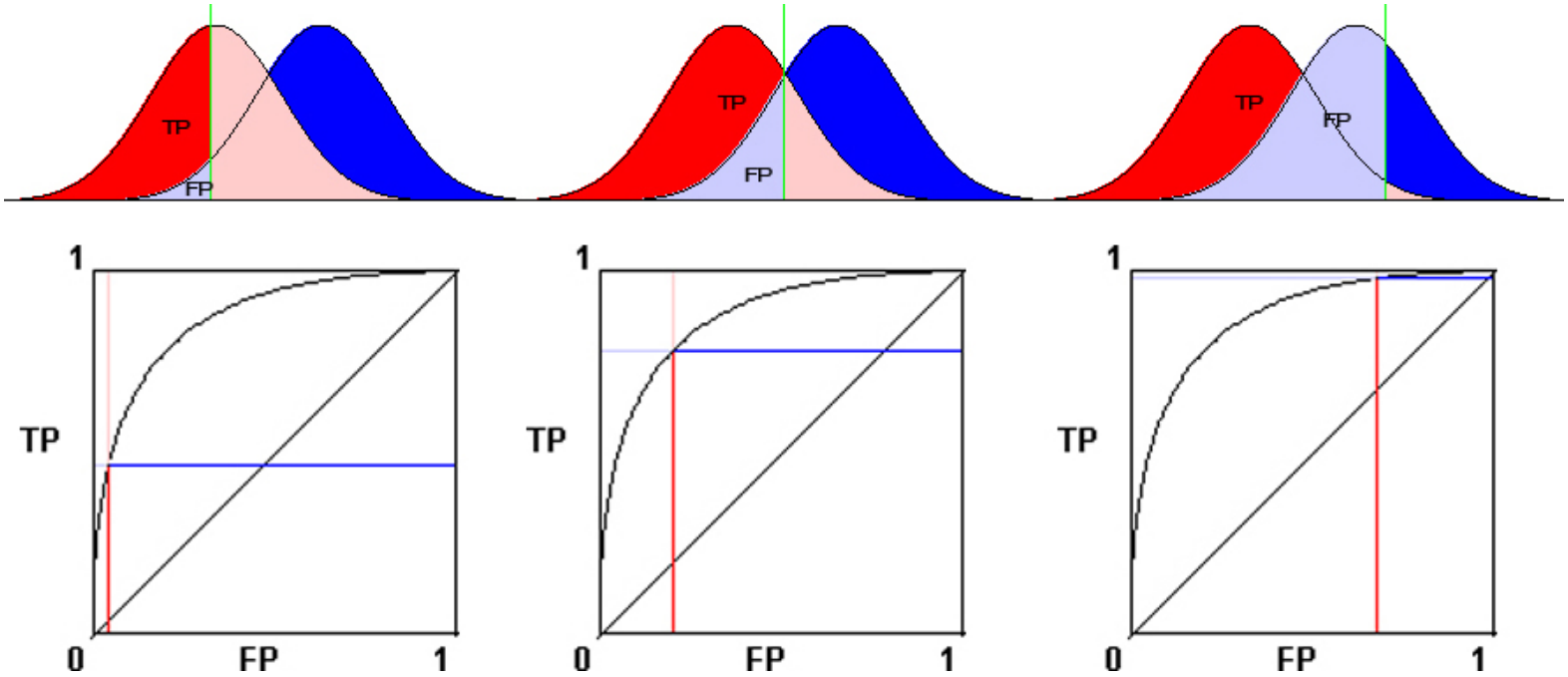
$$TN = 1 - FP \tag{2}$$





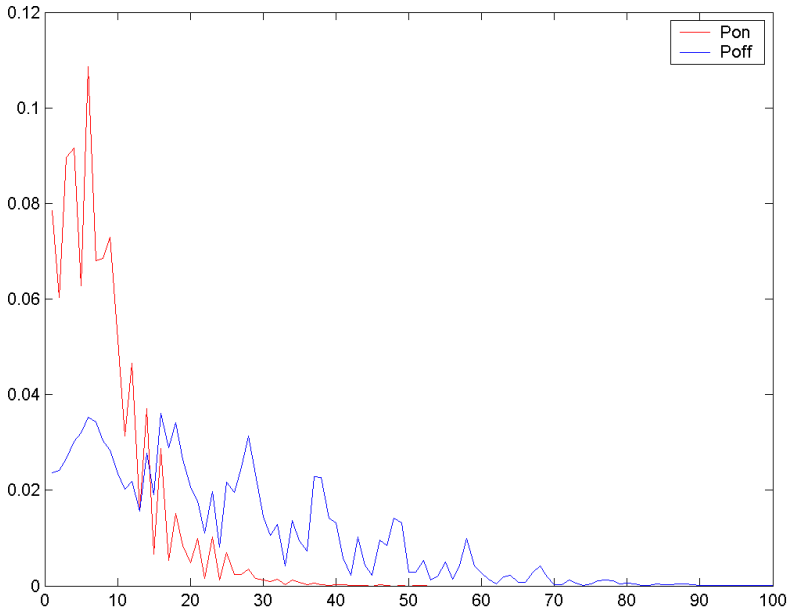
Curve points

- Each point from the curve represents a threshold

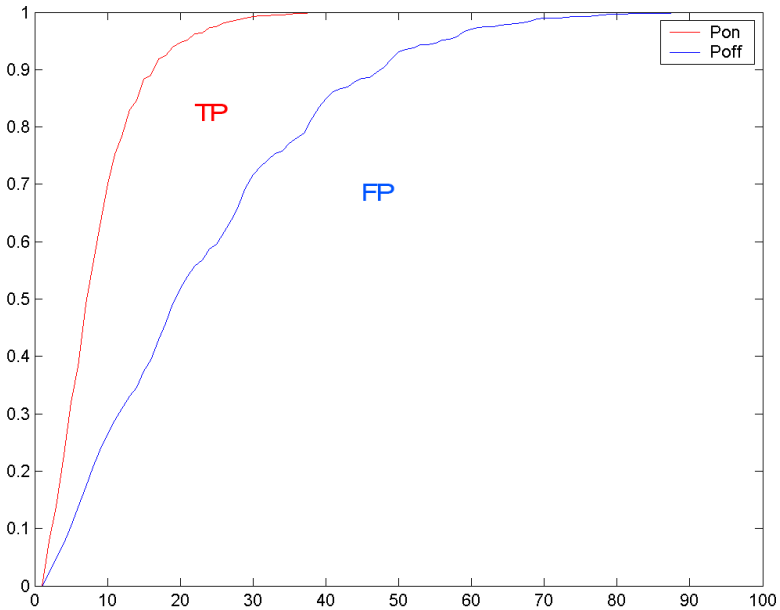


Plotting a curve

- Obtain the CDFs corresponding to PDFs $P(X|on)$ and $P(X|off)$
  - The CDF from $P(X|on)$ corresponds to **TP**
  - The CDF from $P(X|off)$ corresponds to **FP**



