

The trouble with  
high-dimensional  
multi-level  
data

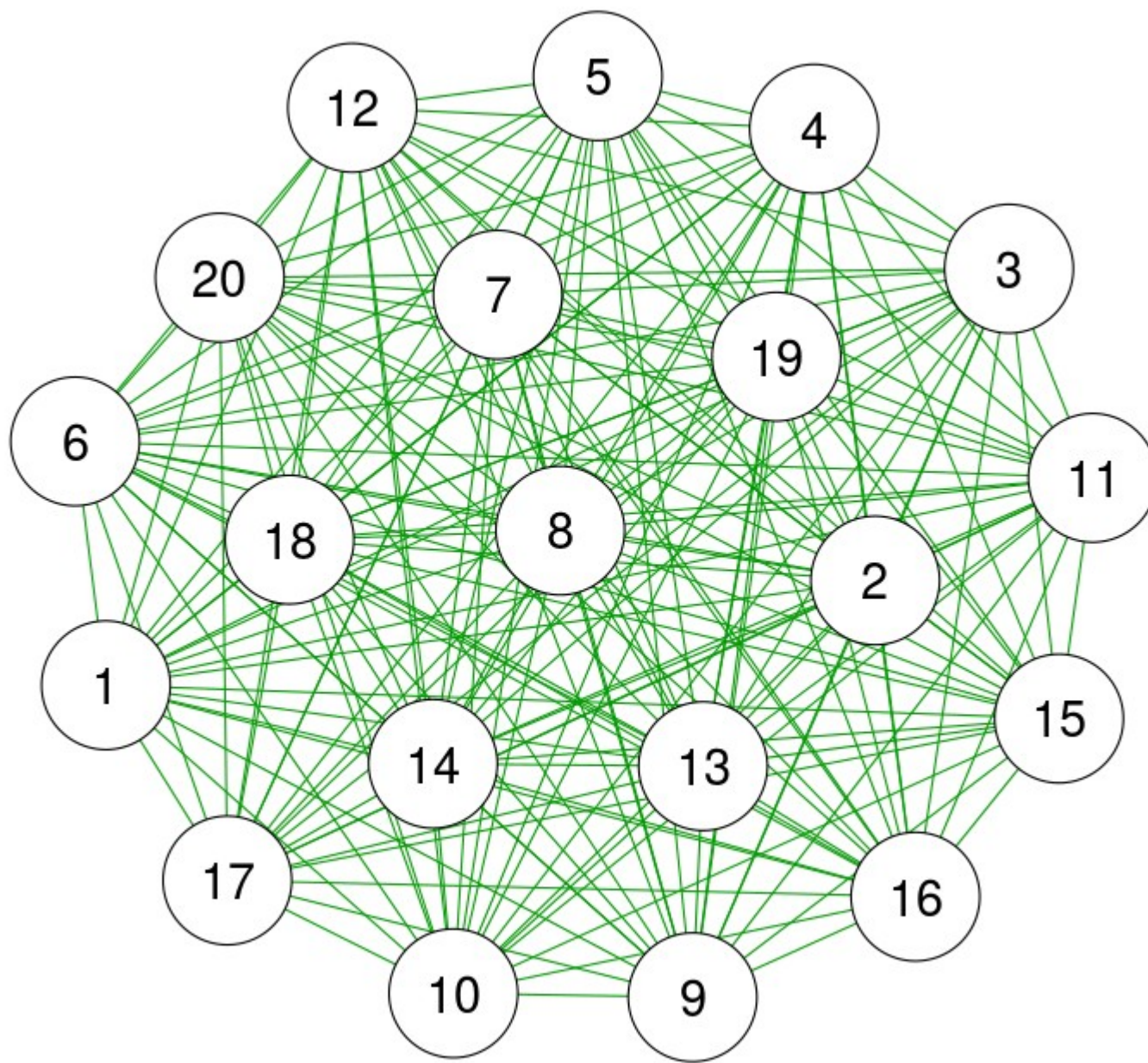
René Mõttus

# The data

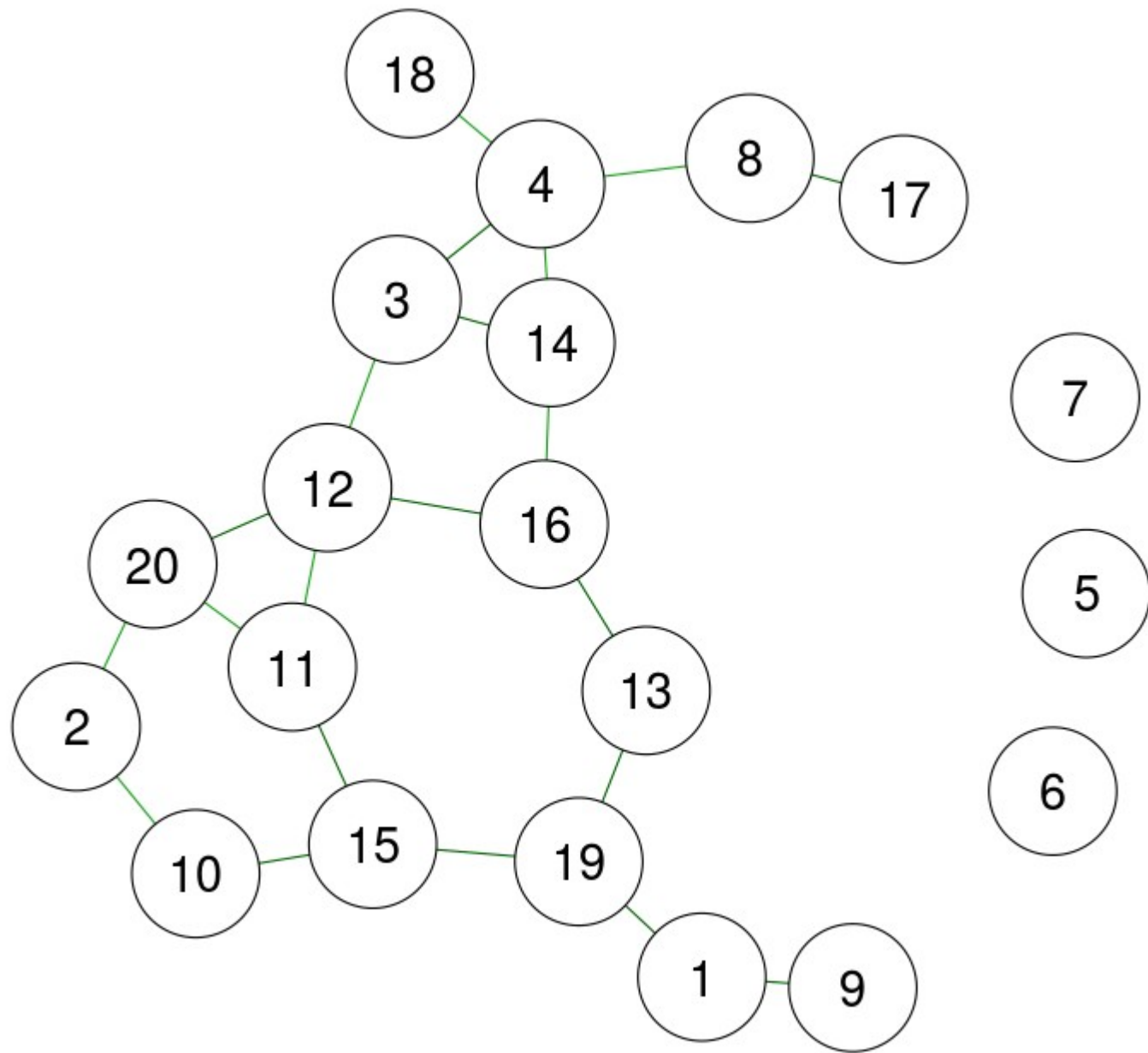
- 'High-dimensional' nested data
  - Say, 20 variables
  - People measured 5 times a day for 10 days
  - Random effects needed
  - Everything correlates with everything
  - Yet all associations ***potentially*** interesting

# The problem

- Univariate associations no go
  - Everything inter-correlating -> uninformative
  - Pervasive spuriousness -> distorted
- We want
  - A sparse pattern
  - Of unique associations



