Journal of the Korean Data & Information Science Society
2014, **25**(5), 1107–1116

http://dx.doi.org/10.7465/jkdi.2014.25.5.1107
한국데이터정보과학회지

Recent developments of constructing adjacency matrix in network analysis†

Younghee Hong[1] · Choongrak Kim[2]

[1,2]Department of Statistics, Pusan National University

Received 28 May 2014, revised 25 June 2014, accepted 2 July 2014

Abstract

In this paper, we review recent developments in network analysis using the graph theory, and introduce ongoing research area with relevant theoretical results. In specific, we introduce basic notations in graph, and conditional and marginal approach in constructing the adjacency matrix. Also, we introduce the Marcenko-Pastur law, the Tracy-Widom law, the white Wishart distribution, and the spiked distribution. Finally, we mention the relationship between degrees and eigenvalues for the detection of hubs in a network.

Keywords: Adjacency matrix, conditional dependency, graph theory, marginal dependency, Tracy-Widom law.

1. Introduction

Recently Big data is one of the most hot issues in many branches of science, and complex network problem is the central issue in Big data. Lots of interest arose in complex networks such as the world-wide web, the internet, biological networks, social networks, and so on. A wonderful scientific fact is that any complex network can be represented by a graph, and any graph can be represented by an adjacency matrix. Further, degree matrix and/or Laplacian matrix are derived by the adjacency matrix.

Graph theory has been developed by mathematicians for a long time, and mathematicians are interested in properties of eigenvalues of adjacency matrix, distribution of eigenvalues (Wigner, 1955; Marcenko and Pastur, 1967) and distribution of the largest eigenvalue (Tracy and Widom, 1996), lower and/or upper bound of eigenvalues, relationship between degree and eigenvalues, and so on. There are numerous books on the graph theory, and one of the recent and best references is Mieghem (2010).

On the other hand, statisticians paid attention to the graph theory quite recently, and they are mainly interested in graphical models and estimation of adjacency matrix using available observations. Recently, statisticians are interested in the case where p (number of variables) is larger than n (sample size) because lots of recent networks such as biological networks and

† This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education(2013012741).
[1] Ph.D. candidate, Department of Statistics, Pusan National University, Busan 609-735, Korea.
[2] Corresponding author: Professor, Department of Statistics, Pusan National University, Busan 609-735, Korea. E-mail: crkim@pusan.ac.kr

social networks reveal $p \gg n$ pattern. The seminal paper in this area is Dempster (1972). Since then, inspired by Dempster (1972), outstanding achievements have been made. Among them, Whittaker (1990), Edward (2000), Meinshausen and Buhlmann (2006), Yuan and Lin (2007), Peng *et al.* (2009), Rothman *et al.* (2010), Bien and Tibshirani (2011), Cai and Yuan (2012) and Chandrasekaran *et al.* (2012) are often referred. Also, noticeable results on the distribution of the largest eigenvalue of the sample covariance matrix under various assumptions are done by Johnstone (2001, 2008), Bickel and Levina (2008), Rothman *et al.* (2009), Cai and Liu (2011), Cai and Zhou (2012) and Birnbaum *et al.* (2013). Recently, Won and Choi (2014) applied the network theory to the comparison of citations in journals.

When a graph is given, scientists are interested in many aspects of a graph. Among them, interesting questions are as follows; (1) How many hubs in a graph and how to find hubs? (2) How many clusters in a graph? (3) How to sample a random graph? etc. In this paper, we review recent developments in network analysis using the graph theory, and introduce ongoing research area with relevant theoretical results. In specific, we introduce basic notations in graph, and conditional and marginal approach in constructing the adjacency matrix. Also, we introduce the Marcenko-Pastur law, the Tracy-Widom law, the white Wishart distribution, and the spiked distribution. Finally, we mention the relationship between degrees and eigenvalues for the detection of hubs in a network.

