trivariate and higher order entropies

for three discrete random variables X, Y and Z the trivariate entropy is bounded

$$H(X,Y) \leq H(X,Y,Z) \leq H(X,Z) + H(Y,Z) - H(Z)$$

the two increments of the inequalities around $H(X,Y,Z)$:

expected joint entropy

$$EJ(X,Y|Z) = H(X,Z) + H(Y,Z) - H(Z) - H(X,Y,Z)$$

non-negative and equal to 0 iff $(X,Y) \perp Z$

association graph
divergence statistic

expected conditional entropy

$$EH(Z|X,Y) = H(X,Y,Z) - H(X,Y)$$

non-negative and equal to 0 iff $X \rightarrow Y$

prediction power

similarily for $H(X,Y,Z,U)$, $H(X,Y,Z,U,V)$, $H(X,Y,Z,U,V,W)$, …

example: network study of corporate law firm

☑ univariate and bivariate entropies of vertex variables

redundancy

```
##          senior status gender office years  age practice lawschool
## senior     6.15   6.15   6.15   6.15  6.15 6.15     6.15      6.15
## status       NA   1.00   1.70   2.08  2.01 2.28     1.98      2.46
## gender       NA     NA   0.82   1.93  2.23 2.38     1.80      2.32
## office       NA     NA     NA   1.12  2.69 2.67     2.09      2.61
## years        NA     NA     NA     NA  1.58 2.75     2.56      3.01
## age          NA     NA     NA     NA    NA 1.58     2.56      2.88
## practice     NA     NA     NA     NA    NA   NA     0.98      2.51
## lawschool    NA     NA     NA     NA    NA   NA       NA      1.53
```

```
# matrix with bivariate entropies
H <- entropy_bivar(dat)
diag(H) # univariate entropies
```

☑ redundant variables

```
##          senior status gender office years age practice lawschool
## senior        0      1      1      1     1   1        1         1
## status        0      0      0      0     0   0        0         0
## gender        0      0      0      0     0   0        0         0
## office        0      0      0      0     0   0        0         0
## years         0      0      0      0     0   0        0         0
## age           0      0      0      0     0   0        0         0
## practice      0      0      0      0     0   0        0         0
## lawschool     0      0      0      0     0   0        0         0
```

```
# detect redundancy
redundancy(dat, dec = 3)
```