- Final project assignment has been posted
- One more analytical HW
- Final project proposal due Oct. 4.

Network Inference

Problem statement: Given vertex attributes x_i infer topology of G

- Link prediction - infer links from vertices/existing links
- Association network inference - infer links from vertices
- Tomographic inference - infer links from subset of vertices

Association Networks

- vertices are linked if there is a sufficient level of association
- Suppose $\{x_i\}_{i=1}^{N_v}$ are nodal attributes; then edges based on $\mathrm{sim}(x_i, x_j)$
- Similarity: corr. partial corr. mutual information
- Choice of Inference: Testing, regression, ad hoc methods
- Choice of parameters: thresholds, smoothing parameters, decision rules

Rmk: these choices matter

Correlation Networks: $\rho_{ij} = \dfrac{\sigma_{ij}}{\sqrt{\sigma_{ii}\,\sigma_{jj}}}$, $E = \{(i,j) : \rho_{ij} \neq 0\}$

typically approached as testing problem

$$H_0: \rho_{ij} = 0 \quad H_1: \rho_{ij} \neq 0$$

• Use empirical correlation — typically transformed so to use other null distributions

$$z_{ij} = \frac{\hat{\rho}_{ij}\sqrt{n-2}}{\sqrt{1-\hat{\rho}_{ij}^2}} \qquad z_{ij} = \frac{1}{2}\log\left(\frac{(1+\hat{\rho}_{ij})}{1-\hat{\rho}_{ij}}\right)$$

(T-dist) (Normal) - "Fisher"

Under the Fisher transform $z_{ij} \sim N(0, 1/_{n-3})$ under the null.

However - multiple testing / correlation

Control for multiple testing: - Bonferroni $\alpha_{ij} = \frac{\alpha}{\binom{N_v}{2}}$ so that

family wise error rate is controlled

\# of false discoveries

- False discovery rate $FDR = \mathbb{E}\left[\frac{R_{false}}{R} \middle| R>0\right] \mathbb{P}(R>0)$

\# edges declared

- Benjamini - Hochberg — provides a rule that says FDR is controlled

and we assign $P_{(i)} \leq r \frac{j}{\binom{N_v}{2}}$

- Also a question of the accuracy in p-values

- Question of validity of distributions under the null

- Lots of issues here

<u>Partial Correlation</u>