A network-based penalized regression method with application to genomic data

Sunkyung Kim\textsuperscript{1}, Wei Pan\textsuperscript{1}, Xiaotong Shen\textsuperscript{2}
\textsuperscript{1}Division of Biostatistics, School of Public Health
\textsuperscript{2}School of Statistics
University of Minnesota

Duke University
May 17, 2013
Outline

• Problem

• Review: Existing penalized methods

• New method
  Pan, Xie and Shen (2010, *Biometrics*);
  Luo, Pan and Shen (2012, *Statistics in Biosciences*);
  Kim, Pan and Shen (2013, *Biometrics*);

• Numerical Results: simulated and real data

• Discussion
Introduction

- Problem: linear model

\[ Y = \sum_{i=1}^{p} X_i \beta_i + \epsilon, \quad E(\epsilon) = 0, \quad (1) \]

Feature: large \( p \), small \( n \).

- Q: variable selection; prediction

- Example 1: Li & Li (2008); Pan, Xie & Shen (2010) ...
  \( Y \): clinical outcome, e.g. survival time;
  \( X_i \): expression level of gene \( i \).

- Example 2: eQTL analysis, Lan et al (2003, 2006); Pan (2009) ...

- Typical approaches: ignore any relationships among \( X_i \)'s.

- In our applications: genes are related ...
e.g. as described \textit{a priori} by

1) gene pathways/sets, e.g. KEGG, GO, etc (Ma et al 2007, 2010, ...; Wang et al 2009; Eng et al 2012; ...)

2) a gene network (here):
Figure 1:
• Various types of gene networks: regulatory; co-expression; protein-protein interaction; pathways ...

• **Network assumption/prior 1:** if two genes $i \sim j$ in a network, then $|\beta_i| \approx |\beta_j|$, or $|\beta_i|/w_i \approx |\beta_j|/w_j$.

  Cluster/pathway-based analysis: force/prefer a common $\beta_i$ or $|\beta_i|$ in a group (Park et al 2007; Eng et al 2012)/(Ma et al 2007; ...).

  Q: too strong?

• **Network assumption/prior 2:** if two genes $i \sim j$ in a network, then more likely to have $I(\beta_i \neq 0) = I(\beta_j \neq 0)$.

• Goal: utilize the network assumption/prior 2.

• How?
Review: Existing Methods

• Penalized methods: for “large $p$, small $n$”

$$\hat{\beta} = \arg \min_{\beta} L(\beta) + p\lambda(\beta),$$

• Lasso (Tibshirani 1996):

$$p\lambda(\beta) = \lambda \sum_{k=1}^{p} |\beta_k|.$$  

Feature: variable selection; some $\hat{\beta}_k = 0$.

• Elastic net (Zou and Hastie 2005)

$$p\lambda(\beta) = \lambda \sum_{k=1}^{p} |\beta_k| + \lambda_2 \sum_{k=1}^{p} \beta_k^2.$$  

But ...
• A network-based penalty of Li and Li (2008): **Grace**

\[ p_\lambda(\beta) = \lambda_1 \sum_{i=1}^{p} |\beta_i| + \lambda_2 \sum_{i \sim j} \left( \frac{\beta_i}{\sqrt{d_i}} - \frac{\beta_j}{\sqrt{d_j}} \right)^2, \quad (2) \]

\( d_i \): degree of node \( i \); two terms for diff purposes ...

Related: Huang et al (2011); Ma et al (2012);
Problem: if \( \beta_i \) and \( \beta_j \) have diff signs ...

• A modification by Li and Li (2010): **aGrace**

\[ p_\lambda(\beta) = \lambda_1 \sum_{i=1}^{p} |\beta_i| + \lambda_2 \sum_{i \sim j} \left( \frac{\text{sgn}(\tilde{\beta}_i)\beta_i}{\sqrt{d_i}} - \frac{\text{sgn}(\tilde{\beta}_j)\beta_j}{\sqrt{d_j}} \right)^2, \quad (3) \]

\( \tilde{\beta}_j \): an initial estimate based on Enet; a 2-step procedure.
• $L_\gamma$-norm with $\gamma > 1$ (Pan, Xie and Shen 2010):

$$p_\lambda(\beta; \gamma, w) = \lambda 2^{1/\gamma} \sum_{i \sim j} \left( \frac{|\beta_i|^{\gamma}}{w_i} + \frac{|\beta_j|^{\gamma}}{w_j} \right)^{1/\gamma} \quad (4)$$

• $w_i$: smooth what?
  1) $w_i = d_i^{(\gamma+1)/2}$: smooth $|\beta_i|/\sqrt{d_i}$, as in Li and Li;
  2) $w_i = d_i$: smooth $|\beta_i|$

Some theory under simplified cases.