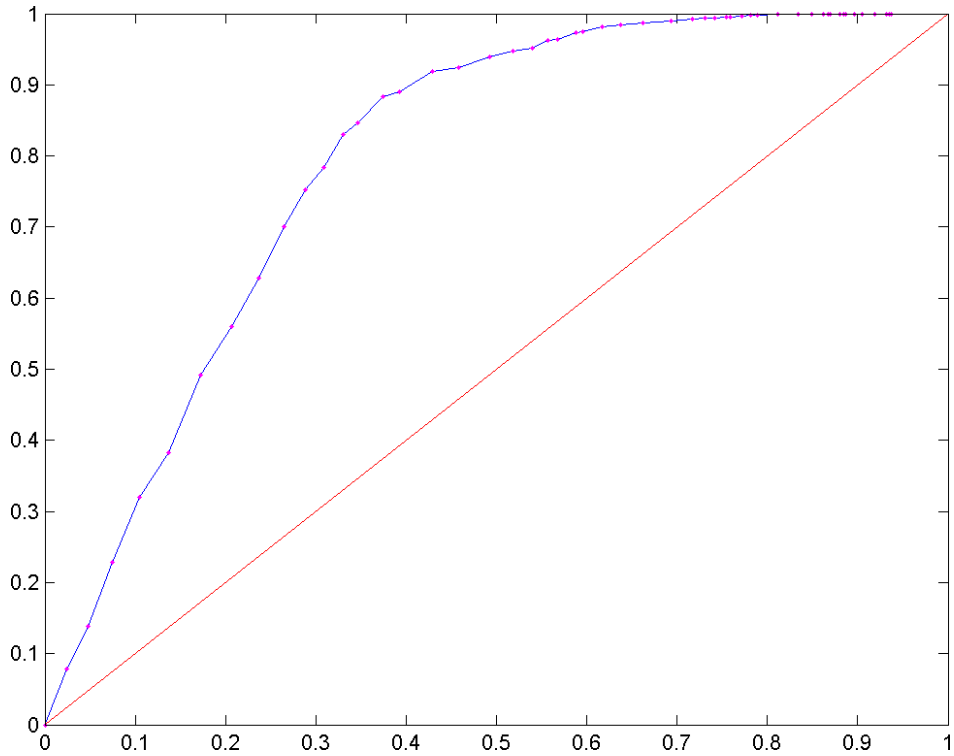
PDFs



CDFs

Plotting a curve

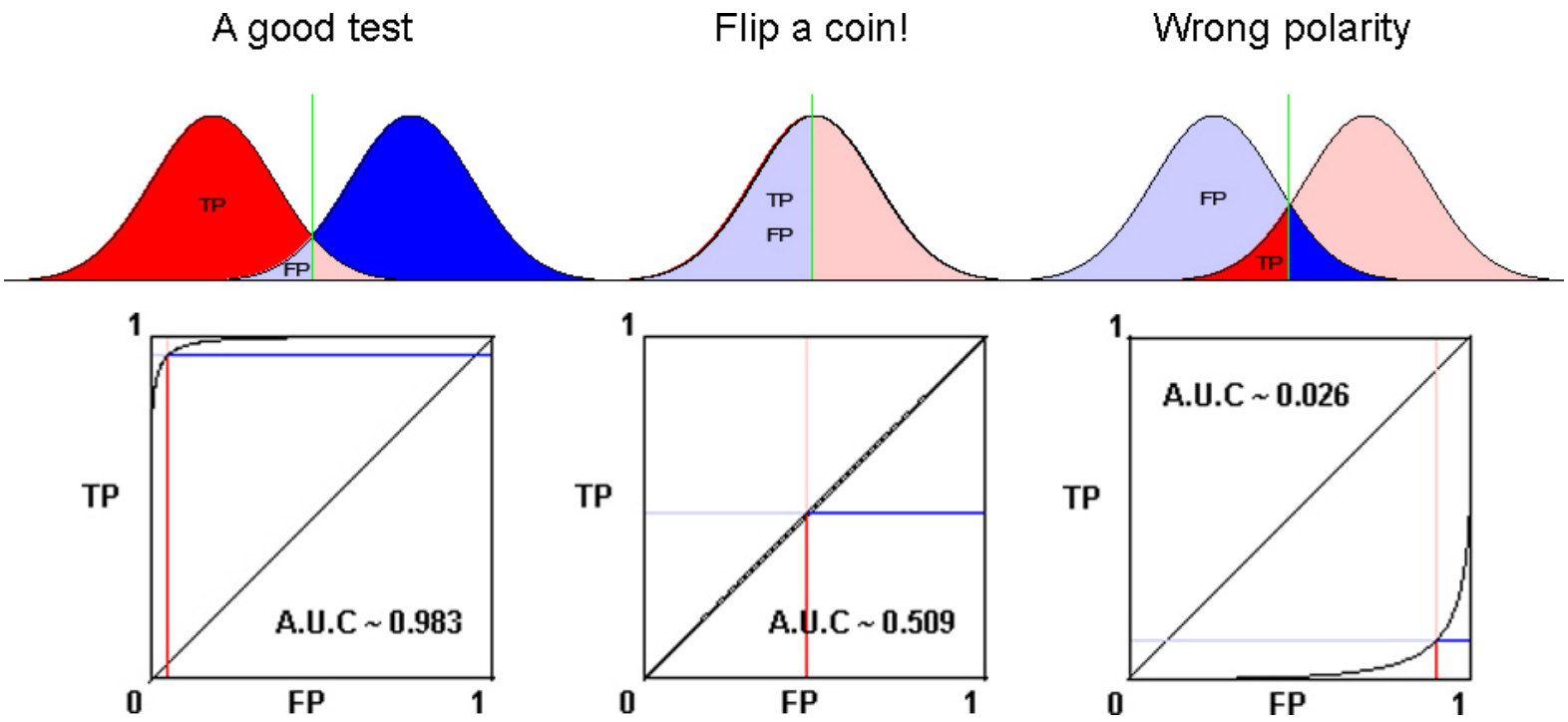
- x coordinates are taken from the $P(X|off)$ CDF (**FP**)
- y coordinates are taken from the $P(X|on)$ CDF (**TP**)



ROC curves are useful for ...