2. Graph theory and networks

2.1. Graph and network

We introduce the notations and examples of graphs. A given network is often represented by a graph $G = G(V, E)$, where $V = \{1, \cdots, p\}$ is the set of nodes (vertices) and E is the set of edges in $V \times V$. Let $a_{ij}, i = 1, \cdots, p, j = 1, \cdots, p$ denotes the connectivity between two nodes i and j. If $(i, j) \in E$, then two nodes i and j are said to be adjacent and a_{ij} has a nonzero arbitrary value. Therefore, a_{ij} denotes the closeness (or adjacency, connectivity) between two nodes i and j, and $\boldsymbol{A} = (a_{ij})$ is often called an adjacency matrix. If a_{ij} takes only either 1 or 0, then it called unweighted adjacency. When a_{ij} can take any real values (usually takes real values between 0 and 1), it called weighted adjacency. A graph is called directed if $a_{ij} \neq a_{ji}$, and called undirected if $a_{ij} = a_{ji}$. Throughout this thesis, we assume that the adjacency matrix is undirected, and therefore, the adjacency matrix is symmetric. The degree of ith node is defined as the magnitude of connectivity, and it is denoted as d_i, i.e., $d_i = \sum_{j=1}^{p} a_{ij}$. In unweighted adjacency case, the degree of ith node d_i is the number of adjacent nodes to ith node. We denote $\boldsymbol{D} = diag(d_1, \cdots, d_p)$ the degree matrix, and the Laplacian matrix is defined as $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$.

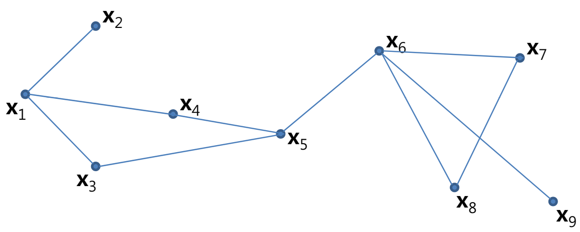

Figure 2.1 Graph of unweighted adjacency

The unweighted graph, given in Figure 2.1, has 9 nodes and 11 edges, and the degrees of each node are 3, 1, 2, 2, 3, 4, 2, 2 and 1, respectively.

2.2. Results on graph theory

There are lots of theoretical results on graphs, however, we only list very important and relevant results to statistics. First, we introduce the Wigners Semicircle law (Wigner, 1955). Let $\boldsymbol{A}$ be a random $p \times p$ real symmetric matrix with independent and identically distributed elements a_{ij} with $Var(a_{ij}) = \sigma^2$ and an eigenvalue of the set of the p real eigenvalues of the scaled matrix $\boldsymbol{A}_p = \boldsymbol{A}/\sqrt{p}$ denoted by $\lambda(\boldsymbol{A}_p)$. Then, the probability density function of $\lambda(\boldsymbol{A}_p)$ converges to

$$f(x) \;=\; \frac{1}{2\pi\sigma^2}\sqrt{4\sigma^2 - x^2}\; I(|x| \leq 2\sigma).$$

Second, the most important result in random matrix theory is the Marcenko-Pastur law (Marcenko and Pastur, 1967), and it can be described as follows. Let $\boldsymbol{C}$ be a random $p \times n$ matrix with independent and identically distributed elements c_{ij} with $E(c_{ij}) = 0$ and $Var(c_{ij}) = \sigma^2$. Let $y = p/n$ as $n \to \infty$ and let $a(y) = \sigma^2(1-\sqrt{y})^2$ and $b(y) = \sigma^2(1+\sqrt{y})^2$, and an eigenvalue of the set of the p real eigenvalues of the scaled Hermitian matrix $\boldsymbol{S} = \boldsymbol{C}\boldsymbol{C}^*/n$ denoted by $\lambda(\boldsymbol{S})$. Then, the probability density function of $\lambda(\boldsymbol{S})$ converges to

$$f(x) = \frac{\sqrt{(x-a(y))(b(y)-x)}}{2\pi x y \sigma^2}\; I(a(y) \leq x \leq b(y)) \;+\; \left(1 - \frac{1}{y}\right)\delta(x) I(y > 1),$$

where $\delta(\cdot)$ is a Dirac function.

