Graphic Lasso: Clustering accuracy for precision matrix model

Jiaxin Hu

January 27, 2021

1 Settings

Let $\rho = 2000$ and $r = 5$. The corresponding model is

$$\Omega^k = \Theta_0 + \sum_{l=1}^{r} u_{kl}\Theta_l,$$

where $\boldsymbol{U} = [\![u_{kl}]\!] \in \mathbb{R}^{K\times r}$ has non-overlapped columns. The following are the results after 50 iterations.

2 Convergency Performance

The algorithm did not converge after 50 iterations. The objective function keep decreasing in general, expect the first few iterations. Note that there exist some fluctuations after about 30 iterations. The objective may act like "3.8 - 3.5 - 3.7 - 3.6 - 3.4". The objective function trajectory will be provided later.

Though there are slight fluctuations, the objective function is decreasing. Therefore, the following results may not be satisfactory. The algorithm is running with more iterations now. I will check the new results tomorrow morning.

3 Membership Results

Based on $\hat{\boldsymbol{U}}$, we obtain the membership result for 53 tissues. See Table 1.

The membership result is not very ideal. Table 1 indicates that multiple kinds of tissues are clustered in group A_0 and A_4. This implies the algorithm failed to separate different tissues. On the other hand, parts of the memberships make sense. The group A_3 just include the skins, and the group A_4 mainly include the brain tissues. This also implies the algorithm is potential to distinguish the tissues. Since we only run 50 iterations and the objective function keeps decreasing, the membership results may be improved with more iterations. trajectory of objective function.

How many genes?
preprocessing is often crucial. eg.
could try separate Glasso for representative tissues.
then eyeball the network similarity

4 Gene Correlation Results

See Figure 1, 2, 3, 4, 5 for the correlation estimation of $\{\Theta_l\}, l \in [r]$.
See Figure 6 for the intercept estimation of Θ_0. We only include the top 30 correlations.
In all figures, the wider edge indicates a larger correlation.

Results seem promissing

$A_0 : u_{kl} = 0, l \in [5]$	Minor Salivary Gland, Brain - Cortex Adrenal Gland, Thyroid, Spleen, Pancreas, Colon - Sigmoid, Small Intestine - Terminal Ileum, Prostate, Testis, Nerve - Tibial Heart - Left Ventricle, Pituitary, Brain - Cerebellum, Cells - Transformed fibroblasts, Artery - Aorta, Cells - EBV-transformed lymphocytes, Liver, Kidney - Cortex, Brain - Anterior cingulate cortex (BA24), Brain - Frontal Cortex (BA9) Brain - Cerebellar Hemisphere, Brain - Hypothalamus, Brain - Amygdala, Artery - Coronary, Fallopian Tube, Bladder, Cervix - Ectocervix Cervix - Endocervix, Heart - Atrial Appendage, Ovary, Uterus
$A_1 : u_{k1} \neq 0$	Whole Blood, Breast - Mammary Tissue Adipose - Subcutaneous, Muscle - Skeletal
$A_2 : u_{k2} \neq 0$	Lung, Stomach, Artery - Tibial, Adipose - Visceral (Omentum)
$A_3 : u_{k3} \neq 0$	Skin - Not Sun Exposed (Suprapubic) Skin - Sun Exposed (Lower leg)
$A_4 : u_{k4} \neq 0$	Vagina, Esophagus - Mucosa, Colon - Transverse, Brain - Hippocampus Brain - Substantia nigra, Brain - Caudate (basal ganglia), Brain - Nucleus accumbens (basal ganglia), Brain - Putamen (basal ganglia), Brain - Spinal cord
$A_5 : u_{k5} \neq 0$	Esophagus - Muscularis, Esophagus - Gastroesophageal Junction

these tissues occur only in the intercept
—> does not belong to any group
—> share global network structure

driving by adipose (fat)

mostly skin

mostly brain

Table 1: Membership result for 53 tissues.