• Feature: each term is an $L_\gamma$ norm, $\gamma \geq 1$

$\implies$ **group** variable selection!; Yuan and Lin 2006, Zhao et al 2007.

$\implies$ tend to realize $\hat{\beta}_i = \hat{\beta}_j = 0$ if $i \sim j$!
Corollary 1 Assume that $X'X = I$. For any edge $i \sim j$, a sufficient condition for $\hat{\beta}_i = \hat{\beta}_j = 0$ is

$$||(\tilde{\beta}_i, \tilde{\beta}_j)||_{\gamma'}^{(1/w_i, 1/w_j)} \leq \lambda 2^{1/\gamma'},$$

(5)

and a necessary condition is

$$||(\tilde{\beta}_i, \tilde{\beta}_j)||_{\gamma'}^{(1/w_i, 1/w_j)} \leq \lambda 2^{1/\gamma'} + d_i + d_j - 2,$$

(6)

where $(\tilde{\beta}_i, \tilde{\beta}_j)$ are OLSEs.
• $\gamma$: a larger $\gamma$ smoothes more;
• $L_\infty$: related to OSCAR (Bondell & Reich 2008)

$$p_\lambda = \lambda \sum_{i \sim j} \max \left( \frac{|\beta_i|}{\sqrt{d_i}}, \frac{|\beta_j|}{\sqrt{d_j}} \right)$$

maximally forces $|\hat{\beta}_i|/\sqrt{d_i} = |\hat{\beta}_j|/\sqrt{d_i}$ if $i \sim j$!

• Other theoretical results (under simplified conditions): shrinkage effects, grouping effects ...

• Computational algorithm of Pan et al (2010):
  Generalized boosted lasso (GBL) (Zhao and Yu 2004); providing approximate solution paths.

• Use CV to choose tuning parameters, e.g. $\lambda$.

• Conclusion of Pan et al (2010): best for variable selection, but not necessarily in prediction (PMSE).
A surprise: $\gamma = \infty$ did not work well!

- Why?

  Use Matlab CVX package; slower but better performance.

- 2) Bias due to group var selection:
  $aL_\infty$: use a 2-step procedure as Grace of Li and Li (2010).
New method

• Relax the smoothness assumption:
  New assumption: neighboring genes are more likely to participate or not participate at the same time; no assumption on the smoothness of regression coefficients.

• Prior: if $i \sim j$, more likely to have $I(\beta_i \neq 0) = I(\beta_j \neq 0)$ just for variable selection

• How to approximate the discontinuous $I(\beta_j \neq 0)$?
  Truncated Lasso Penalty (Shen, Pan & Zhu 2012, JASA):
  \[
  J_\tau(\beta_j) = \min(1, |\beta_j|/\tau) \to I(\beta_j \neq 0)
  \]
  as $\tau \to 0^+$; see Fig:
- TLP: related to SCAD (Fan and Li 2001), MCP (Zhang 2010), SELO (Dicker et al 2012; Li, Wang & Lin 2012), ..., but ...

Figure 3:
• Use a new penalty to approximate $\sum_{i \sim j} |I(\beta_i \neq 0) - I(\beta_j \neq 0)|$:

$$p_\lambda(\beta; \tau) = \lambda \sum_{i \sim j} |J_\tau(\beta_i) - J_\tau(\beta_j)|.$$  

• But $p_\lambda(\beta; \tau)$ is not convex; use difference convex (DC) programming (Tao & An 1998)! related to MM (Hunter & Lange 2010).

• Two tricks:
  1) $J_\tau(z) = \frac{1}{\tau}(|z| - \max(|z| - \tau, 0));$
  2) $|u - v| = 2\max(u, v) - (u + v)$.

