

# Supplementary Materials to Biomarker discovery from high-throughput data by connected network-constrained support vector machine

Lingyu Li<sup>a</sup>, Zhi-Ping Liu<sup>a,\*</sup>

<sup>a</sup>*School of Control Science and Engineering, Shandong University, Jinan, Shandong 250061, China*

## 1. Some notations

Consider a prior gene network documented in RegNetwork [1] that is represented by a simple graph  $G = (V, E)$  without self-loops or parallel edges, where  $V$  is a finite set of vertices that correspond to  $p$  genes,  $E = \{e_{kj} \mid k \sim j\}$  is a set of edges indicating that gene  $k$  and  $j$  are linked in the knowledge-based gene network [2]. Graph  $G$  is an undirected graph iff the set  $E$  has symmetry or adjoint property. So let  $w(e_{kj})$  be the weight of the edge  $e_{kj}$  and abbreviate it as  $w_{kj}$  with

$$w_{kj} = \begin{cases} 1, & \text{if } k \text{ and } j \text{ are adjacent,} \\ 0, & \text{otherwise.} \end{cases} \quad (\text{S1})$$

Then it holds  $w_{kj} = w_{jk}$ . Correspondingly, its degree function  $d : V \rightarrow \mathbb{R}^*$  at vertex  $k$  can be defined as

$$d_k := \sum_{k \sim j} w_{kj}. \quad (\text{S2})$$

Given a dataset  $\mathcal{D} = (\mathbf{X}, \mathbf{y})$ , let  $W \in \mathbb{R}^{p \times p}$  be the weight matrix of graph  $G$ . For a given prior network (e.g., RegNetwork for gene interactions [1]) of variables, define the degree matrix  $D$  with element  $d_k$ ,

$$d_k = \begin{cases} \sum_{k \sim j} w_{kj}, & \text{if } k \text{ and } j \text{ are adjacent,} \\ 0, & \text{if } k \text{ is an isolated vertex,} \end{cases} \quad (\text{S3})$$

which represents the degree of vertex  $v_k \in V$  ( $k = 1, \dots, p$ ). Obviously,  $D$  is a diagonal matrix with non-diagonal elements of 0.

The normalized version  $\mathcal{L}$  [3] associated with graph  $G$  is introduced to illustrate that the  $p$  explanatory features are measured in a graph. For the normalized Laplacian matrix  $\mathcal{L}$ , it is made up of  $\hat{l}_{kj}$  ( $k, j = 1, 2, \dots, p$ ) elements, which are defined by

$$\hat{l}_{kj} = \begin{cases} 1 - \frac{w_{kk}}{d_k}, & \text{if } k = j \text{ and } d_k \neq 0, \\ -\frac{w_{kj}}{\sqrt{d_k d_j}}, & \text{if } k \text{ and } j \text{ are adjacent,} \\ 0, & \text{otherwise.} \end{cases} \quad (\text{S4})$$

---

\* Corresponding author

Email addresses: lingyuli@mail.sdu.edu.cn (Lingyu Li), zpliu@sdu.edu.cn (Zhi-Ping Liu)

## 2. Some definitions

We will propose the connectivity of a graph/network mainly based on the definitions of the  $kj$ -node cut  $\mathcal{S}$  at first [4, 5]. Secondly, before describing the connectivity, we also need to introduce the definition of  $kj$ -node cut set  $\Gamma$  [6]. Finally, we give modeling connectivity theorem based on the *Dirac* measure widely used in measure theory [7].

**Definition 2.1.** ( *$kj$ -node cut*) Given that vertices  $k, j \in U$  that are nonadjacent ( $e_{kj} \notin E$ ), the subset  $\mathcal{S} \subseteq U \setminus \{k, j\}$  is a  $kj$ -node cut if there is no path between  $k$  and  $j$  in  $G[U \setminus \mathcal{S}]$ .