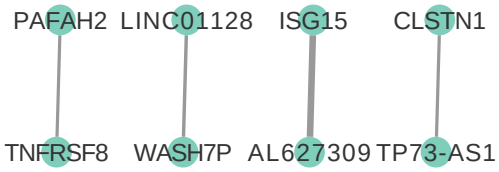

Figure 1: Non-zero correlations in Θ_1.

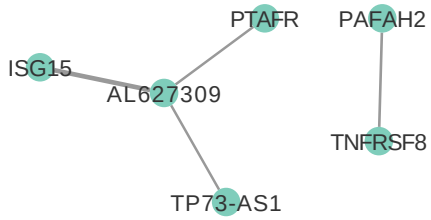

Figure 2: Non-zero correlations in Θ_2.

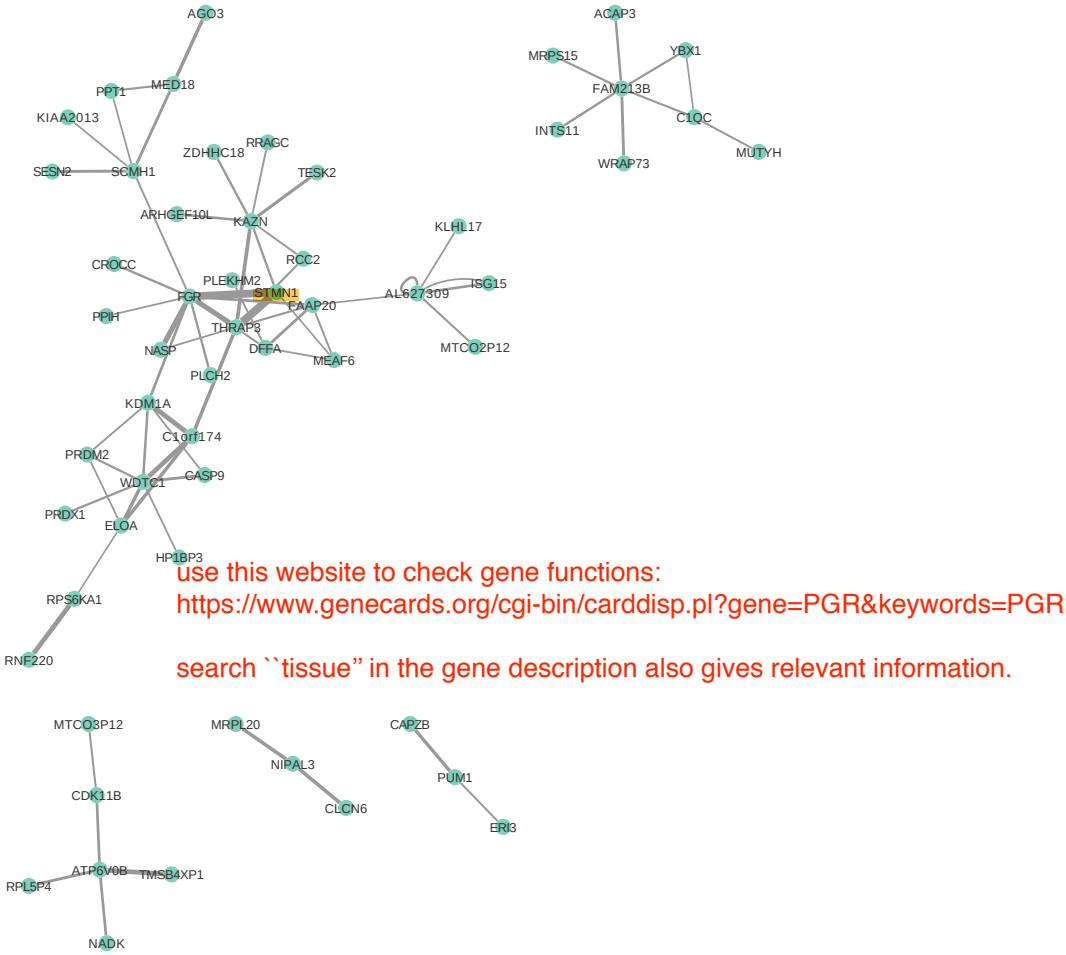

Figure 3: Non-zero correlations in Θ_3.