• $TTLP_I$:

$$p(\beta) = \lambda_1 \sum_{j=1}^{p} J_\tau(\beta_j) + \lambda_2 \sum_{j \sim j'} \left| J_\tau \left( \frac{\beta_j}{w_j} \right) - J_\tau \left( \frac{\beta_{j'}}{w_{j'}} \right) \right|,$$  

16
• **LTLP**

\[
p(\beta) = \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j \sim j'} \left| J_\tau \left( \frac{|\beta_j|}{w_j} \right) - J_\tau \left( \frac{|\beta_{j'}|}{w_{j'}} \right) \right|, \quad (8)
\]

• **LTLP**

\[
p(\beta) = p_1(\beta) - p_2(\beta),
\]

\[
p_1(\beta) = \frac{1}{\tau} \left( \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j \sim j'} 2\max(u_j, v_j) \right),
\]

\[
p_2(\beta) = \frac{1}{\tau} \left( \lambda_2 \sum_{j \sim j'} (u_j + v_j) \right),
\]

\[
u_j = \frac{|\beta_j|}{w_j} + \max\left( \frac{|\beta_{j'}|}{w_{j'}} - \tau, 0 \right) \text{ and } v_j = \frac{|\beta_{j'}|}{w_{j'}} + \max\left( \frac{|\beta_j|}{w_j} - \tau, 0 \right).
\]

• Linearizing \( p_2 \) at a current estimate \( \hat{\beta}^{(m-1)} \) and ignoring terms
independent of $\beta$, we obtain a convex approximation of $S(\beta)$:

$$S^{(m)}(\beta) = \frac{1}{2} \|Y - X\beta\|^2 + \frac{\lambda_1}{\tau} \sum_{j=1}^{p} |\beta_j| + \frac{\lambda_2}{\tau} \sum_{j \sim j'} 2\max(u_j, v_j)$$

$$-\frac{\lambda_2}{\tau} \sum_{j \sim j'} \left( \frac{\beta_j}{w_j} \mathrm{Sgn}(\hat{\beta}_j^{(m-1)})(m-1) |1 + I\left(\frac{\hat{\beta}_j^{(m-1)}}{w_j} > \tau\right)\right)$$

$$+ \frac{\beta_{j'}}{w_{j'}} \mathrm{Sgn}(\hat{\beta}_{j'}^{(m-1)})(m-1) |1 + I\left(\frac{\hat{\beta}_{j'}^{(m-1)}}{w_{j'}} > \tau\right)\right),$$

which is minimized to obtain an updated estimate $\hat{\beta}^{(m)}$.

- Since $S^{(m)}(\beta)$ is convex, we use Matlab package CVX.

- **Theorem**: the above DC algorithm monotonically converges to a local minimum in finite steps.

- Use grid search and CV to determine the choice of $(\tau, \lambda_1, \lambda_2)$.  


• Simulation set-ups:
  network: 10 subnetworks, each with one TF connects to ist 10 targets (Li and Li 2008);
  \( n = 50, \ p = p_1 + p_0 = 44 + 66; \)

• True \( \beta \): for \( j \sim j' \),
  Set-up 1: \( \beta_j / \sqrt{d_j} = \beta_{j'} / \sqrt{d_{j'}}; \)
  Set-up 2: \( |\beta_j| / \sqrt{d_j} = |\beta_{j'}| / \sqrt{d_{j'}}; \)
  Set-up 3: \( |\beta_j| / \sqrt{d_j} \neq |\beta_{j'}| / \sqrt{d_{j'}} \) but \( I(\beta_j \neq 0) = I(\beta_{j'} \neq 0) \).

• Use \( w_j = \sqrt{d_j} \) (and \( w_j = 1 \), not shown).