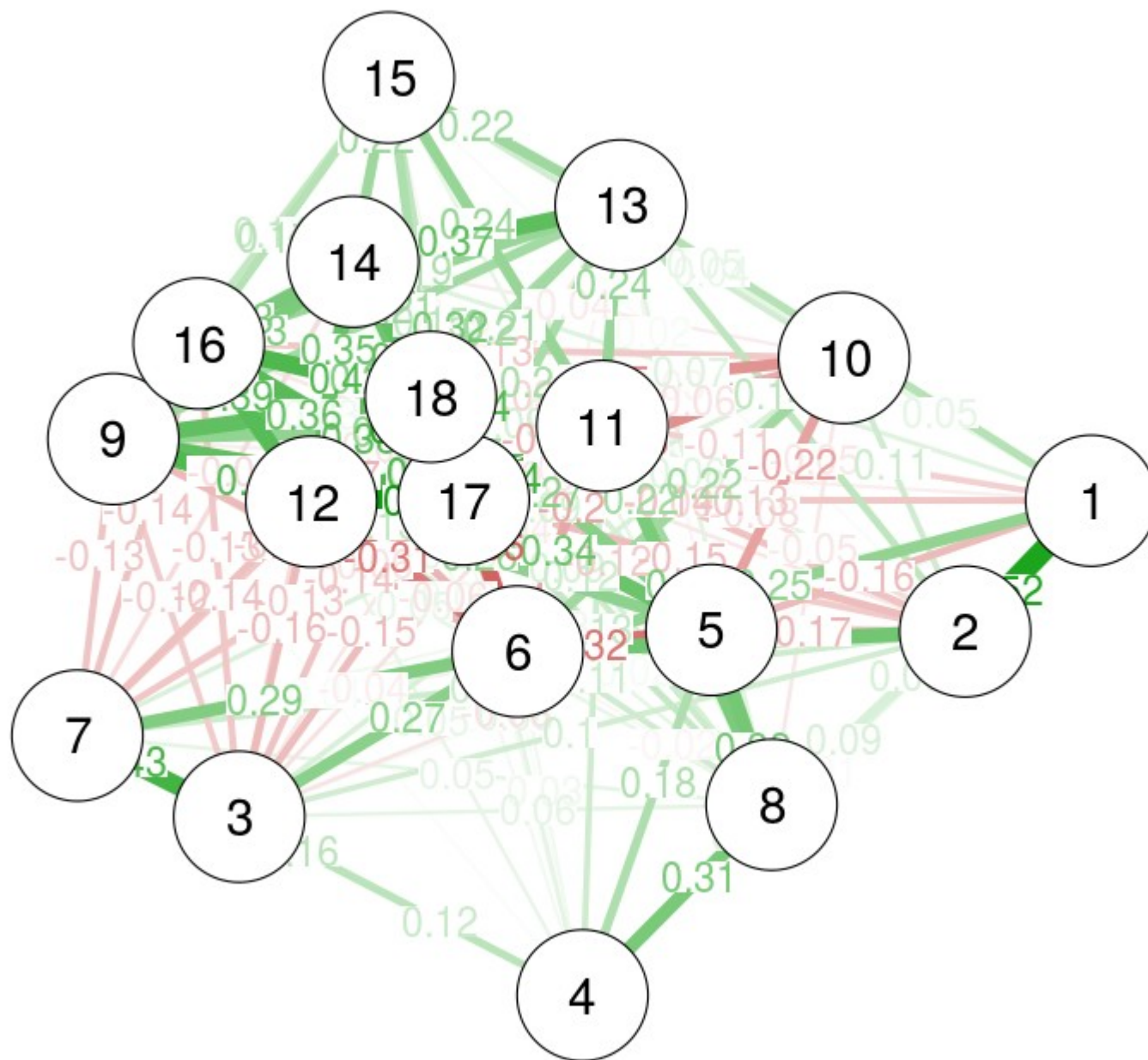
```
> require(qgraph)  
> qgraph(yourMatrix)
```



