同濟大學
TONGJI UNIVERSITY

地址：中国上海市四平路1239号　邮编：200092
1239 SIPING ROAD SHANGHAI CHINA 200092
电话（TEL）：+86 21-　传真（FAX）：+86 21-
网址（WEB）：www.tongji.edu.cn

Network Data with power enhancement

1. 改进前提 ① 用于比较两个网络数据的多重检验方法，且不假定数据服从正态分布，只是施加一些矩条件（在network data中易满足），但数据类型特别是 $p\times p$ 矩阵（网络有 p 个 nodes）且样本量十分小，只有数十个，本文比较两总体网络均值之间的差异。我们无法观测到 $X_{d,l}$，只能直接观察到 $S_{d,l}$

② 贡献 (1) 分析的数据格式特别，直接检验网络的links.

(2) 全局统计量易于获得极限分布，但多重检验复杂得多（在我们这么弱的假定下）在Brain Analysis环境下，数据中没有明显的 correlation structure，故本文不讨论任何相依结构下的检验。

(3) 多重检验中，我们提出的power enhancement 有广泛应用。

③ 本文的思想源于GAP方法，都为利用辅助统计量获取稀疏信息，进而分组调权检验，但与之有明显的不同。

(1) PE应用于网络数据，GAP应用于向量数据

(2) GAP要求参数空间 s_1, s_2 都稀疏，PE要求 s_1-s_2 稀疏，如在二元模型中，s_1, s_2 的参数都是非负的。

(3) 依据辅助统计量分组的范围不同，GAP中，C_1, C_2 固定为 $-4, 4$，而PE的可以调节。

(4) 分组意义不同，在GAP中，假定 $K=3$，其假设为 s_1, s_2 都稀疏，故 H_1 (non-null) 一般体现在 G_1 和 G_3 当中，即（负得很多，正得很多），而在PE中，假定的是 s_1-s_2 稀疏，故表现出来为同时小或同时大，故 H_1 (non-null) 一般体现在 G_2.

2. 理论假设.

(1) 矩条件 (A1: Sub-Gaussian tail)，设 $\log p = o(n^{\frac{1}{5}})$，$\exists \eta>0, K>0$，对于 $d=1,2$，有

$$E\{\exp\{\eta (S_{d,l,i,j}-s_{d,i,j})^2/\mathrm{Var}(S_{d,l,i,j})\}\} \le K, \quad 1\le i<j\le p,\ l=1,\cdots,n_d.$$

(A2: Polynomial-tail) 设 $p\le cn^{r_0}$，对某个 r_0 和 $c>0$，$\exists \varepsilon>0, K>0$ s.t. $d=1,2$ 有

$$E\{|(S_{d,l,i,j}-s_{d,i,j})/\mathrm{Var}(S_{d,l,i,j})^{\frac{1}{2}}|^{4r_0+2+\varepsilon}\} \le K, \quad 1\le i<j\le p,\ l=1,\cdots,n_d.$$

NOTE：该矩条件在高维情形中常常满足，且弱于高斯假定

(2) 全局检验的假定 (B1) $C_0^{-1} \le \lambda_{\min}(\mathcal{R}_d) \le \lambda_{\max}(\mathcal{R}_d) \le C_0$，$C_0>0$，$d=1,2$.

(B2) $\max\limits_{d=1,2}\ \max\limits_{1\le k<j\le p} |r_{d,i,j}| < r < 1$，$0<r<1$.

NOTE：(B1)在高维情形中也常见，(B2)很弱，因为如果 $r_{d,i,j}=1$ 则 $\mathcal{R}_d$ 为奇异阵。

(3) 多重检验的假定 (C1) $|S_\rho| \ge (\frac{1}{\sqrt{\pi}\alpha}+\delta)(\log q)^{\frac{1}{2}}$，$q=\frac{p(p-1)}{2}$，$S_\rho=\{(i,j): 1\le i<j<p, \frac{|s_{1,i,j}-s_{2,i,j}|}{\sqrt{\frac{\mathrm{Var}(S_{1,l,i,j})}{n_1}+\frac{\mathrm{Var}(\,)}{n_2}}} \ge (\log q)^{\frac{1}{2}+\rho}\}$

(C2) $\max\limits_{1\le i\le q} |A_i(\xi)| = O(q^{\nu})$，$\xi>0$，$0<\nu<\frac{(1-r)}{1+r}$.

(C3) $q_0=|\mathcal{H}_0| \ge c_1 q$，$c_1>0$.