• \( ME = (\beta - \hat{\beta})' E(X'X)(\beta - \hat{\beta}); \)
  PE: prediction mean squared error for \( Y \); \( PE=ME+c; \)
  \( TP = |\{j : \beta_j \neq 0, \hat{\beta}_j \neq 0\}|; \) (max \( TP=22 \))
  \( FP = |\{j : \beta_j = 0, \hat{\beta}_j \neq 0\}|; \)
### Set-up 1: mean[median](sd)

<table>
<thead>
<tr>
<th>Method</th>
<th>ME(sd)</th>
<th>PE(sd)</th>
<th>TP</th>
<th>FP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lasso</td>
<td>44.2(13.2)</td>
<td>66.2(13.1)</td>
<td>13.5<a href="3.2">14</a></td>
<td>16.8<a href="19.2">13</a></td>
</tr>
<tr>
<td>Enet</td>
<td>34.2(13.1)</td>
<td>65.0(13.5)</td>
<td>16.5<a href="3.7">17</a></td>
<td>22.2<a href="16.6">18</a></td>
</tr>
<tr>
<td>Grace</td>
<td>4.7(3.6)</td>
<td>39.7(5.8)</td>
<td>22.0<a href="0.1">22</a></td>
<td>59.5<a href="21.2">63</a></td>
</tr>
<tr>
<td>aGrace</td>
<td>23.9(16.4)</td>
<td>55.6(14.4)</td>
<td>17.6<a href="4.1">18</a></td>
<td>29.4<a href="22.3">23.5</a></td>
</tr>
<tr>
<td>$L_\infty$</td>
<td>14.2(8.0)</td>
<td>50.4(11.2)</td>
<td>22.0<a href="0.0">22</a></td>
<td>9.7<a href="6.8">8</a></td>
</tr>
<tr>
<td>a$L_\infty$</td>
<td>4.3(4.1)</td>
<td>38.8(6.0)</td>
<td>22.0<a href="0.0">22</a></td>
<td>4.1<a href="5.4">2</a></td>
</tr>
<tr>
<td>TTLPI</td>
<td>12.4(12.0)</td>
<td>45.4(9.1)</td>
<td>21.5<a href="2.7">22</a></td>
<td>20.2<a href="28.3">1</a></td>
</tr>
<tr>
<td>LTLP_I</td>
<td>9.6(8.5)</td>
<td>43.4(8.5)</td>
<td>21.7<a href="1.4">22</a></td>
<td>23.4<a href="17.0">22</a></td>
</tr>
</tbody>
</table>
### Set-up 2: mean[median](sd)

<table>
<thead>
<tr>
<th>Method</th>
<th>ME(sd)</th>
<th>PE(sd)</th>
<th>TP</th>
<th>FP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lasso</td>
<td>34.6(8.8)</td>
<td>67.9(11.4)</td>
<td>10.2<a href="3.0">9.5</a></td>
<td>13.4<a href="15.4">9.0</a></td>
</tr>
<tr>
<td>Enet</td>
<td>34.8(8.5)</td>
<td>68.2(11.4)</td>
<td>13.2<a href="4.3">13.0</a></td>
<td>24.4<a href="22.1">18</a></td>
</tr>
<tr>
<td>Grace</td>
<td>27.1(5.7)</td>
<td>59.8(9.0)</td>
<td>18.5<a href="3.4">19</a></td>
<td>45.1<a href="25.1">43.5</a></td>
</tr>
<tr>
<td>aGrace</td>
<td>25.3(10.9)</td>
<td>58.4(11.6)</td>
<td>17.5<a href="5.0">19</a></td>
<td>41.9<a href="24.1">39.5</a></td>
</tr>
<tr>
<td>$L_{\infty}$</td>
<td>34.5(10.2)</td>
<td>65.1(12.2)</td>
<td>20.9<a href="2.6">22</a></td>
<td>15.2<a href="11.0">13</a></td>
</tr>
<tr>
<td>a$L_{\infty}$</td>
<td>20.7(9.9)</td>
<td>53.5(11.6)</td>
<td>20.7<a href="3.1">22</a></td>
<td>8.3<a href="10.7">5</a></td>
</tr>
<tr>
<td>$TTLP_I$</td>
<td>28.5(11.0)</td>
<td>59.5(11.3)</td>
<td>21.0<a href="3.3">22</a></td>
<td>26.7<a href="28.6">15</a></td>
</tr>
<tr>
<td>$LTLP_I$</td>
<td>23.2(8.1)</td>
<td>55.3(9.3)</td>
<td>21.4<a href="2.2">22</a></td>
<td>37.2<a href="21.4">33</a></td>
</tr>
</tbody>
</table>
## Set-up 3: mean[median](sd)