# Matrix of fixed effects, univariate lmer()

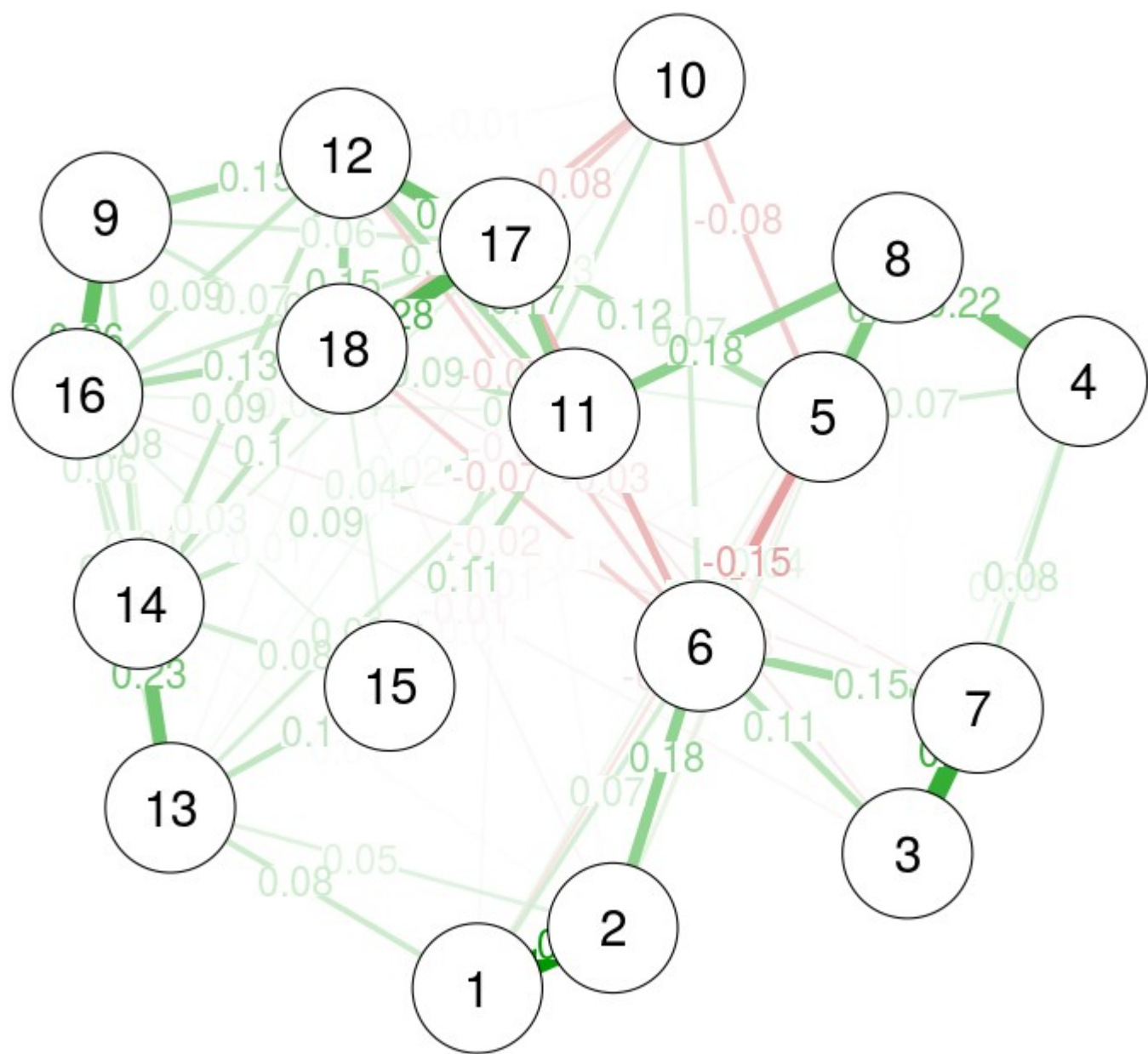
```
## d: 18 personality & situation items plus ID (8,188 observations)
## Standardized within individuals

> k = ncol(d) - 1
> r = matrix(ncol=k, nrow=k, 1)
> for(a in 1:k){
+   for(b in a:k){
+     if(b != a){
+       r[b,a] = fixef(lmer(d[,a] ~ d[,b] + (-1 + d[,b]|ID),d))[-1]
+     }
+   }
+ }
```



```
> r[upper.tri(r)] = t(r)[upper.tri(r)]  
> qqgraph(r, edge.labels=T)
```





```
> qqgraph(r, edge.labels=T, graph="glasso", sampleSize=nrow(Data))
```