- Measuring test accuracy.
  - Area under the curve (**AUC**).
  - AUC is non-parametrical.
  - Probability of classifying successfully both a positive and a negative example.
- Comparing two tests.
  - Compare the shapes of their corresponding curves.
- Choosing a suitable threshold.
  - A point near to the upper left corner.

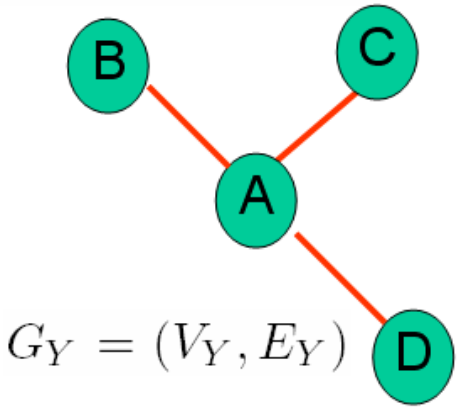
Analyzing results



Adjacency matrices

- Two undirected and unweighted graphs G_x and G_y are defined by their adjacency matrices.

$G_X = (V_X, E_X)$



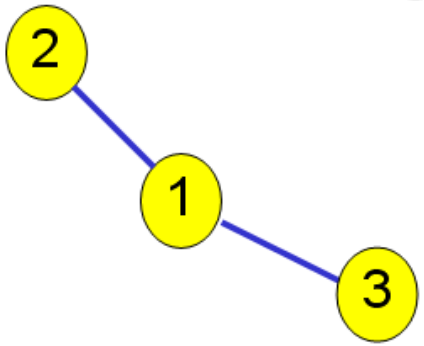
$V_X = \{A, B, C, D\}$

$E_X = \{(A, B), (A, C), (A, D)\}$

$$X_{ab} = \begin{cases} 1 & \text{if } (a, b) \in E_X \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{X} = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

$G_Y = (V_Y, E_Y)$



$V_Y = \{1, 2, 3\}$

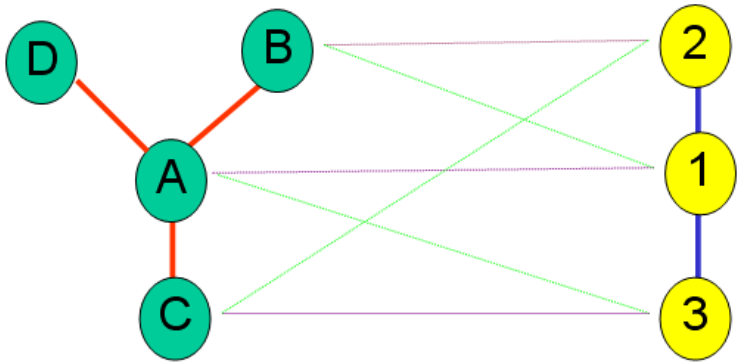
$E_Y = \{(1, 2), (1, 3)\}$

$$Y_{ij} = \begin{cases} 1 & \text{if } (i, j) \in E_Y \\ 0 & \text{otherwise .} \end{cases}$$

$$\mathbf{Y} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

Matching matrices

- Encode the correspondence between vertices in the first graph (rows) and vertices in the second one (columns).
  - The solutions to matching.
- Constraints:
  - A given vertex in the first graph can match either a unique vertex in the second or none of them, and viceversa.



$$\mathbf{M}_1 = \begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \end{array} \qquad \mathbf{M}_2 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

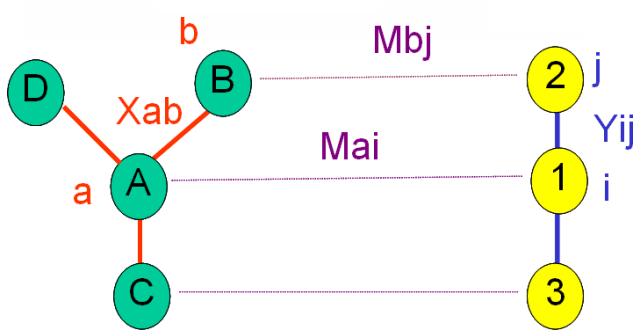
$$\forall\, a \ \sum_{i=1}^{I} M_{ai} \leq 1, \quad \forall\, i \ \sum_{a=1}^{A} M_{ai} \leq 1,$$

$$\forall\, ai \ M_{ai} \in \{0, 1\}$$

Gold & Rangarajan

- Gold & Rangarajan proposed a cost function for quantifying the rectangle rule.

