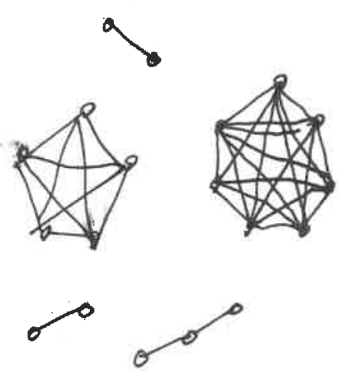
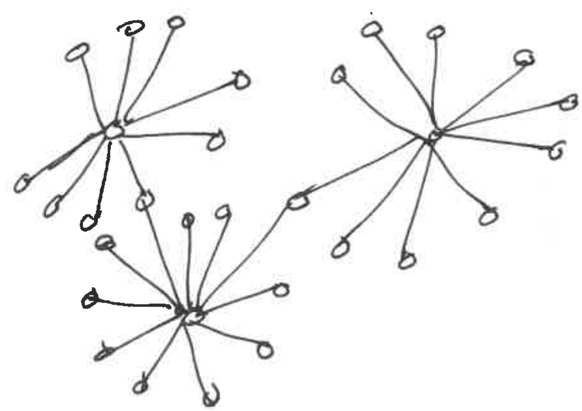
Degree Corellations

- Attempts to answer the question: Does the degree of a vertex affect the degree of its neighbours. Three possibilities
  - <u>assortative</u>: If deg(v) is large, then the degree of v's neighbours is likely to be large.
  - <u>neutral</u>: deg(v) has no effect on the degrees of nodes in N(v).
  - <u>disassortive</u>: If deg(v) is large, then the degree of v's neighbours is likely to be small



assortive



disassortive.

Basic Experiment.

$$n_\ell = |\{v \in V(G) : \deg(v) = \ell\}|.$$
$$d_\ell = \frac{n_\ell}{n}$$

- Remember G is a graph with $|V(G)| = n$ and $|E(G)| = m$.
- Choose a uniformly random edge $e \in E(G)$.
- Choose a uniformly random endpoint v of e (let w be the other endpoint of e).

- This gives a uniform probability space of size $2m$.

$$q_{\ell_1} = \Pr(\deg(v) = \ell_1)$$
$$p(\ell_1, \ell_2) = \Pr(\deg(v) = \ell_1 \text{ and } \deg(w) = \ell_2)$$
$$p(\ell_2 \mid \ell_1) = \Pr(\deg(w) = \ell_2 \mid \deg(v) = \ell_1) = \frac{p(\ell_1, \ell_2)}{q_{\ell_1}}$$

requires $q_{\ell_1} > 0$.

$|\{x \in V(G) : \deg(x) = \ell_1\}|$ $\quad n_{\ell_1}/n$

$$q_{\ell_1} = \frac{\ell_1 \cdot n_{\ell_1}}{2m} = \frac{\ell_1 \cdot n_{\ell_1}/n}{2m/n} = \frac{\ell_1 \cdot d_{\ell_1}}{\text{avg-deg}(G)}.$$

~~G~~ must have at least 1 vertex of degree ℓ_1.

All the information we want is given by the degree correlation matrix:

$$P = \left(p(\ell_1, \ell_2)\right)_{\ell_1, \ell_2 \in [\Delta]} \quad \text{where } \Delta = \max\{\deg(v) : v \in V(G)\}.$$

Uncorrelated Graph: $p(\ell_1,\ell_2) = q_{\ell_1} \cdot q_{\ell_2}$ for each $\ell_1, \ell_2 \in [\Delta]$.

The events "$\deg(v) = \ell_1$" and "$\deg(w) = \ell_2$" are independent.

Notation: $\hat{p}(\ell_1,\ell_2) = q_{\ell_1} \cdot q_{\ell_2}$

$$\hat{p}(\ell_2 | \ell_1) = \frac{\hat{p}(\ell_1,\ell_2)}{q_{\ell_1}} = q_{\ell_2}$$

Uncorrelated graphs correspond to the "neutral" case.

$$\underbrace{K_{nn}(\ell) = \sum_{\ell' \in [\Delta]} \ell' \cdot p(\ell'|\ell) = \mathbb{E}\big(\deg(w) \,\big|\, \deg(v) = \ell\big)}_{K_{nn} : [\Delta] \to \mathbb{R}}$$

In an uncorrelated graph we get

$$\hat{K}_{nn}(\ell) = \sum_{\ell' \in [\Delta]} \ell' \cdot \hat{p}(\ell'|\ell) = \sum_{\ell' \in [\Delta]} \ell' \cdot q_{\ell'} = \sum_{\ell' \in [\Delta]} \ell' \cdot \frac{\ell' n_{\ell'}}{2m}$$

$$= \sum_{\ell' \in [\Delta]} \frac{\ell' \cdot \ell' \cdot n_{\ell'}/n}{2m/n} = \sum_{\ell' \in [\Delta]} \frac{(\ell')^2 \cdot d_{\ell'}}{\text{avg-deg}(G)} = \frac{\mathbb{E}\big((\deg(u))^2\big)}{\mathbb{E}(\deg(u))}$$

when u is a uniformly random vertex in $V(G)$.

