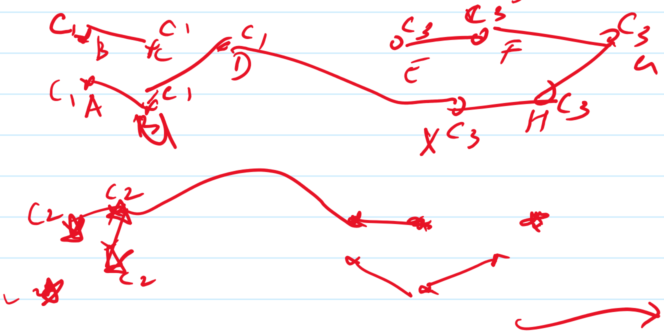
Assortative Measure - Modularity.

17 April 2023 09:57

A n/w w' different types

$c_i \rightarrow$ type of vertex i



$$\begin{array}{c|c} & A \; B \; C \; D \ldots U \\ \hline A & \\ \vdots & \\ U & \end{array}$$

$n_c = 4$

Let c_i be the type for vertex i
Let n_c be the total no. of types

Actual / Expected.

Actual

$$\sum_{(i,j)} \underbrace{\delta(c_i, c_j)}_{\text{Kronecker delta}} = \frac{1}{2} \sum_{ij} A_{ij} \, \delta(c_i, c_j) \rightarrow A$$

if $c_i = c_j$
$\delta(c_i, c_j) = 1$
else
$\delta(c_i, c_j) = 0$

$A - U$
$U - A$ } 2

Example $A - U$ $\quad A_{AU} = 1$
$C_A = C_U$
$\delta(C_A, C_U) = 1$

$U - X$
$C_U \neq C_X$
$\delta(C_U, C_X) = 0$

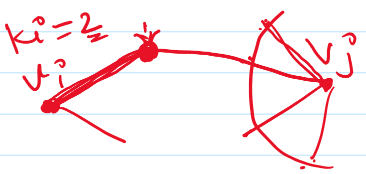
Expected

$A - U$
$U - A$ }

Total no. of edges $= \underline{2m}$

Let v_i be vertex w' degree k_i

Let v_j be another vertex w' degree k_j



$k_i = 2$ $\quad k_j = 4$

$\frac{k_j}{2m} \rightarrow$ because we assume undirected

$\frac{k_j}{2m}$ → undirected

Prob. that one end of ith edge of v_i is connected to an end of the edge of v_j

Prob. that one edge end of v_i is connected to an edge end of $v_j = k_j/2m$

How many edges for $v_i = 2 * k_j/2m$

Total prob. that edges of v_i connected to edge ends of v_j $= k_i * k_j/2m$

Total expected no. of edges betw all pairs of vertices of same type

$$\frac{1}{2} \sum_{i,j} \frac{k_i k_j}{2m} \delta(c_i, c_j) \longrightarrow B$$

Measure of homophily/Assortative Mixing

Actual Expected

A — B formula.

$$\frac{1}{2} \sum_{i,j} A_{ij} \delta(c_i, c_j) - \frac{1}{2} \sum_{i,j} \frac{k_i k_j}{2m} \delta(c_i, c_j)$$

$$= \frac{1}{2} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m}\right] \delta(c_i, c_j)$$

Total no. of edges $= m$.

To get fraction

$$= \frac{\frac{1}{2} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m}\right] \delta(c_i, c_j)}{m}$$

$$Q = \frac{1}{2m} \sum \left(A_{ij} - \frac{k_i k_j}{2m}\right) \delta(c_i, c_j) \rightarrow C$$

Modularity (A measure for assortative mixing)

Modularity → (A measure for assortative mixing)
Z
Unnormalize value

Normalized Modularity $= \dfrac{Q}{Q_{max}}$
Z

Perfectly Mixed N/W : All the vertices are connected to other vertices of same/similar type

ie All edges are b/w vertices of same type

$$Q_{max} = \frac{1}{2}\underbrace{\boxed{\sum A_{ij}\,\delta(c_i, c_j)}}_{2m} - \frac{1}{2}\sum \frac{k_i k_j}{2m}\,\delta(c_i, c_j)$$