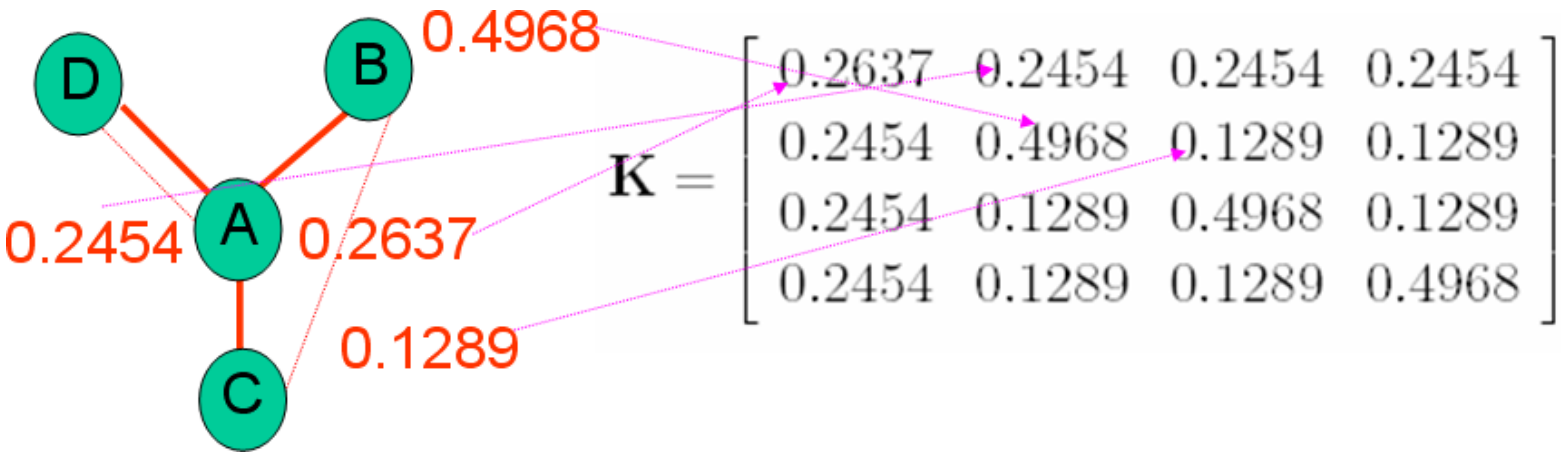
$$F(M) = \sum_{a=1}^{A} \sum_{i=1}^{I} \sum_{b=1}^{A} \sum_{j=1}^{I} M_{ai} M_{bj} \underbrace{X_{ab} Y_{ij}}_{C_{abij}} \tag{3}$$



- G&R proposed a continuation method for maximizing $F(M)$ which updates all assignments iteratively.

Diffusion Kernels

- We use the values from the graph kernel as attributes for non-attributed graphs.



- Probabilistic interpretation:
  - K_{ij} is the probability that a lazy random walk starting at vertex i reaches vertex j. Such a probability decays with the distance between i and j.

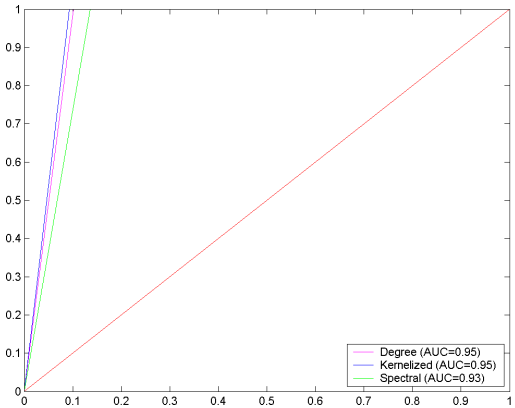
Kernelized cost function

- The G&R cost function is weighted by information coming from distributional attributes:

