Centralities in graphs

(1) Degree:

$$\vec{C}_d = \frac{\vec{d}}{2E} = \frac{A\vec{1}}{2E}$$

(2) Betweenness centrality:

(3) Closeness:

$$\vec{C}_c(i) = \frac{1}{\frac{1}{N}\sum_j d_{ij}}$$

(4) Eigenvector centrality:

$$\vec{C}_E \qquad \vec{C}_E(i) = \alpha \sum_{j\sim i} A_{ij} \vec{C}_E(j)$$

$$A\vec{C}_E = \lambda_1 \vec{C}_E$$

(5) Page rank.

$$\vec{C}_{PR_{t+1}} = \alpha \left(A D^{-1}\right) \vec{C}_{PR_t} + (1-\alpha) \frac{\vec{1}}{N}$$

$\alpha < 1$

$\alpha = 0.85$

Pagerank is the stationary eigenvector of this process:

$$\vec{C}^{*}_{PR} = \alpha \left(A D^{-1}\right) C^{*}_{PR} + (1-\alpha) \frac{\vec{1}}{N}$$

→ Comparing clusterings:

- Extension of measures we considered when looking at confusion matrices to multiclass clusterings with unequal numbers of clusters.

ARI
AMI

Comparing clusterings: ARI

The contingency table [edit]

Given a set S of n elements, and two groupings or partitions (*e.g.* clusterings) of these elements, namely $X = \{X_1, X_2, \ldots, X_r\}$ and $Y = \{Y_1, Y_2, \ldots, Y_s\}$, the overlap between X and Y can be summarized in a contingency table $[n_{ij}]$ where each entry n_{ij} denotes the number of objects in common between X_i and Y_j : $n_{ij} = |X_i \cap Y_j|$.

$X \backslash Y$	Y_1	Y_2	$\ldots$	Y_s	Sums
X_1	n_{11}	n_{12}	$\ldots$	n_{1s}	a_1
X_2	n_{21}	n_{22}	$\ldots$	n_{2s}	a_2
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
X_r	n_{r1}	n_{r2}	$\ldots$	n_{rs}	a_r
Sums	b_1	b_2	$\ldots$	b_s	

Definition [edit]

The original Adjusted Rand Index using the Permutation Model is

$$\overbrace{ARI}^{\text{Adjusted Index}} = \frac{\sum_{ij}\binom{n_{ij}}{2} - [\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}]/\binom{n}{2}}{\frac{1}{2}[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}] - [\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}]/\binom{n}{2}}$$

where n_{ij}, a_i, b_j are values from the contingency table.

Imperial College London

Comparing clusterings: AMI

$$MI(U,V) = \sum_{i=1}^{R} \sum_{j=1}^{C} P(i,j) \log \frac{P(i,j)}{P(i)P'(j)}$$

where *P(i,j)* denotes the probability that a point belongs to both the cluster U_i in *U* and cluster V_j in *V*:

$$P(i,j) = \frac{|U_i \cap V_j|}{N}$$

Adjustment for chance

$$AMI(U,V) = \frac{MI(U,V) - E\{MI(U,V)\}}{\max\{H(U), H(V)\} - E\{MI(U,V)\}}.$$

where

$$H(U) = -\sum_{i=1}^{R} P(i) \log P(i)$$

$$E\{MI(U,V)\} = \sum_{i=1}^{R} \sum_{j=1}^{C} \sum_{n_{ij}=(a_i+b_j-N)^+}^{\min(a_i,b_j)} \frac{n_{ij}}{N} \log\left(\frac{N \cdot n_{ij}}{a_i b_j}\right) \times \frac{a_i! b_j! (N-a_i)! (N-b_j)!}{N! n_{ij}! (a_i - n_{ij})! (b_j - n_{ij})! (N - a_i - b_j + n_{ij})!}$$