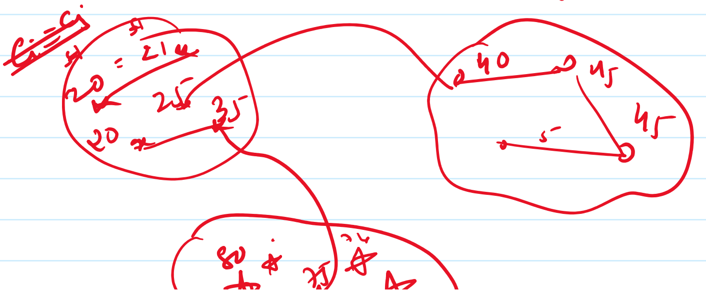
$$= \frac{1}{2}\times 2m - \frac{1}{2}\sum \frac{k_i k_j}{2m}\,\delta(c_i, c_j)$$

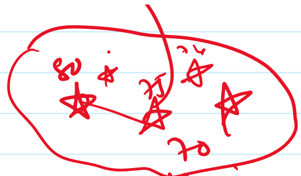
$$= \frac{\frac{1}{2}\left(2m - \sum \frac{k_i k_j}{2m}\,\delta(c_i, c_j)\right)}{m} \rightarrow D$$

$$\text{Normalized } Z = \frac{Q}{Q_{max}} = \frac{C}{D}$$

$$= \frac{\cancel{\tfrac{1}{2}}\sum\left(A_{ij} - \frac{k_i k_j}{2m}\right)\delta(c_i, c_j)}{\cancel{\tfrac{1}{2}}\ 2m - \sum \frac{k_i k_j}{2m}\,\delta(c_i, c_j)}$$

Numerical attributes → age / cgpa / salary
↳ Cont. values.





Variance = Spread of the data.

Mean

$$\begin{bmatrix} 20 \\ 20.5 \\ 21 \\ 22 \\ 23 \end{bmatrix}$$

$20 - \mu$
$20.5 - \mu$

Covariance

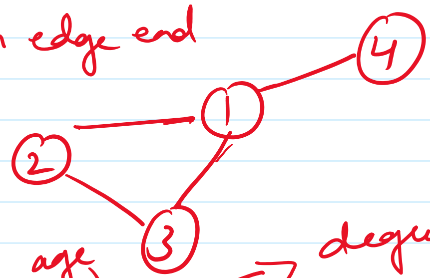
$$Cov(X, Y) = E\left[(X - E[X]) \cdot (Y - E[Y])\right]$$

$$\overset{x}{\begin{bmatrix} \\ \\ \end{bmatrix}} \overset{y}{\begin{bmatrix} \\ \\ \end{bmatrix}}$$

L, R

L → one edge end

R → other edge end



L	R
1 —	2
1 —	3
1 —	4
2 —	3
2 —	1
3 —	1
4 —	1
3 —	2

age → degree of vertex

$L = \{x_1, x_1, x_1, x_2, x_2, x_3, x_4\}$

$R = \{x_2, x_3, x_4, x_3, x_1, x_1, x_2\}$

How many x_1 in $L = 3$
$R = 3$

x_2 $L = 2$
$R = 2$

$\mu_L = \mu_R$ = same.

$\sigma_L = \sigma_R$ → std. dev.

$$Cov(L, R) = \frac{\sum A_{ij}(x_i - \mu_L)(x_j - \mu_R)}{\sum A_{ij}}$$

unnormalized x_i

$$\mu_L = \frac{\sum A_{ij}\, x_i}{\sum A_{ij}} = \frac{\sum x_i \sum A_{ij}}{\sum_i \sum_j A_{ij}} = \boxed{\frac{\sum k_i k_i}{\sum k_i}}$$

numerical value

$\mathrm{Cov}_{max}\}$