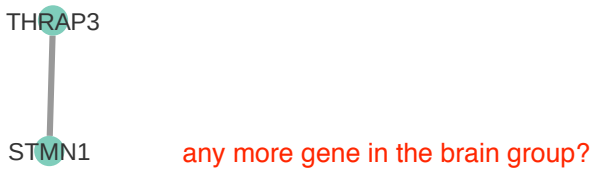

Figure 4: Non-zero correlations in Θ_4.

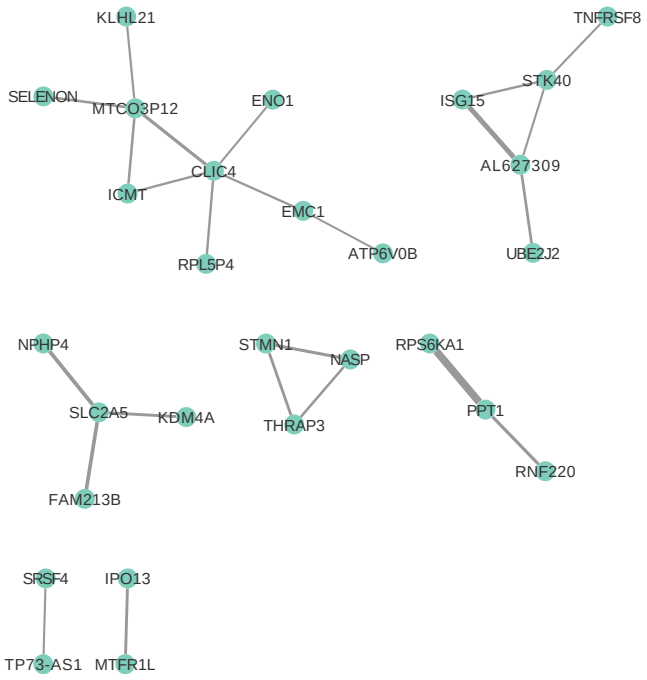

Figure 5: Non-zero correlations in Θ_5.

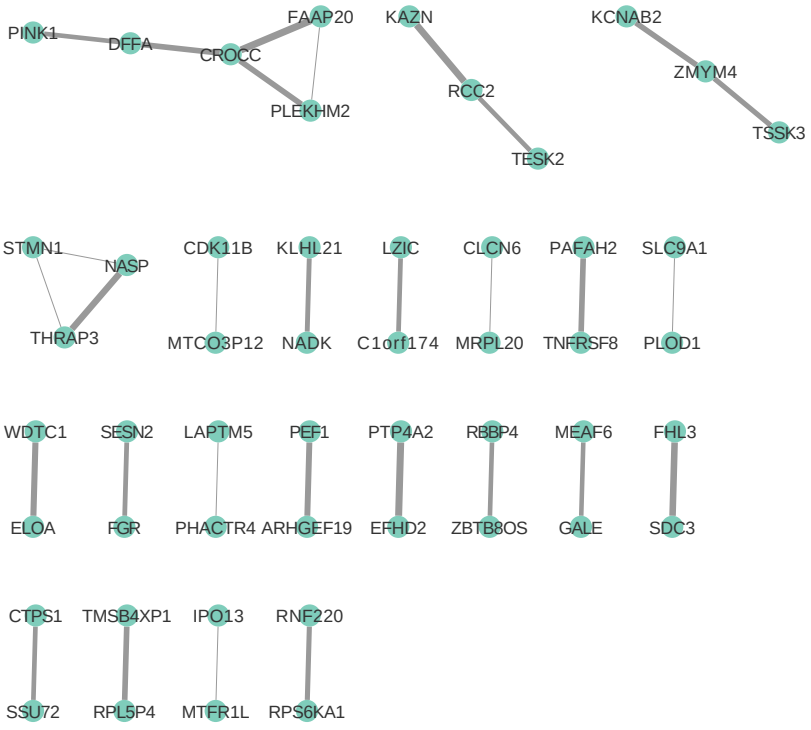

Figure 6: Non-zero correlations in Θ_0.