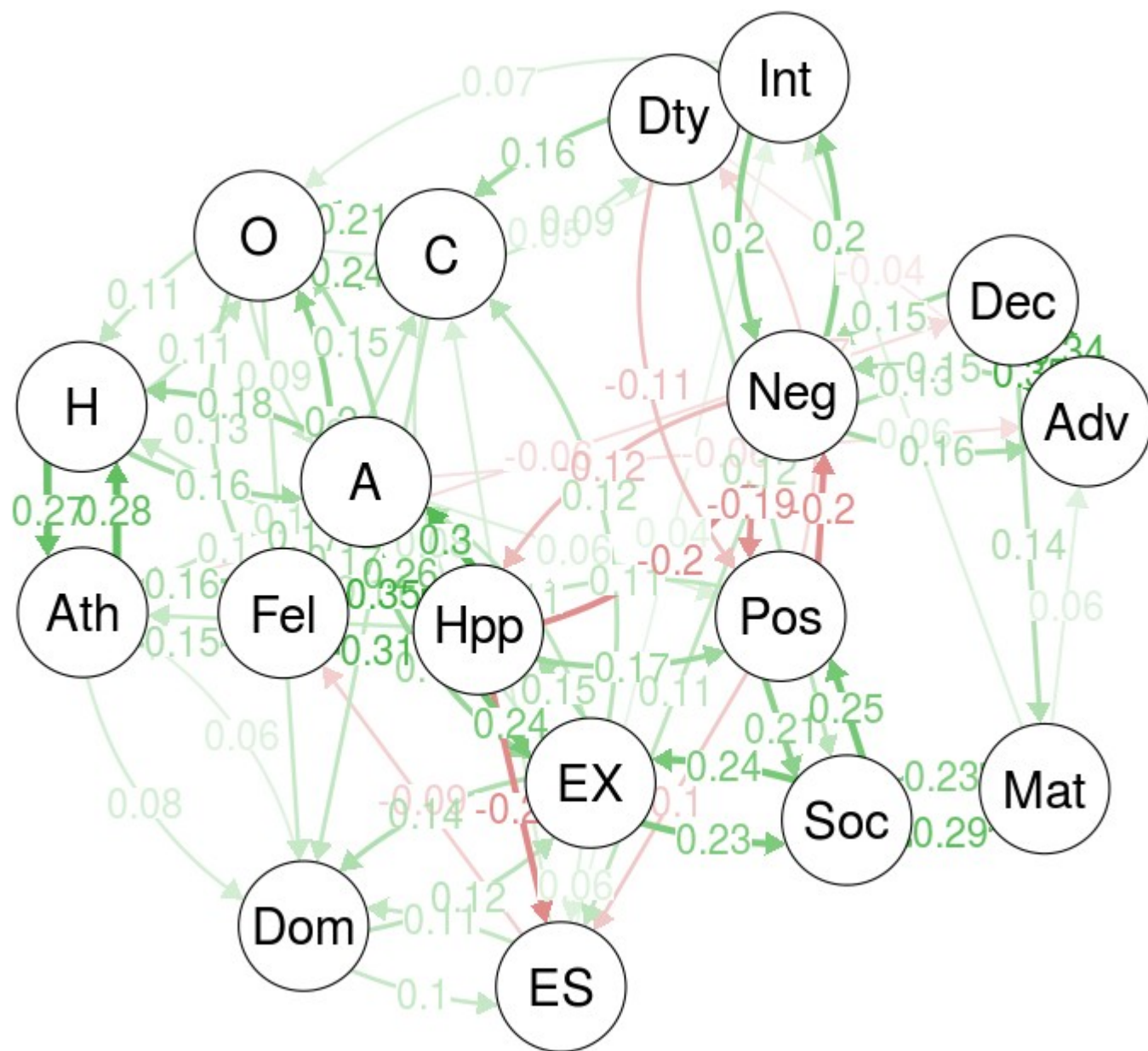


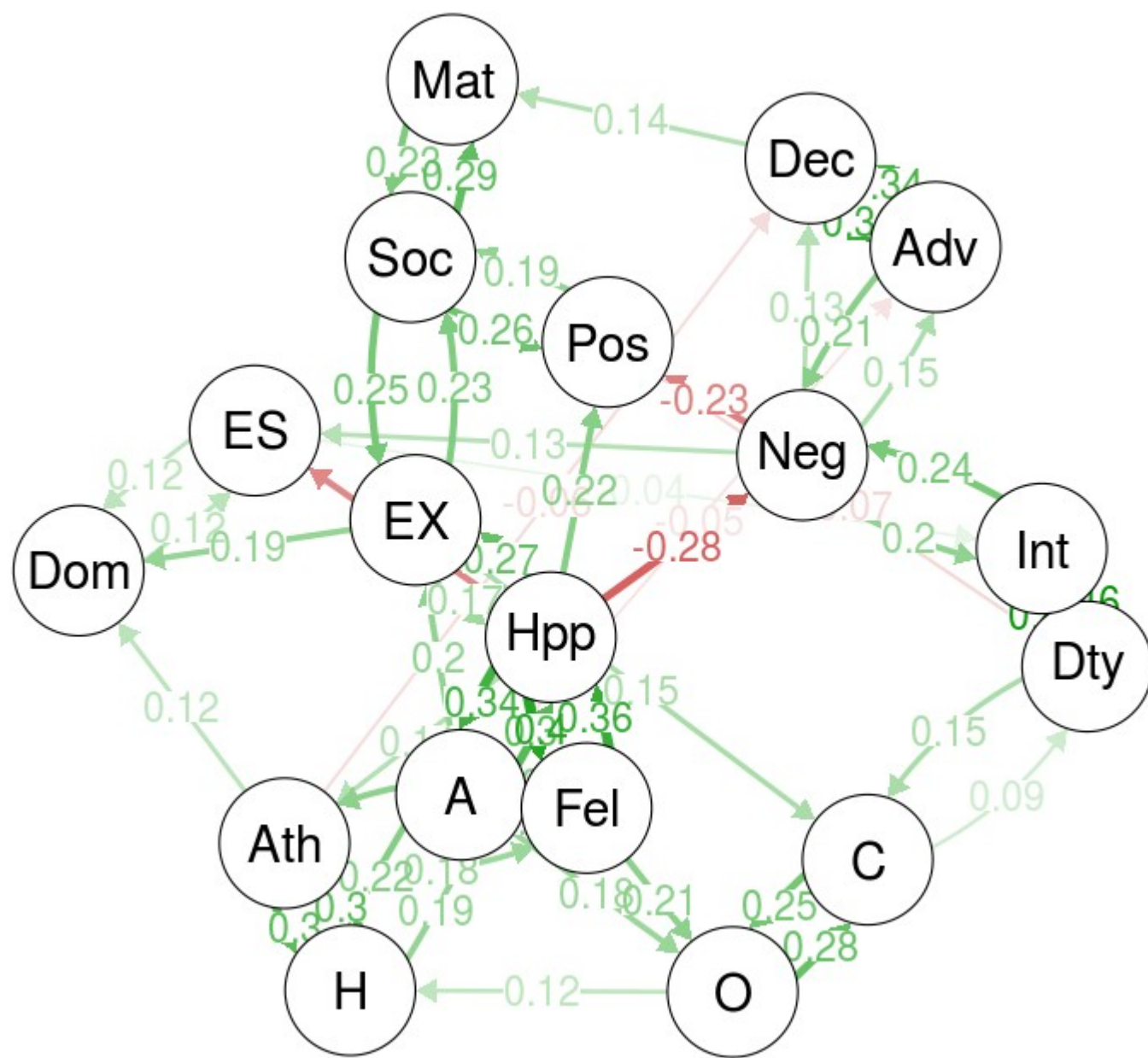
Let every variable be predicted by  
all variables at once

- Hoping that most associations would become very small
  - Convergence issues unless random effects uncorrelated (not realistic)

# Step-wise lmer (really no package for this?!)

- Iterative model selection
  - AIC, BIC, R-squared
- Everything still predicts everything
  - Fit always improves
  - No sparsity, convergence problems
  - Manual limit for predictors (top 3, 5, 7)





Ideas welcome!

- Pulling out **most unique associations**  
**from high-dimensional multi-level data**

`rene.mottus@ed.ac.uk`