**Definition 2.2.** (*Minimal  $kj$ -node cut set*) For  $\forall e_{kj} \notin E$ , define

$$\Gamma(k, j) = \{\mathcal{S} \subseteq U \setminus \{k, j\} : \mathcal{S} \text{ is a minimal } kj\text{-node cut}\} \quad (\text{S5})$$

then  $\Gamma(k, j)$  is called the minimal  $kj$ -node cut set.

**Definition 2.3.** (*Dirac measure*) [8] Let  $C$  be a nonempty set and let  $x \in C$ . For every  $U \subseteq C$ , define

$$\delta_x(U) = \begin{cases} 1, & x \in U, \\ 0, & x \in C \setminus U, \end{cases} \quad (\text{S6})$$

on set  $C$ , then  $\delta_x$  is called the Dirac measure at point  $x$ .

**Theorem 2.1.** (*Modeling connectivity*) To impose  $U$  be connected, the following inequalities suffice,

$$\sum_{l \in \mathcal{S}} \delta_l(U) \geq \delta_k(U) + \delta_j(U) - 1, \\ \forall \mathcal{S} \in \Gamma(k, j), \forall k, j \in C, e_{kj} \notin E, \forall U \subseteq C \subseteq G, \quad (\text{S7})$$

where  $\delta_x(U)$  is 1 if  $x \in U$  and 0 if  $x \in C \setminus U$ .

In the connectivity constraints (S7), we can see that if  $\delta_k(U) = \delta_j(U) = 1$ , (i.e., if  $k$  and  $j$  are both selected), then they force at least one vertex  $l$  in  $\mathcal{S}$  to be selected as well [6, 9]. Thus for the Reg-SVM model, if both coefficients  $\theta_k$  and  $\theta_j$  are all non-zeros simultaneously, then  $\theta_l$  is also non-zero.

### 3. Theorem and proof

**Theorem 3.1.** Let  $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n]^T \in \mathbb{R}^{n \times p}$ , and let

$$\lambda_{\max} = \max\{\lambda_1, \lambda_2, \dots, \lambda_m\},$$

where  $m \in \mathbb{N}^*$ . Given the dataset  $\mathcal{D} = \{\mathbf{X}, \mathbf{y}\}$ , the maximum value of  $\lambda$  in CNet-SVM model with the solution  $\boldsymbol{\theta}^* \neq \mathbf{0}$  satisfies

$$\lambda_{\max} = \frac{1}{\alpha} \|\mathbf{X}^T \mathbf{y}\|_{\infty}. \quad (\text{S8})$$

This section gives the proof of Theorem 3.1.

*Proof.* Let

$$\mathcal{J}(\boldsymbol{\theta}, \lambda, \alpha) = \sum_{i=1}^n \left[ 1 - y_i (\mathbf{X}_i^T \boldsymbol{\theta} + \theta_0) \right]_+ + \lambda \alpha \|\boldsymbol{\theta}\|_1 + \lambda(1 - \alpha) \boldsymbol{\theta}^T \mathcal{L} \boldsymbol{\theta}. \quad (\text{S9})$$

When  $\boldsymbol{\theta} \neq \mathbf{0}$ , computing the gradient of  $\mathcal{J}(\boldsymbol{\theta}, \lambda, \alpha)$  with respect to  $\boldsymbol{\theta}$ , we get

$$\frac{\partial \mathcal{J}(\boldsymbol{\theta}, \lambda, \alpha)}{\partial \boldsymbol{\theta}} = -\mathbf{X}^T \mathbf{y} + \lambda \alpha \text{sign}(\boldsymbol{\theta}) + 2\lambda(1 - \alpha) \mathcal{L} \boldsymbol{\theta}. \quad (\text{S10})$$

From KKT conditions, we obtain that if  $\boldsymbol{\theta} = \mathbf{0}$ , then  $\lambda$  makes Equation (S10) zero. That is to say

$$\left. \frac{\partial \mathcal{J}(\boldsymbol{\theta}, \lambda, \alpha)}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta} = \mathbf{0}} = -\mathbf{X}^T \mathbf{y} + \lambda \alpha \mathbf{1}_p = \mathbf{0}_p, \quad (\text{S11})$$

where  $\mathbf{1}_p$  and  $\mathbf{0}_p$  are the  $p$ -dimensional column vectors which contain one and zero elements, respectively.