Third, the most relevant result to statisticians is the Tracy-Widom law (Tracy and Widom, 1996) studied by Tracy and Widom (1996, 2000), Johnstone (2001, 2008), Birnbaum *et al.* (2013), and many others. The main result is as follows; Suppose that $\boldsymbol{X} = (X_{ij})_{p\times n}$ has entries which are iid $N(0,1)$. Let the sample eigenvalues of the white Wishart matrix $\boldsymbol{X}\boldsymbol{X}^T$ be $\lambda_1 > \cdots > \lambda_p$. Also, let

$$\mu_{np} = (\sqrt{n-1} + \sqrt{p})^2,$$

$$\sigma_{np} = (\sqrt{n-1} + \sqrt{p})\;\left(\frac{1}{\sqrt{n-1}} + \frac{1}{\sqrt{p}}\right)^{1/3}.$$

Then, the Tracy-Widom law of order 1 has the distribution function defined as

$$F_1(t) = exp\;\left\{-\frac{1}{2}\int_t^{\infty} q(x) + (x-t)q^2(x)dx\right\},$$

where the function q solves the Painleve II differential equation

$$q^{''}(x) = xq(x) + 2q^3(x)$$

and $q(x)$ is the Airy function. This result was first found by Tracy and Widom (1996) as the limiting distribution of the largest eigenvalue of the Gaussian symmetric matrix. Johnstone (2001) showed that if $n/p \to \gamma \geq 1$ then $(\lambda_1 - \mu_{np})/\sigma_{np}$ converges in distribution to F_1. Since the distribution F_1 cannot be expressed as an analytic form, Tracy and Widom (2000) evaluated F_1 numerically. They showed that F_1 is unimodal and asymmetric with mean -1.21 and standard deviation 1.27. Recently, Pillai and Yin (2012) showed that the largest eigenvalue of the correlation matrix also follows the same distribution. Recently, Birnbaum *et al.* (2013) studied minimax bounds in the sparse spiked distribution, where the covariance matrix $\boldsymbol{\Sigma}$ is not an identity matrix but $\boldsymbol{\Sigma} = diag(\tau_1^2, \cdots, \tau_M^2, 1, \cdots, 1)$, and M is an unknown parameter. In the setup of spiked distribution, it is assumed that there are M hubs. Therefore, the white Wishart corresponds to the null hypothesis (no hubs), and the spiked distribution corresponds to the alternative hypothesis (M hubs). See references in Birnbaum *et al.* (2013) for studies in the spiked distribution.

3. Construction of adjacency matrix

Note that the Laplacian matrix is defined by the adjacency matrix, and the adjacency between two nodes i and j, denoted by a_{ij}, reveals the strength of connectivity between two nodes. Therefore, network analysis starts with efficient estimation of adjacency matrix. In most statistical literature, a_{ij} is estimated as either 0 or 1, i.e., unweighted adjacency case only. This assumption is very restrictive and unrealistic because there exists weak or strong relation between two nodes. Hence, it is reasonable to consider the weighted adjacency case. To deal with the weighted adjacency matrix, we start with the investigation of correlation matrix which is a basic building block in constructing the weighted adjacency matrix.

3.1. Interpretations of correlation matrix

Let $\boldsymbol{X} = (X_1, \cdots, X_p)$ be a p-dimensional random vector distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_p)$ and $\boldsymbol{\Sigma} = (\sigma_{ij})$. Now, we mention that the jth diagonal element of the inverse of correlation matrix is

$$1/(1 - R_j^2), \tag{3.1}$$

where R_j^2 is the coefficient of determination when regressing the jth variable (i.e., response variable) to other variables (i.e.,covariates). This result implies that if the jth diagonal element of the inverse of correlation matrix is large, then the jth variables is highly correlated with others. In networks, this result can be interpreted as follows; If the jth diagonal element of the inverse of correlation matrix is large, then d_j, the jth diagonal element of the degree matrix, is large. Therefore, the diagonal elements of the inverse of correlation matrix show the dependency of each node to others. For the unweighted adjacency matrix case, each diagonal element of the inverse of correlation matrix denotes the number of connected nodes to each node.