NOTE：(C1)弱，只要求 s_1-s_2 的标准化差异有很小一部分阶为 $(\log q)^{\frac{1}{2}+\rho}$，(C2)弱，只要求了不能有很多强相关的元素。(C3)自然且弱，否则，$q_0=o(q)$，则我们可以拒绝所有假设就有一个好结果。

同济大学
TONGJI UNIVERSITY

地址：中国上海市四平路1239号　邮编：200092
1239 SIPING ROAD SHANGHAI CHINA 200092
电话（TEL）：+86 21-　传真（FAX）：+86 21-
网址（WEB）：www.tongji.edu.cn

3.方法步骤.

Step 1：初始化

(1)：计算检验和辅助统计量$\{(T_i, A_i), i=1,\cdots,q\}$

(2)：计算p值：$p_i = 2(1-\Phi(|T_i|))$, $i=1,\cdots,q$.

(3)：输入预设参数K, C_1, C_2, N.

(4)：计算格点集 $J=\{\frac{C_1N-1}{N}\sqrt{\log q}, C_1\sqrt{\log q}, \cdots, \frac{C_2N-1}{N}\sqrt{\log q}, C_2\sqrt{\log q}\}$.

Step 2：对每个$J_k=\{\lambda_1,\cdots,\lambda_{K-1}\}$, $\lambda_0=-\infty$, $\lambda_K=+\infty$.

(1)：建立 $G_k=\{i: 1\le i\le q, \lambda_{k-1}<A_i\le\lambda_k\}$. $k=1,\cdots,K$.

(2)：$\forall G_k$, 计算 $q_k=|G_k|$

(3)：$\forall G_k$, 估计non-null比例 $\hat{\pi}_k$,

(4)：计算调整p值：$p_i^w=\min\{\frac{p_i}{w_i}, 1\}$, $i=1,\cdots,q$.

其中 $w_i=\left(\sum_{k=1}^{K}\frac{q_k\hat{\pi}_k}{1-\hat{\pi}_k}\right)^{-1}\frac{q\hat{\pi}_k}{1-\hat{\pi}_k}$, $i\in G_k$, $1\le k\le K$. $\Rightarrow$ $\hat{\pi}_k$小, w_i小, 进而p_i^w大
$\hat{\pi}_k$大, w_i大, 进而p_i^w小.

(5)：对调整后的p值应用BH-procedure, 并记录J_k拒绝数.

Step 3：获得调整拒绝域.

(1)：选取J_k, 使得拒绝数最大

(2)：计算相应J_k的调整p值. p_i^w, $1\le i\le q$.

(3)：对p_i^w排序. $p_{(1)}^w\le\cdots\le p_{(q)}^w$

(4)：输出拒绝域 $\{i: i<\hat{r}\}$. 其中 $\hat{r}=\max\{i: p_{(i)}^w\le\frac{i}{q}\alpha\}$.

4.参数说明：① 分组数K一般取为3. $K\ge4$数值表明power没有显著提升, 但计算成本高

② C_1, C_2可以取为使得 $C_1\sqrt{\log q}$, $C_2\sqrt{\log q}$ 等于最小(大)的辅助统计量A_i. 但在$\pm16\sqrt{\log q}$上做truncate, 即若 $\min A_i\le-16\sqrt{\log q}$, 取$C_1=16$. 若 $\max A_i\ge16\sqrt{\log q}$, 取$C_2=16$

③ N可取为任意整数, N越大, 结果越准确, 但相应的计算成本高

5.理论结果：$FDP_{adj}=\frac{\sum_{i\in H_0}I\{p_i^w\le p_{(\hat{r})}^{(w)}\}}{\sum_{i=1}^{q}I\{p_i^w\le p_{(\hat{r})}^{w}\}\vee1}$, $FDR_{adj}=E(FDP_{adj})$

定理一(FDR有效) 设B2, C1−C3, A_1或A_2成立, $\log q=o(n^{c})$, $c>5$. 则有

$$\lim_{(n_1,n_2,q)\to\infty}\frac{FDR_{adj}}{\alpha q_0/q}=1, \quad \lim_{(n_1,n_2,q)\to\infty}\frac{FDP_{adj}}{\alpha q_0/q}\to1$$

定义算法一(由CAP lemma 3知等价于BH)和PE的power分别为Ψ, Ψ_{adj}. $\Psi=ETP$.

定理二(power有效) 同上条件, 有. $\Psi_{adj}\ge\Psi+o(1)$, $q\to+\infty$.