Using Equation (S11), we obtain

$$\lambda \alpha \mathbf{1}_p = \mathbf{X}^T \mathbf{y}. \quad (\text{S12})$$

Taking  $L_1$ -norm on both sides of Equation (S12), it derives to

$$p\lambda\alpha = \sum_{j=1}^p |(\mathbf{X}^T \mathbf{y})_j| \leq \sum_{j=1}^p \max_{1 \leq j \leq p} |(\mathbf{X}^T \mathbf{y})_j| = \sum_{j=1}^p \|\mathbf{X}^T \mathbf{y}\|_{\infty} = p\|\mathbf{X}^T \mathbf{y}\|_{\infty}, \quad (\text{S13})$$

From Equation (S13), we get  $\lambda \leq \frac{\|\mathbf{X}^T \mathbf{y}\|_{\infty}}{\alpha}$ . Therefore it holds

$$\lambda_{\max} = \frac{\|\mathbf{X}^T \mathbf{y}\|_{\infty}}{\alpha}. \quad (\text{S14})$$

This completes the proof of Theorem 3.1.  $\square$

#### 4. Figure S1

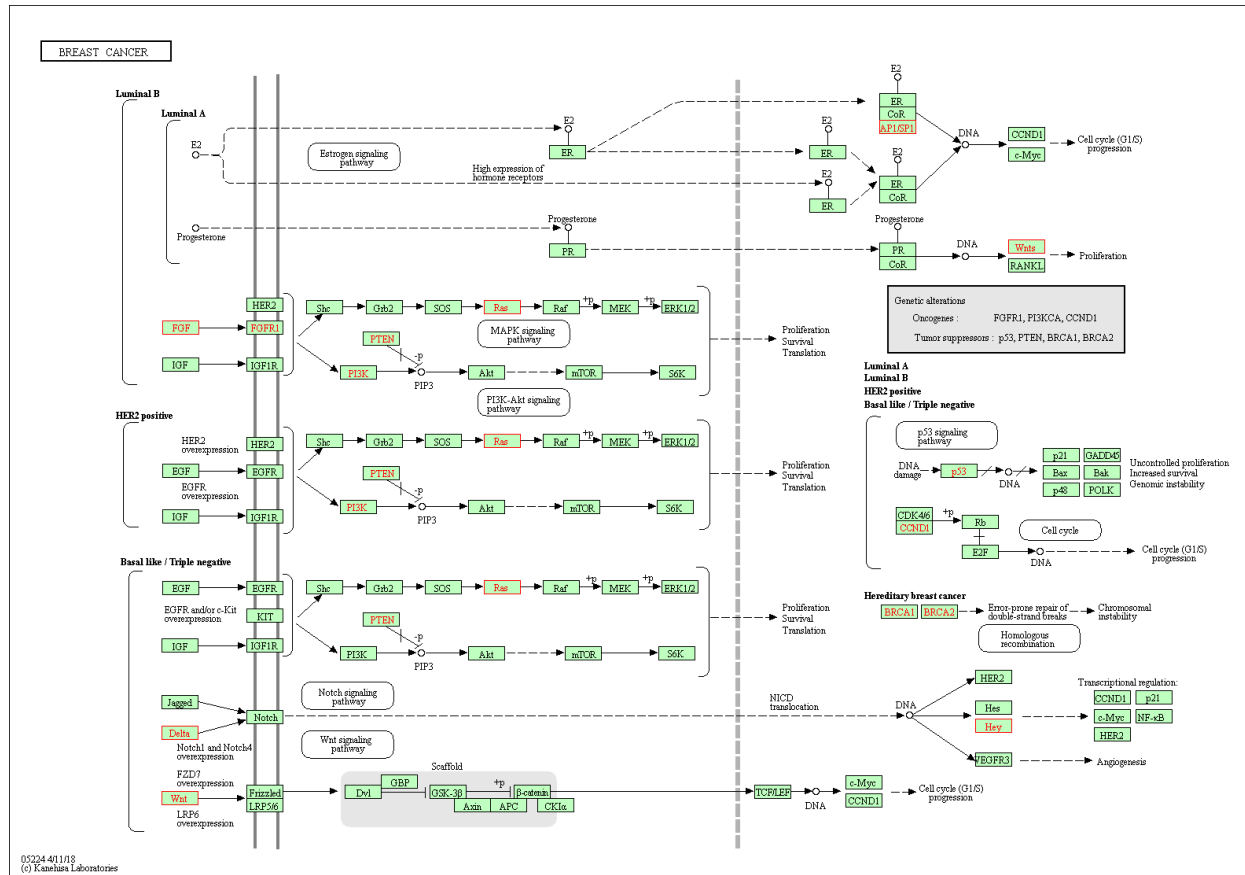


Figure S1: The Breast Cancer pathway is the top pathway enriched by our selected biomarkers.