To verify (3.1), without loss of generality, let $j = 1$ and let

$$\begin{aligned} \boldsymbol{\gamma} &= (\sigma_{12}, \cdots, \sigma_{1p})^t \\ \boldsymbol{\Sigma}_{(11)} &= \begin{pmatrix} \sigma_2^2 & \cdots & \sigma_{2p} \\ & \cdots & \\ \sigma_{2p} & \cdots & \sigma_p^2 \end{pmatrix}. \end{aligned}$$

Now, the population version of R_1^2 is

$$R_1^2 = \frac{SSR}{SST} = \frac{\boldsymbol{\beta}^t\,\boldsymbol{\Sigma}_{(11)}\,\boldsymbol{\beta}}{\sigma_1^2}.$$

Now, $\boldsymbol{\beta} = \boldsymbol{\Sigma}_{(11)}^{-1}\boldsymbol{\gamma}$, and therefore $R_1^2 = \boldsymbol{\gamma}^t\boldsymbol{\Sigma}_{(11)}^{-1}\gamma/\sigma_1^2$. Note that

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \sigma_1^2 & \boldsymbol{\gamma}^t \\ \boldsymbol{\gamma} & \boldsymbol{\Sigma}_{(11)} \end{pmatrix}^{-1}$$

and use the lemma for the inverse of partitioned matrix, i.e.,

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix}^{-1} = \begin{pmatrix} (\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{Q}^{-1} \\ -\boldsymbol{Q}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} & \boldsymbol{Q}^{-1} \end{pmatrix},$$

where $\boldsymbol{Q} = \boldsymbol{D} - \boldsymbol{C}\,\boldsymbol{A}^{-1}\,\boldsymbol{B}$. Hence, the first diagonal element of $\boldsymbol{\Sigma}^{-1}$ is $(\sigma_1^2 - \boldsymbol{\gamma}^t\,\boldsymbol{\Sigma}_{(11)}^{-1}\,\gamma)^{-1}$. Therefore, $1/(1-R_1^2)$ is equivalent to the first diagonal element of the inverse of correlation matrix.

The next issue in correlation matrix is marginal and conditional independency, and it is a crucial aspect of network theory. For notational convenience, we consider $p = 3$ case only. Let $\boldsymbol{X} = (X_1, X_2, X_3)^t$ be distributed as a multivariate normal with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, where

$$\boldsymbol{\mu} = (\,\mu_1,\,\mu_2,\,\mu_3\,)^t$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix}.$$

Note that the conditional distribution of (X_1, X_2) given X_3 is a bivariate normal with mean vector $\boldsymbol{\mu}^*$ and covariance matrix $\boldsymbol{\Sigma}^*$, where

$$\boldsymbol{\mu}^* = \left(\ \mu_1 + \rho_{13}\sigma_1{\sigma_3}^{-1}(x_3 - \mu_3), \quad \mu_2 + \rho_{23}\sigma_2{\sigma_3}^{-1}(x_3 - \mu_3)\ \right)^t$$

$$\boldsymbol{\Sigma}^* = \begin{pmatrix} (1-\rho_{13}^2)\sigma_1^2 & (\rho_{12} - \rho_{13}\rho_{23})\sigma_1\sigma_2 \\ (\rho_{12} - \rho_{13}\rho_{23})\sigma_1\sigma_2 & (1-\rho_{23}^2)\sigma_2^2 \end{pmatrix}.$$