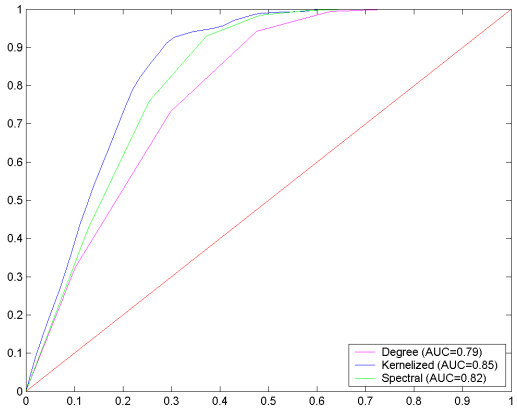
$$F(M) = \sum_{a=1}^{A} \sum_{i=1}^{I} \sum_{b=1}^{A} \sum_{j=1}^{I} M_{ai} M_{bj} C_{abij} \tag{4}$$

$$C_{abij} = X_{ab} Y_{ij} \exp -[(H_a^{K_x} - H_i^{K_y})^2 + (H_b^{K_x} - H_j^{K_y})^2] \tag{5}$$

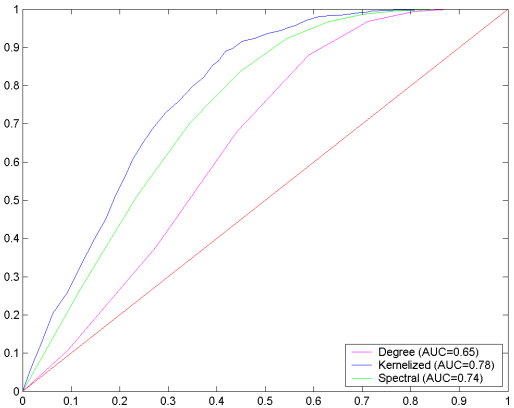
Attribute comparative



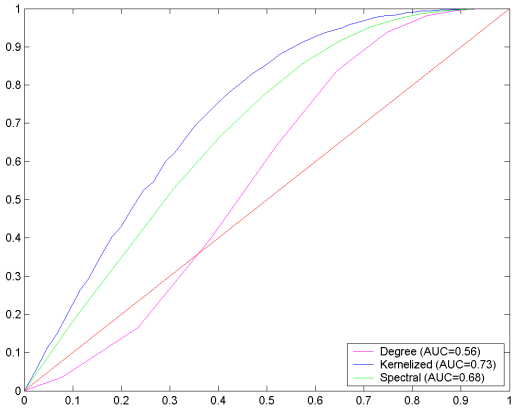
Node noise: 0%



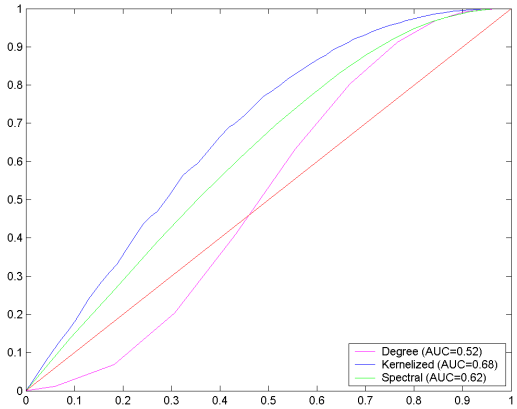
Node noise: 10%



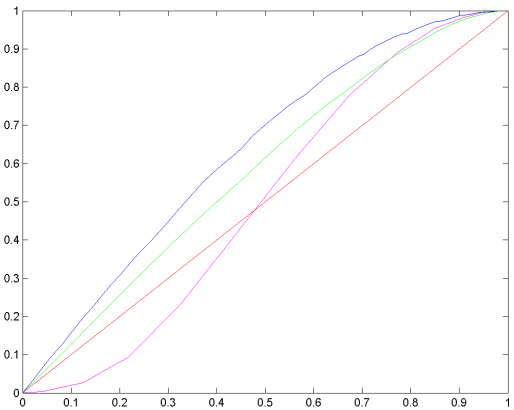
Node noise: 20%



Node noise: 30%



Node noise: 40%



Node noise: 50%