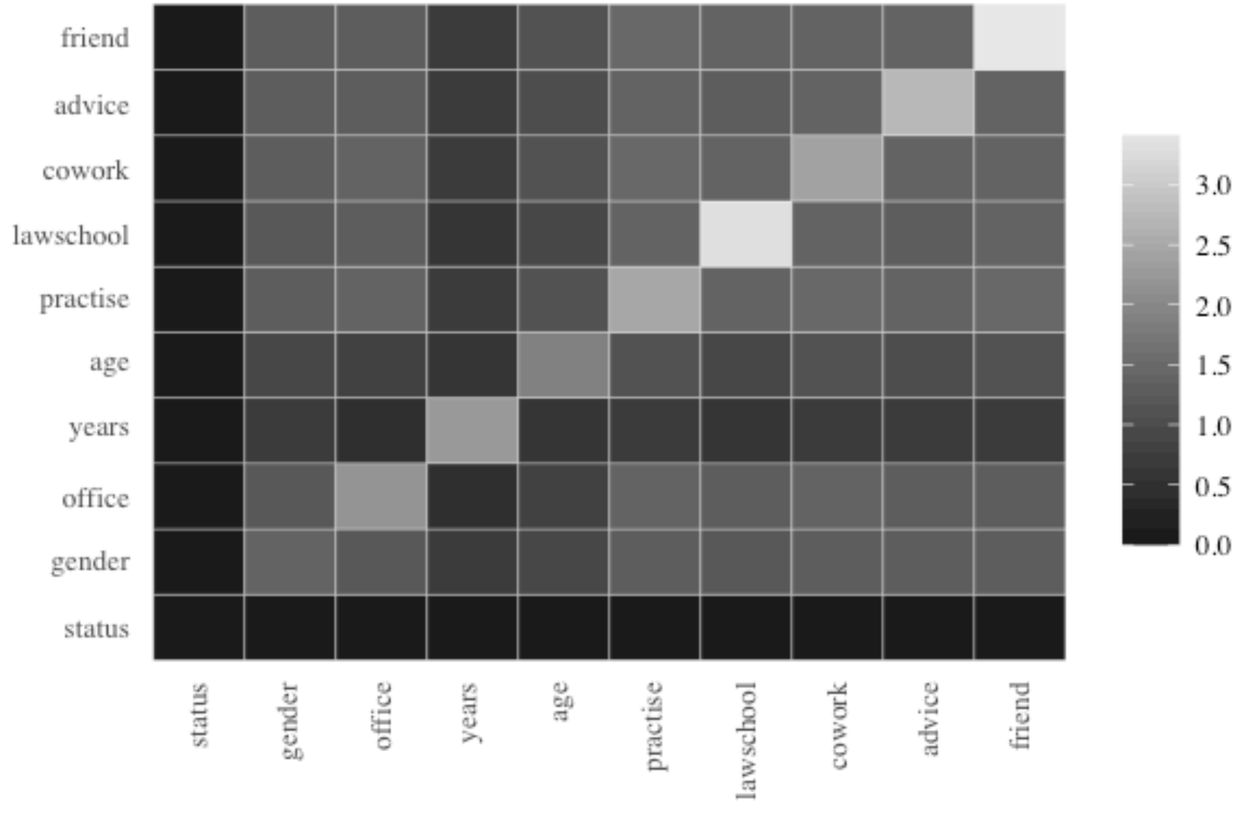
example: network study of corporate law firm

☑ prediction power based one expected conditional entropy $EH(Z|X,Y)$

finding good predictors:
variables (almost) uniquely determined
by combinations of other

```
# prediction power matrix with E(Z|X,Y)
pp <- prediction_power(var, dat)
diag(pp) # single variable prediction EH(Z|X)
```

prediction power visualized using ggplot:



best predictors of 'status':

(years, office)

(age, years)

(lawschool, years)

divergence tests of goodness of fit

goodness of fit tests of hypothetical multivariate discrete distributions
(as suggested by association graphs)

$p =$ general model based on empirical distribution with estimated likelihood function $L(p)$

$q =$ data follows a specified probability model with estimated likelihood function $L(q)$

☑ log likelihood ratio test statistic with d degrees of freedom (for large n)

$$2 \log \frac{L(p)}{L(q)} = 2nD(p,q) \underset{approx}{\sim} \chi^2 \left(d \right)$$

where

$D(p,q)$ is the information divergence (expected log likelihood ratio) with

$d = d(p) - d(q)$ degrees of freedom (numbers of parameters estimated to get p and q)

☑ critical region with approximately 95% confidence level (for large n)

$$\chi^2(d) \geq d + 2\sqrt{2d} = d + \sqrt{8d}$$