Then, the conditional correlation of (X_1, X_2) given X_3 is

$$Corr(X_1, X_2|X_3\,) = \frac{(\rho_{12} - \rho_{13}\rho_{23})}{\sqrt{(1-\rho_{13}^2)(1-\rho_{23}^2)}}.$$

On the other hand, the marginal correlation of X_1 and X_2 is

$$Corr(X_1, X_2) = \rho_{12}.$$

Therefore, $Corr(X_1, X_2|X_3\,) = Corr(X_1, X_2)$ only when $\rho_{13} = \rho_{23} = 0$. Note that $Corr(X_1, X_2)$ is large when ρ_{12} is large irrespective of the magnitudes of ρ_{13} and ρ_{23}, however, $Corr\,(\,X_1, X_2|X_3\,)$ is large only when two conditions are satisfied; First, ρ_{12} is large and both ρ_{13} and ρ_{23} are small. Second, ρ_{12} is small and both ρ_{13} and ρ_{23} are large.

Now, the above arguments can be generalized to the p-dimensional case. Consider a scaled version of the inverse of the correlation matrix, called scaled inverse correlation matrix, in which diagonals have unit entries. Then, the off diagonal elements of the scaled inverse correlation matrix are the negative of the conditional (also called partial) correlation coefficients between the corresponding pair of variables given the remaining variables. Here is an illustration based on a real data set.

Example 3.1 Table 3.1 shows the sample covariance matrix based on marks in five mathematics exams (mechanics, vector, algebra, analysis, statistics) for 88 students (Mardia *et al.*, 1979)

Table 3.1 The sample covariance matrix.

	mechanics	vector	algebra	analysis	statistics
mechanics	302.29				
vector	125.78	170.88			
algebra	100.43	84.19	111.60		
analysi s	105.07	93.60	110.84	217.88	
statistics	116.07	97.89	120.49	153.77	294.37

From the sample covariance matrix, we can easily compute the correlation matrix (Table 3.2) and the inverse correlation matrix (Table 3.3).

Table 3.2 The sample correlation matrix

	mechanics	vector	algebra	analysis	statistics
mechanics	1.0				
vector	0.55	1.0			
algebra	0.55	0.61	1.0		
analysi s	0.41	0.49	0.71	1.0	
statistics	0.39	0.44	0.66	0.61	1.0

Table 3.3 The inverse correlation matrix

	mechanics	vector	algebra	analysis	statistics
mechanics	1.60				
vector	-0.56	1.80			
algebra	-0.51	-0.66	3.04		
analysi s	0.00	-0.15	-1.11	2.18	
statistics	-0.04	-0.04	-0.86	-0.52	1.92

Off-diagonal elements of the correlation matrix show the marginal correlation between two variables. For example, the marginal correlation between vector and statistics is 0.44. Also, each diagonal element of the inverse correlation matrix is related to the coefficient of determination when the variable is regressed to the remaining variables, i.e., each diagonal element is $1/(1-R_j^2),\ j=1,\cdots,5$. Therefore, $R_j^2=(f_j-1)/f_j, j=1,\cdots,5$, where f_j is the jth diagonal element of the inverse correlation matrix. For example, $R_3^2=(3.04-1)/3.04=0.667$, and the algebra is most predictable variable, i.e., the algebra is most explained by other variables. While, the mechanics is least predictable. Hence, each diagonal element of the inverse correlation matrix shows the degree of connectivity to other variables. On the other hand, off-diagonal elements of the scaled inverse correlation matrix (Table 3.4) show negatives of the conditional correlation between two variables given others. For example, the conditional correlation between vector and statistics given other variables is 0.02 which is totally different from the marginal correlation 0.44.