<table>
<thead>
<tr>
<th>Method</th>
<th>ME(sd)</th>
<th>PE(sd)</th>
<th>$TP$</th>
<th>$FP$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lasso</td>
<td>36.2(9.4)</td>
<td>67.0(11.3)</td>
<td>10.0<a href="3.3">10</a></td>
<td>13.6<a href="16.3">10</a></td>
</tr>
<tr>
<td>Enet</td>
<td>34.9(7.9)</td>
<td>65.8(10.3)</td>
<td>12.7<a href="3.8">12</a></td>
<td>22.7<a href="19.2">17</a></td>
</tr>
<tr>
<td>Grace</td>
<td>34.9(7.8)</td>
<td>65.4(10.6)</td>
<td>13.6<a href="4.2">14</a></td>
<td>24.8<a href="19.3">19</a></td>
</tr>
<tr>
<td>aGrace</td>
<td>36.2(8.4)</td>
<td>63.1(9.0)</td>
<td>15.2<a href="5.6">15</a></td>
<td>32.0<a href="24.3">24</a></td>
</tr>
<tr>
<td>$L_\infty$</td>
<td>33.9(8.1)</td>
<td>65.1(10.3)</td>
<td>15.3<a href="4.6">15</a></td>
<td>13.8<a href="11.5">11</a></td>
</tr>
<tr>
<td>a$L_\infty$</td>
<td>37.6(9.2)</td>
<td>66.0(12.1)</td>
<td>15.0<a href="4.7">15</a></td>
<td>9.7<a href="11.0">7.5</a></td>
</tr>
<tr>
<td>$TTLPI$</td>
<td>34.2(10.1)</td>
<td>63.9(10.9)</td>
<td>19.1<a href="5.2">22</a></td>
<td>20.1<a href="22.7">13</a></td>
</tr>
<tr>
<td>$LTLPI$</td>
<td>31.3(7.4)</td>
<td>61.1(9.6)</td>
<td>20.5<a href="3.7">22</a></td>
<td>39.2<a href="21.9">44</a></td>
</tr>
</tbody>
</table>
Example

- $n = 286$ breast cancer patients (Wang et al 2005); (time to) metastasis within a 5-year follow-up after surgery; 106 events;

- $n = 295$ breast cancer patients (van de Vijver et al 2002); (time to) metastasis within a 5-year follow-up after surgery; 78 events;

- Consider three tumor suppressor genes, $BRCA1$, $BRCA2$, $TP53$, and their direct neighbors in a PPI network (Chuang et al 2007);

- Fit a linear model
  $Y$: binary; $X$: expression levels of $p = 294$ genes;

