Assortativity
Modularity

SYSM 6302

CLASS 13

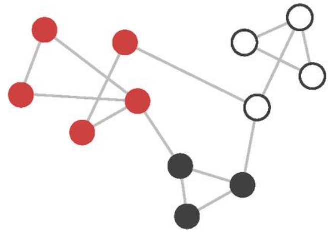
Assortativity measures how "well-mixed" a network is

↑ or lack there of!

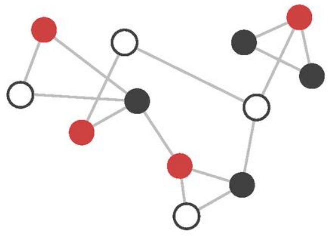
→ Consider: enumerative classes – gender, race, type

or

scalar (real) values – age, income, degree



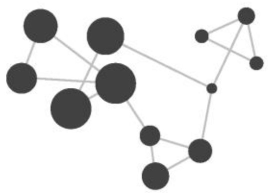
...... random mixing ......



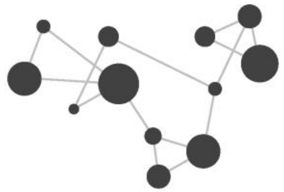
assortive "mixing"

disassortive "mixing"

We use random mixing as our baseline



assortive "mixing"



disassortive "mixing"

Assortativity by Enumerative Characteristics

Modularity = $\underbrace{\text{fraction of edges that connect nodes of the same "type"}}_{\parallel}$ − $\underbrace{\text{fraction of such edges if the edges were positioned at random}}_{\text{random baseline}}$

kronecker delta

$$\underbrace{\frac{1}{2m}\left[\sum_{\substack{\text{edges}\\(i,j)}} \delta(c_i,c_j)\right]}_{\#\text{ of edges}} = \frac{1}{2m}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\,\delta(c_i,c_j)\right]$$

fraction $= \frac{\#}{2m}$

c_i is the class or type of node i → $\delta(c_i,c_j)=1$ if node i & j are of the same class

Assortativity by Enumerative Characteristics

Modularity = fraction of edges that connect nodes of the same "type" − fraction of such edges if the edges were positioned at random

random baseline

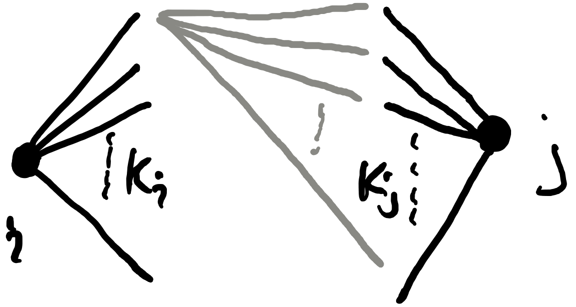
preserving degree (structure)

unfair to compare to a regular graph!

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

expected fraction of edges between nodes i and j (should be ≤ 1)

Counting only nodes of the same type



$k_i k_j$ = # of possible connections between i and j

Assortativity by Enumerative Characteristics

Modularity = fraction of edges that connect nodes of the same "type" − fraction of such edges if the edges were positioned at random

$$Q = \frac{1}{2m} \sum_{i=1}^{n} \sum_{j=1}^{n} \underbrace{\left(A_{ij} - \frac{k_i k_j}{2m} \right)}_{= B_{ij} \text{ Modularity Matrix}} \delta(c_i, c_j)$$

$Q > 0$: Assortative Mixing

$Q < 0$: Disassortative Mixing

$Q = 0$: perfectly random

$Q \leq 1$ since it is a fraction of edges, but most networks can not reach $Q = 1$ even in "best case"

$$Q_{max} = \frac{1}{2m} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\underline{A_{ij}} - \frac{k_i k_j}{2m} \right) \underline{\delta(c_i, c_j)}$$

every edge is between nodes of same type: if $A_{ij} = 1$ then $\delta(c_i, c_j) = 1$