Table 3.4 The scaled inverse correlation matrix

	mechanics	vector	algebra	analysis	statistics
mechanics	1.0				
vector	-0.33	1.0			
algebra	-0.23	-0.28	1.0		
analysi s	0.00	-0.08	-0.43	1.0	
statistics	-0.02	-0.02	-0.36	-0.25	1.0

3.2. Conditional and marginal dependency

Let $\boldsymbol{X} = (X_1, \cdots, X_p)$ be distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{X}_{-(i,j)}$ denote $p-2$ dimensional vector without X_i and X_j, i.e., $\boldsymbol{X}_{-(i,j)} = \{X_k | 1 \leq k \neq i, j \leq p\}$. Also, let σ^{ij} be the ijth component of the inverse covariance matrix $\boldsymbol{\Sigma}^{-1}$. Then, it is well known that the conditional (partial) correlation between the ith variable and the jth variable given other variables is represented by

$$\rho^{ij} = Corr(X_i, X_j | X_{-(i,j)}) = -\frac{\sigma^{ij}}{\sqrt{\sigma^{ii}\sigma^{jj}}}.$$

In fact, ρ^{ij} is the ijth component of the scaled inverse correlation matrix. Therefore, nonzeros in the inverse of covariance matrix (called concentration matrix or precision matrix) imply conditional dependence between variables. For this reason, covariance selection is called that the selection problem of nonzeros in inverse of covariance matrix. This approach, which consists of identification and estimation of nonzero entries in concentration matrix, is a very useful method in detecting associations among a set of random variables. The seminal paper in covariance selection is Dempster (1972) which mitigate the situation by reducing the effective number of parameters through imposing sparsity in concentration matrix. Inspired by Dempster (1972), many authors studied the covariance selection problem. Whittaker (1990) and Edward (2000) noted that traditional methods do not work when p (number of variables) is larger than n (sample size). Since then, several authors studied the covariance selection problem when $p \gg n$. Among them, Meinshausen and Buhlmann (2006) suggested an algorithm for identification of zeros in inverse of covariance matrix using lasso. Yuan and Lin (2007) discussed penalized maximum likelihood with a lasso penalty on inverse of covariance matrix. Peng *et al.* (2009) suggested an algorithm called SPACE (Sparse PArtial Correlation Estimation) for selecting nonzero partial correlations and hub identification by the lasso in high dimensional setting. All the approaches mentioned above are called conditional dependency, and its corresponding graphical model is called a Markov network.

On the other hand, marginal dependency approach is based on $\boldsymbol{\Sigma}$ rather than $\boldsymbol{\Sigma}^{-1}$. Butte *et al.* (2000) studied interactions between genes in a graphical model by estimating $\boldsymbol{\Sigma}$ which represents covariance between genes. Chaudhuri *et al.* (2007) suggested a method of estimating $\boldsymbol{\Sigma}$ when zero patterns of the graph is prespecified. Rothman *et al.* (2010) studied a shrinkage method to obtain a sparse estimate of covariance, and Bien and Tibshirani (2011) suggested a penalized likelihood method for estimating $\boldsymbol{\Sigma}$. Recently, Cai and Yuan (2012) proposed the method of estimation of covariance matrix by block thresholding when $p >> n$.

Chandrasekaran *et al.* (2012) gave natural conditions under which latent variable graphical models are identifiable.

4. Issues in network analysis

4.1. Clustering and classification

The $p \gg n$ (small n, large p) problem often occurs in cDNA microarray data representing the gene expression, and they can be expressed as $p \times n$ matrix

$$\boldsymbol{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{pmatrix}$$

where x_{ij} denotes genetic information for the ith gene and the jth sample. Usually, p is thousands and n is tens. For example, x_{ij} represents gene expression of the ith gene and the jth patient in leukemia. With this example, we might be interested either in clustering or classification depending on type of data - unsupervised or supervised, respectively. There are numerous methods on clustering and classification; see Speed (2003) among others for the analysis of microarray data. Recently, Kim *et al.* (2008) suggested a simultaneous approach to clustering and classification. In microarray data, one may be interested in two ways of inference. To do this row-wise normalization is required, i.e., $\sum_j^n x_{ij} = 0$ and $\sum_j^n x_{ij}^2 = 1$. If we let row-wise normalized matrix be $\boldsymbol{X}^*$, then the corresponding adjacency matrix is either $\boldsymbol{X}^*\boldsymbol{X}^{*T}$ for the marginal dependency or the inverse of $\boldsymbol{X}^*\boldsymbol{X}^{*T}$ for the conditional dependency. One is inference on variables, i.e., investigating networks of genes such as finding hub genes and clustering of genes.