- Goal: variable selection
  $Q$: which genes’ expression levels predict the survival time?
• Among $p = 294$ genes, 18 cancer (CA) genes.
• Split the sample into $n = 95, 95, 96$ for training, tuning, testing;
  repeat 20 times.
<table>
<thead>
<tr>
<th>Method</th>
<th>PE</th>
<th># CA</th>
<th># Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lasso</td>
<td>0.235(0.004)</td>
<td>0.30<a href="0.13">0.00</a></td>
<td>8.80<a href="1.91">8.00</a></td>
</tr>
<tr>
<td>Final</td>
<td>-</td>
<td>1</td>
<td>30</td>
</tr>
<tr>
<td>Enet</td>
<td>0.227(0.003)</td>
<td>0.20<a href="0.09">0.00</a></td>
<td>9.90<a href="2.60">1.00</a></td>
</tr>
<tr>
<td>Final</td>
<td>-</td>
<td>2</td>
<td>51</td>
</tr>
<tr>
<td>Grace</td>
<td>0.227(0.003)</td>
<td>0.70<a href="0.16">1.00</a></td>
<td>9.50<a href="2.38">2.50</a></td>
</tr>
<tr>
<td>Final</td>
<td>-</td>
<td>2</td>
<td>49</td>
</tr>
<tr>
<td>aGrace</td>
<td>0.229(0.003)</td>
<td>1.30<a href="0.25">1.00</a></td>
<td>10.20<a href="2.10">6.00</a></td>
</tr>
<tr>
<td>Final</td>
<td>-</td>
<td>2</td>
<td>52</td>
</tr>
<tr>
<td>$L_{inf}$</td>
<td>0.236(0.005)</td>
<td>0.10<a href="0.07">0.00</a></td>
<td>10.35<a href="1.97">7.50</a></td>
</tr>
<tr>
<td>Final</td>
<td>-</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>$aL_{inf}$</td>
<td>0.239(0.005)</td>
<td>0.10<a href="0.07">0.00</a></td>
<td>10.20<a href="2.43">7.50</a></td>
</tr>
<tr>
<td>Final</td>
<td>-</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>TTLP</td>
<td>0.282(0.015)</td>
<td>2.90<a href="0.34">3.00</a></td>
<td>12.00<a href="2.68">8.00</a></td>
</tr>
<tr>
<td>Final</td>
<td>-</td>
<td>4</td>
<td>30</td>
</tr>
<tr>
<td>LTLP</td>
<td>0.256(0.009)</td>
<td>1.35<a href="0.28">1.50</a></td>
<td>11.10<a href="2.07">8.00</a></td>
</tr>
<tr>
<td>Final</td>
<td>-</td>
<td>4</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td># Freq of selecting BRCA1, BRCA2 and TP53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lasso</td>
<td>BRCA1 (1), BRCA2 (0), TP53 (1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enet</td>
<td>BRCA1 (0), BRCA2 (0), TP53 (0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grace</td>
<td>BRCA1 (7), BRCA2 (2), TP53 (2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>aGrace</td>
<td>BRCA1 (10), BRCA2 (4), TP53 (9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$L_\infty$</td>
<td>BRCA1 (0), BRCA2 (0), TP53 (0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$aL_\infty$</td>
<td>BRCA1 (0), BRCA2 (0), TP53 (0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$TTLP_I$</td>
<td>BRCA1 (20), BRCA2 (10), TP53 (20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$LTLP_I$</td>
<td>BRCA1 (9), BRCA2 (5), TP53 (9)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 4: The final models by $TTLP_I$. 5 genes in hexagons: in both models; triangles/big circles: in only one; 5 red ones: BC genes.
Discussion

- Bayesian approaches (Moni and Li 2009; Li and Zhang 2009; Tai, Pan & Shen 2010):
  prior prob’s \( Pr(\beta_i \neq 0) \) modeled by a network-induced MRF.

- A new penalty (Zhu, Shen & Pan 2013, JASA):

  \[
p_\lambda(\beta; \tau) = \lambda \sum_{i \sim j} [J_\tau(\beta_i + \beta_j) + J_\tau(\beta_i - \beta_j)],
  \]

  aiming for

  \[
  \sum_{i \sim j} ||\beta_i| - |\beta_j||.
  \]

- Another application: eQTL mapping (Pan 2009)

  \[
  Y_g = X\beta_g + \epsilon_g, \quad E(\epsilon_g) = 0, \tag{9}
  \]

  for \( g = 1, \ldots, G \).

  \( X \): DNA markers; obs \((Y_1, \ldots, Y_G, X)\).
Q: which markers are associated with $Y_g$?
$\Rightarrow$ variable selection or ...

- Typical approaches:
  Gene-by-gene, separately, with possible var selection (Broman and Speed 2002; Wang et al 2011; ...)

- BUT, genes are related...
  e.g. as described by pathways or clusters (Lan et al 2003; Chun and Keles 2009; Zhang et al 2010; ...) or by a co-expression network (Pan 2009).
  $\Rightarrow Y_g$'s are correlated, and more likely to be co-regulated!

- Network assumption/prior: if two genes $g \sim h$ in a network, then $|\beta_g| \approx |\beta_h|$, or, $I(\beta_g \neq 0) = I(\beta_h \neq 0)$.
- Goal: utilize the above assumption/prior.
- How?
• Reformulate the original multiple regressions to a single regression:

\[ Y_c = (Y_1', ..., Y_G')', \]

\[ X_c = diag(X, ..., X), \]

\[ \beta = (\beta_1', ..., \beta_G')', \]

\[ Y = X\beta + \epsilon, \quad E(\epsilon) = 0, \quad (10) \]
Acknowledgement: This research was supported by NIH.

You can download our papers from
http://sph.umn.edu/ex biostatistics/techreports.php?

Thank you!