## 5. Figure S2

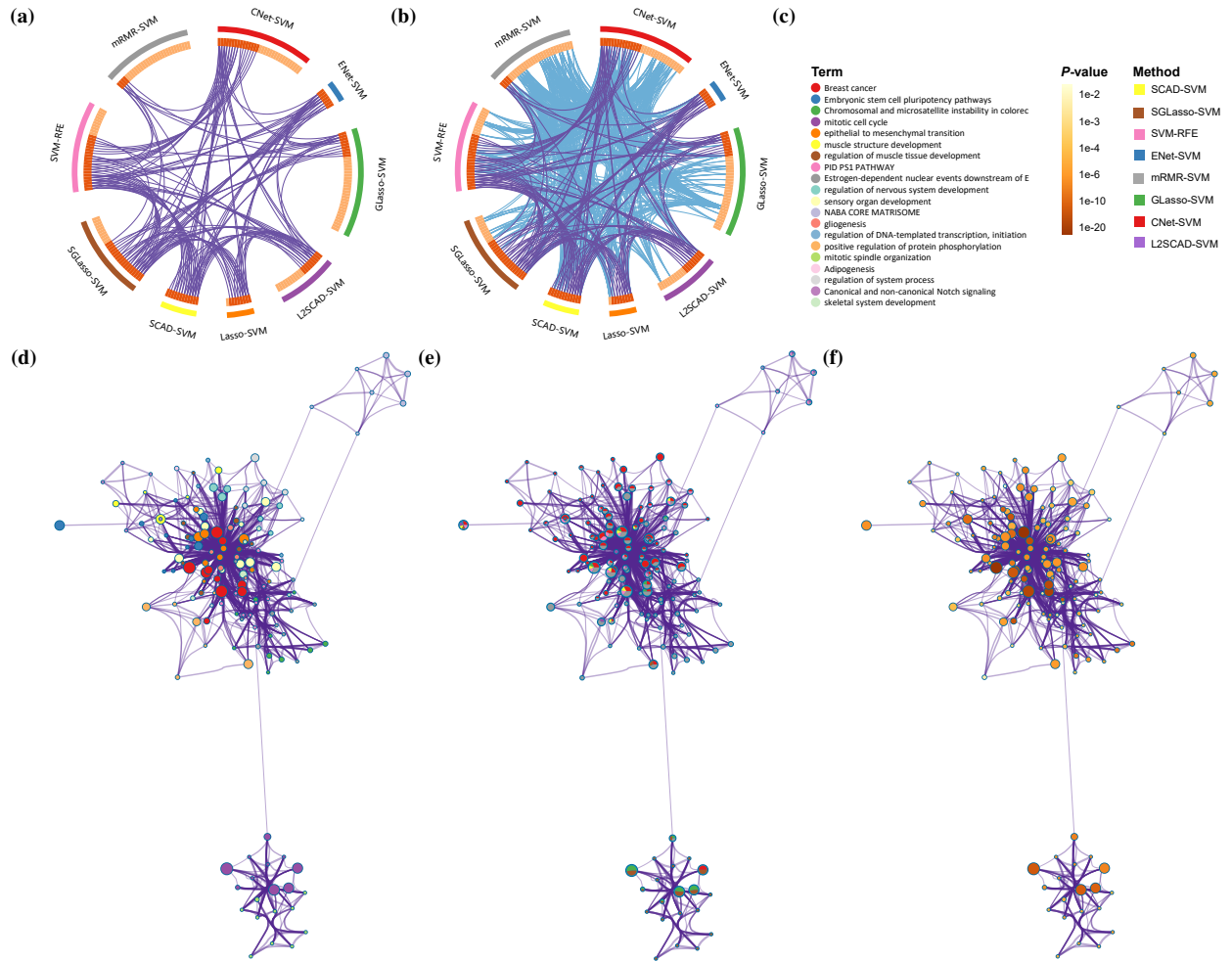


Figure S2: The functional enrichment analysis of biomarkers respectively identified by nine methods. (a) The same genes shared by multiple gene subsets are linked by purple lines. (b) The different genes where they fall into the same ontology term are linked by blue links. (c) The legends for (d), (e) and (f). (d) The enriched ontology clusters are colored by cluster labels. (e) The same enrichment network, with nodes colored by *P*-value. (f) The same enrichment network, with nodes displayed as pies.

## References

- [1] Z.-P. Liu, C. Wu, H. Miao, H. Wu, RegNetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse, *Database* 2015 (2015) 1–12. <https://doi.org/10.1093/database/bav095>.
- [2] L. Chen, J. Xuan, R. B. Riggins, R. Clarke, Y. Wang, Identifying cancer biomarkers by network-constrained support vector machines, *BMC Systems Biology* 5 (2011) 1–20. <https://doi.org/10.1186/1752-0509-5-161>.
- [3] S. Bougleux, A. Elmoataz, M. Melkemi, Local and nonlocal discrete regularization on weighted graphs for image and mesh processing, *International Journal of Computer Vision* 84 (2009) 220–236. <https://doi.org/10.1007/s11263-008-0159-z>.