The other is inference on samples and the primary interest is classification or clustering of patients. To do this column-wise normalization is required, i.e., $\sum_i^n x_{ij} = 0$ and $\sum_i^n x_{ij}^2 = 1$. If we let column-wise normalized matrix be $\tilde{\boldsymbol{X}}$, then the corresponding adjacency matrix is either $\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{X}}$ for the marginal dependency or the inverse of $\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{X}}$ for the conditional dependency.

4.2. Detection of hub

As far as we know, there does not exist an explicit and rigorous definition of hub in networks literature. In fact, in most literature, hub is loosely defined as, for example, vertices with unusually high degree (Newman, 2010). Some relevant measures such as centrality were studied and suggested by Katz (1953) and Bonacich (1987), however, it is somewhat different measure. Therefore, it is highly required for a mathematical definition of hub, and it would be useful if explicit tools are developed for detection of hubs if they exist. Also, if there exist hubs in a given network, then the estimation of the number of hubs are in the given network is also very important.

As a relevant result for the detection of hub, we introduce preliminary results on the relationship between degrees and eigenvalues. There are interesting relationships between degrees and eigenvalues of the Laplacian matrix $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$. For simplicity, let the ordered degrees be

$$d_1 \geq d_2 \geq \cdots \geq d_p.$$

Also, let the eigenvalues of $\boldsymbol{L}$ be ordered as

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{p-1} > \lambda_p = 0.$$

The first relation can be obtained by the spectral decomposition of $\boldsymbol{L}$, i.e.,

$$\boldsymbol{L} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^t = \sum_{i=1}^{p} \lambda_i \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i^t$$

where $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_p)$, $\boldsymbol{\Lambda} = diag\,(\lambda_1, \lambda_2, \cdots, \lambda_p)$. Here, $\boldsymbol{\gamma}_i$ is the ith eigenvector corresponding to the ith eigenvalue λ_i. Then, by comparing the jth diagonal component of both sides, we have

$$d_j = \sum_{i=1}^{p} \lambda_i \gamma_{ij}^2.$$

The second relationship is (Mieghem, 2010)

$$\sum_{j=1}^{k} d_j \leq \sum_{j=1}^{k} \lambda_j, \quad k = 1, 2, \cdots, p,$$

which is very interesting because

$$\sum_{j=1}^{p} d_j = \sum_{j=1}^{p} \lambda_j.$$

5. Concluding remarks

In this paper we reviewed recent developments in the network analysis using the graph theory. We introduced interesting research area in the network analysis with some important theoretical results. Especially, we reviewed marginal and conditional approach in estimating the adjacency matrix, and argued that methods of defining and finding hubs in a network are highly recommended. Studies on the the Tracy-Widom law should be pursued in the case of the scaled inverse correlation matrix and the Laplacian matrix. Also, for the spiked distribution, in which the variance of some variables are far away from 1, the corresponding Tracy-Widom law is also worth pursuing.

References

Bickel, P. J. and Levina, E. (2008). Covariance regularization by thresholding. *The Annals of Statistics*, **36**, 2577-2604.

Bien, J. and Tibshirani, R. J. (2011). Sparse estimation of a covariance matrix. *Biometrika*, **98** , 807-820.

Birnbaum, A., Johnstone, I. M., Nadler, B. and Paul, D.(2013). Minimax bounds for sparse PCA with noisy high-dimensional data. *The Annals of Statistics*, **41**, 1055-1084.