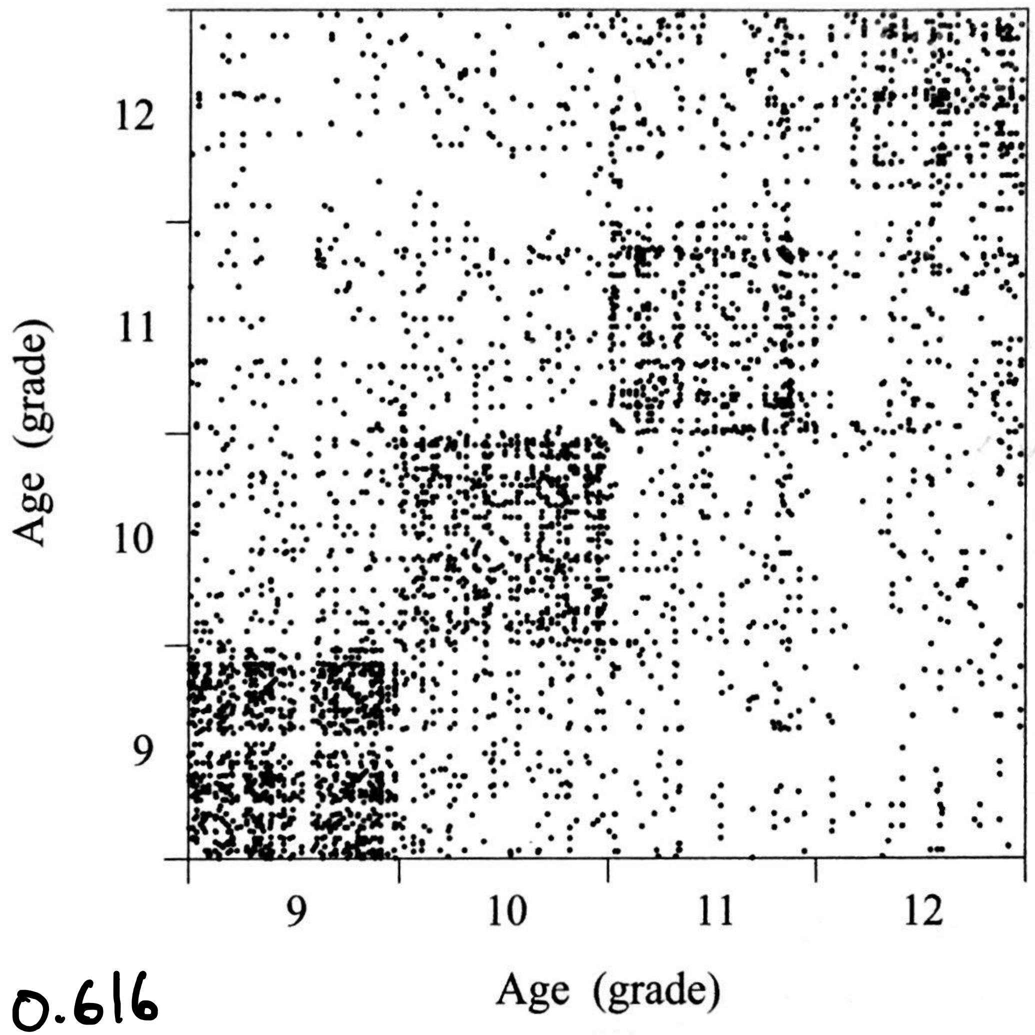
$$\Downarrow$$

$$\sum_i \sum_j A_{ij} \delta(c_i, c_j) = 2m$$

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

Scalar Assortativity

reorder adjacency A in order of scalar characteristic



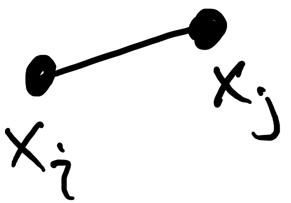
r=0.616

⇒ Could create "bins", but this might separate two people that are very close together

Usually not this clear! Consider income instead!

This looks a lot like covariance

→ A measure of how much two variables vary together away from the mean



The nodes at the end of this edge have an average value of: $\frac{x_i + x_j}{2}$

$$\mu = \frac{1}{2m} \sum_{\substack{\text{edges} \\ (i,j)}} \frac{x_i + x_j}{2} = \frac{1}{4m} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(A_{ij} x_i + A_{ij} x_j \right)$$

mean value of x_i at the end of an edge (some x_i appear at the end of more edges!)

recall: $k_i = \sum_{j=1}^{n} A_{ij}$ (row sum) and $k_j = \sum_{i=1}^{n} A_{ij}$ (column sum)

$$\mu = \frac{1}{4m} \left[\sum_{i=1}^{n} k_i x_i + \sum_{j=1}^{n} k_j x_j \right]$$

$$\mu = \frac{1}{2m} \sum_{i=1}^{n} k_i x_i$$

← weighted average of scalar node values, proportional to their degree

Covariance of x_i and x_j: $\mathrm{cov}(x_i, x_j) = \underline{(x_i - \mu)}\,\underline{(x_j - \mu)}$

how much x_i is different from the mean

how much x_j is different from the mean

"Covariance of the network" = average covariance over all the edges:

$$R = \frac{1}{2m}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\,\mathrm{cov}(x_i, x_j) = \frac{1}{2m}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\underbrace{(x_i-\mu)(x_j-\mu)}$$

$$\mu = \frac{1}{2m}\sum_{\ell=1}^{n} k_\ell x_\ell$$

$$= x_i x_j - \mu x_i - \mu x_j + \mu^2$$

insert effort!

$$= \frac{1}{2m}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij} x_i x_j - \mu^2$$

$$\mu^2 = \frac{1}{4m^2}\left(\sum_{i=1}^{n} k_i x_i\right)\left(\sum_{j=1}^{n} k_j x_j\right)$$

$$= \frac{1}{2m}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(A_{ij} - \frac{k_i k_j}{2m}\right) x_i x_j$$

$\Leftarrow$ Similar to Modularity!

$R > 0$ assortative

$R < 0$ disassortative

Perfect assortativity if $x_i = x_j$ for all edges:

$$R_{max} = \frac{1}{2m} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(A_{ij} x_i^2 - \frac{k_i k_j}{2m} x_i x_j \right) = \frac{1}{2m} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(k_i \delta_{ij} - \frac{k_i k_j}{2m} \right) x_i x_j$$

Assortativity Coefficient: $r = \frac{R}{R_{max}}$

$\Rightarrow$ Assortative Mixing by degree is a specific example of a scalar characteristic

$x_i = k_i$

social networks tend to have positive assortativity by degree