- [4] E. Althaus, M. Blumenstock, A. Disterhoft, A. Hildebrandt, M. Krupp, Algorithms for the maximum weight connected  $k$ -induced subgraph problem, in: *International Conference on Combinatorial Optimization and Applications*, Springer, 2014, pp. 268–282. [https://doi.org/10.1007/978-3-319-12691-3\\_21](https://doi.org/10.1007/978-3-319-12691-3_21).
- [5] E. Álvarez-Miranda, M. Sinnl, A Relax-and-Cut framework for large-scale maximum weight connected subgraph problems, *Computers & Operations Research* 87 (2017) 63–82. <https://doi.org/10.1016/j.cor.2017.05.015>.
- [6] R. Carvajal, M. Constantino, M. Goycoolea, J. P. Vielma, A. Weintraub, Imposing connectivity constraints in forest planning models, *Operations Research* 61 (2013) 824–836. <https://doi.org/10.1287/opre.2013.1183>.
- [7] L. Li, Z.-P. Liu, A connected network-regularized logistic regression model for feature selection, *Applied Intelligence* 52 (2022) 1–31. <https://doi.org/10.1007/s10489-021-02877-3>.
- [8] M. M. Rao, *Measure theory and integration*, CRC Press, 2018.
- [9] Y. Wang, A. Buchanan, S. Butenko, On imposing connectivity constraints in integer programs, *Applications of Management Science: in Productivity, Finance, and Operations* 166 (2017) 241–271. <https://doi.org/10.1007/s10107-017-1117-8>.