# ESTIMATION OF NETWORK FROM MEANINGFUL PART OF CROSS-CORRELATION MATRIX

**W. Souma**[a], **J. Maskawa**[b], and **K. Fukuda**[c]

[a]NiCT/ATR CIS Applied Network Science Laboratory
2-2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0288, Japan (ATR)
souma@atr.jp

[b]Department of Business Administration, Fukuyama Heisei University
Fukuyama, Hisochima, 720-0001, Japan
maskawa@heisei-u.ac.jp

[c]National Institute of Informatics
2-1-2, Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430, Japan
kensuke@nii.ac.jp

When we investigate networks, we must care about missing links. Almost all of network data is incomplete, therefore there are missing links. Characteristics of a network are decisively changed by a lack of one link. This susceptibility to a lack of links distinguishes statistics of quantities having relation with networks from that having no relation with networks. A lack of data is not so problem when we are interested in the distribution of quantities having no relation with networks. For example, in the case of the distribution of the height of human being, a lack of data for one person does not decisively change the distribution.

In some systems, even if network data is incomplete, dynamics of nodes are recorded as time series data. We can construct a cross-correlation matrix from time series data of each node. It is natural to consider that the cross-correlation matrix reflects the underlying network structure. Hence it is expected that we can estimate the underlying network from the cross-correlation matrix, and that we can estimate missing links. However the cross-correlation matrix corresponds to an weighted adjacency matrix of a complete graph. In such a complete graph, many links are meaningless, and are originated from noise components of the cross-correlation matrix. Hence, it is necessary to filter out noise components from the cross-correlation matrix.

We construct the cross-correlation matrix for the time series data of internet traffic in the Science Information NETwork (SINET) in Japan [1], and calculate eigenvalues and eigenvectors of the cross-correlation matrix. By comparing the distribution of eigenvalues and eigenvectors with the theoretical distribution functions derived from the random matrix theory, we extract the meaningful part of the cross-correlation matrix [2]. In addition, we decompose the cross-correlation matrix, $\mathbf{C}$, as

$$\mathbf{C} = \mathbf{C}_{\mathrm{w}} + \mathbf{C}_{\mathrm{g}} + \mathbf{C}_{\mathrm{n}},$$

where $\mathbf{C}_{\mathrm{w}}$, $\mathbf{C}_{\mathrm{g}}$, and $\mathbf{C}_{\mathrm{n}}$ indicate the whole effect, the group structure, and the random noise, respectively [3], and we call them as pseudo-correlation matrices.

As the first step to estimate networks from the cross-correlation matrix and the pseudo-correlation matrices, we apply the method of minimum spanning tree (MST). Figures 1 (a), (b), (c), and (d) show MSTs corresponding to $\mathbf{C}$, $\mathbf{C}_{\mathrm{w}}$, $\mathbf{C}_{\mathrm{g}}$, and $\mathbf{C}_{\mathrm{n}}$, respectively. We can find the hub in Fig. 1 (b). This hub corresponds to the node that connects SINET with other

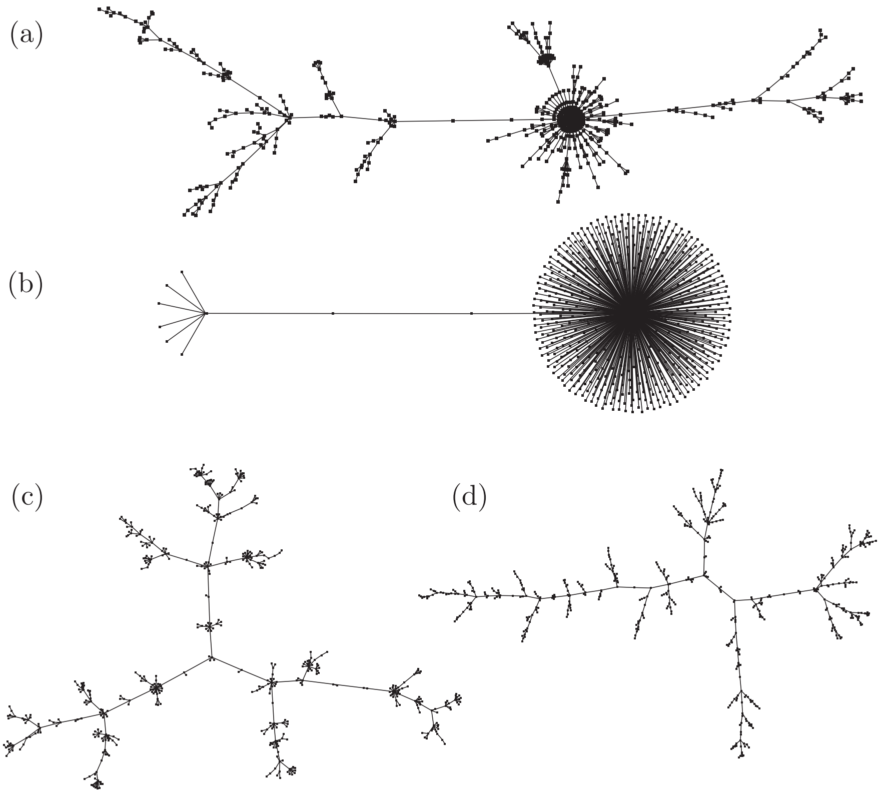


Figure 1: The MSTs corresponding to $\mathbf{C}$, $\mathbf{C}_{\rm w}$, $\mathbf{C}_{\rm g}$, and $\mathbf{C}_{\rm n}$.

autonomous systems. We can also find the group structure in Fig. 1 (c). On the other hand, we cannot find special structure in Fig. 1 (d).

We apply same method to the time series data of stock market and earthquake, and estimate corresponding networks. To estimate networks, we also apply the method of threshold [4] and the method of the planar maximally filtered graph [5].

**Keywords**
cross-correlation matrix, random matrix theory, complex network, network estimation

**References**
[1] Science Information NETwork (SINET), http://www.sinet.ad.jp/
[2] V. Plerou, et al., "Random matrix approach to cross correlations in financial data," *Phys. Rev. E*, v. 65, p. 066126, 2002.
[3] D.-H. Kim, and H. Jeong, "Systematic analysis of group identification in stock markets," *Phys. Rev. E*, v. 72, p. 046133, 2005.
[4] J.-P. Onnela, et al., "Dynamic asset trees and portfolio," *Eur. Phys. J. B*, v. 30, p. 285-288, 2002.
[5] M. Tumminello, et al., "A tool for filtering information in complex systems," *PNAS*, v. 102, p. 10421-10426, 2005.