Bonacich, P. (1987). Power and centrality: A family of measures. *The American Journal of Sociology*, **92**, 1170-1182.

Butte, A. J., Tamayo, P., Slonim, D., Golub, T. R. and Kohane, I. S.(2000). Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. *Proceedings of the National Academy of Sciences*, **97**, 12182-12186.

Cai, T. T. and Liu, W. (2011). A direct estimation approach to sparse linear discriminant analysis. *Journal of the American Statistical Association*, **106**, 1566-1577.

Cai, T. T. and Yuan, M. (2012). Adaptive covariance matrix estimation through block thresholding. *The Annals of Statistics*, **40**, 2014-2042.

Cai, T. T. and Zhou, H. (2012). Minimax estimation of large covariance matrices under $L1$ norm (with discussion). *Statistica Sinica*, **22**, 1319-1378.

Chandrasekaran, V., Parrilo, P. A. and Willsky, A. S. (2012). Latent variable graphical model selection via convex optimization. *The Annals of Statistics*, **40**, 1935-1967.

Chaudhuri, S., Alur, R. and Cerny, P. (2007). Model checking on trees with path equivalences. *Tools and Algorithms for the Construction and Analysis of Systems Lecture Notes in Computer Science*, **4424**, 664-678.

Dempster, A. P. (1972). Covariance selection. *Biometrics*, **28**, 157-175.

Edward, D. (2000). *Introduction to graphical modelling,* 2nd ed., Springer, New York.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal component analysis. *The Annals of Statistics*, **29**, 295-327.

Johnstone, I. M. (2008). Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy-Widom limits and rates of convergence. *The Annals of Statistics*, **36**, 2638-2716.

Katz, L.(1953). A new status index derived from sociometric analysis. *Psychometrika*, **18**, 393.

Kim, C., Cheon, M., Kang, M. and Chang, I. (2008). A simple and exact Laplacian clustering of complex networking phenomena: Application to gene expression profiles. *Proceedings of the National Academy of Science*, **105**, 4083-4087.

Marcenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Mathmatics of the USSR - Sbornik*, **1**, 457-483.

Meinshausen, N. and Buhlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, **34**, 1436-1462.

Mieghem, P. V. (2010). *Graph spectra for complex networks*, Cambridge University Press, New York.

Newman, M. (2010), *Networks; An introduction*, Oxford University Press, New York.

Peng, J., Wang, P., Zhou N. and Zhu, J. (2009). Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association*, **104**, 735-746.

Pillai, N. S. and Yin, J. (2012). Edge universality of correlation matrices. *The Annals of Statistics*, **40**, 1737-1763.

Rothman, A. J., Levina, E. and Zhu, J. (2009). Generalized thresholding of large covariance matrices. *Journal of the American Statistical Association*, **104**, 177-186.

Rothman, A. J., Levina, E. and Zhu, J. (2010). Sparse multivariate regression with covariance estimation. *Journal of Computational and Graphical Statistics*, **19**, 947-962.

Speed, T. (2003). *Statistical analysis of gene expression microarray data*, Chapman and Hall, New York.

Tracy, C. A. and Widom, H. (1996). On orthogonal and symplectic matrix ensembles. *Communications in Mathematical Physics*, **177**, 727-754.

Tracy, C. A. and Widom, H. (2000). The distribution of the largest eigenvalue in the Gaussian ensembles; $\beta = 1, 2, 4$. *CRM Series in Mathematical Physics*, **4**, 461-472.

Whittaker, J. (1990). *Graphical models in applied mathematical multivariate statistics*, John Wiley & Sons. New York.

Wigner, E. P. (1955). Characteristic vectors of bordered matrices with infinite dimensions. *The Annals of Mathematics*, **62**, 548-564.

Won, D. and Choi, K. (2014). Network analysis and comparing citation index of statistics journals. *Journal of the Korean Data and Information Science Society*, **25**, 317-325.

Yuan, M. and Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model, *Biometrika*, **94**, 19-35.