Importantly: $\hat{K}_{nn}(\ell)$ does not depend on ℓ.

Recall that $\mathbb{E}(X^2) \geq (\mathbb{E}(X))^2$ and this inequality is strict unless $X(\omega) = \mathbb{E}(X)$ for all $\omega \in S$.

So, in all interesting cases $\mathbb{E}(X^2) > (\mathbb{E}(X))^2$.

Therefore
$$\frac{\mathbb{E}((\deg(u))^2)}{\mathbb{E}(\deg(u))} > \frac{(\mathbb{E}(\deg(u)))^2}{\mathbb{E}(\deg(u))} = \mathbb{E}(\deg(u)).$$

Feld's Friendship Paradox: Most people have fewer friends than their friends have.

<u>Example:</u> $G_{n,p}$. Pick a random u, $E(\deg(u)) = p\cdot(n-1)$.

For each $u' \in N(u)$, $E(\deg(u')) = \overset{\text{counts } u}{1} + p\cdot(n-2) = (1-p) + p(n-1)$
$$= (1-p) + E(\deg(u))$$

For two random variables X and Y, the covariance is

$$\mathrm{Cov}(X,Y) \overset{\text{def}}{=} E\big((X-E(X))(Y-E(Y))\big) = E(X\cdot Y) - E(X)\cdot E(Y).$$

Note that $\mathrm{Cov}(X,X) = E\big((X-E(X))^2\big) = \mathrm{Var}(X)$.

$$\mathrm{Var}(X+Y) = \cancel{E((X+Y)^2 - E(X+Y)}$$

$$E\big((X+Y)^2\big) - \big(E(X+Y)\big)^2$$

$$= E(X^2) + 2\cdot E(XY) + E(Y^2) - (E(X))^2 - E(X)\cdot E(Y) - (E(Y))^2$$

$$= \mathrm{Var}(X) + 2\,\mathrm{Cov}(X,Y) + \mathrm{Var}(Y).$$

Note: If X and Y are independent, then

$$E(X\cdot Y) = E(X)\cdot E(Y)$$

$$E(X-E(X)) = E(X) - E(X) = 0.$$

So, if X and Y are independent then

$$\mathrm{Cov}(X,Y) = E\big((X-E(X))\cdot(Y-E(Y))\big) = E(X-E(X))\cdot E(Y-E(Y)) = 0\cdot 0 = 0.$$

Roughly: $\mathrm{Cov}(X,Y)$ measures how dependent X and Y are.

Pearson's Correlation Coefficient normalizes this into $[-1,1]$:

$$\rho_{X,Y} \overset{\text{def}}{=} \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\cdot\mathrm{Var}(Y)}}$$

If $X = \deg(v)$ and $Y = \deg(w)$, then we can test degree correlations using covariance, and normalize this use Pearson's correlation test.

$$r = \frac{\mathbb{E}(X\cdot Y) - \mathbb{E}(X)\cdot\mathbb{E}(Y)}{\sqrt{\mathrm{Var}(X)\cdot\mathrm{Var}(Y)}} = \frac{\mathbb{E}(X\cdot Y) - \overbrace{(\mathbb{E}(X))^2}^{\text{'Cause } \mathbb{E}(X)=\mathbb{E}(Y)}}{\underbrace{\mathrm{Var}(X)}_{\text{'Cause } \mathrm{Var}(X) = \mathrm{Var}(Y)}}$$

$$\mathbb{E}(X\cdot Y) = \frac{1}{2m}\sum_{e=vw\in E(G)} \left(\deg(v)\cdot\deg(w) + \deg(w)\cdot\deg(v)\right)$$

$$= \frac{1}{m}\sum_{vw\in E(G)} \deg(v)\cdot\deg(w) = \sum_{l_1,l_2\in[\Delta]} l_1\cdot l_2\cdot p(l_1,l_2).$$

Degree Correlation Coefficient:

$$r = \frac{\displaystyle\sum_{l_1,l_2\in[\Delta]} l_1\cdot l_2\cdot p(l_1,l_2) - \overbrace{\left(\sum_{l\in[\Delta]} l\, q_l\right)^2}^{(\mathbb{E}(X))^2}}{\underbrace{\displaystyle\sum_{l\in[\Delta]} l^2 q_l - \left(\sum_{l\in[\Delta]} l\, q_l\right)^2}_{\mathrm{Var}(X)}}$$

If $r > 0$ then G is assortative.
If $r = 0$ then G is neutral
If $r < 0$ then G is disassortative.