Name of the Presenter: Eonyong Han
Title: Topological normalization of pathway unit correlational network for fair comparative study

Affiliation of the Presenter: Department of Computer Science and Engineering, Kyungpook National University, Daegu, South Korea

Topological normalization of pathway unit correlational network for fair comparative study

Eonyong Han[1], Inuk Jung[1*]

[1] School of Computer Science and Engineering, Kyungpook National University, Daegu

[*]*Corresponding authors:*
E-mail: inukjung@knu.ac.kr; phone.: +82539505552

Abstract. Biological pathways are frequently used to explain or comprehend the biological mechanisms of a disease or a certain phenotype. Methods, such as GSVA or GSEA, are utilized to measure the activity of pathways in means to cluster or classify sample groups. Pathways are comprised of proteins that are related to each other under a specific biological mechanism in action. Thus, the discovery of such relational data often involves network analysis. For example, in the protein level, the protein-protein interaction data is used. In the RNA level, gene co-expression network analysis methods, such as WGCNA, is used to figure out important gene regulatory relationships.

From such network analysis, we expect to find certain gene sets, or pathways, that well explain a clinical feature of interest. However, constructing such a gene co-expression network several aspects need to be considered. First, a single large network is difficult to interpret in terms of association with the clinical features, since the outcome may be too broad. Second, if a network is constructed independently for each group of interest, it is a non-trivial task to compare the two networks in a fair manner. This is because the two networks may significantly differ in size or topology, which may incur a bias in the result.

Here, we made effort to provide a way to normalize the networks that are to be compared for identifying important changes in gene-gene correlation, which are expected to explain association between the groups and their clinical features. In sense of biology, such co-expression networks are known to follow a scale-free network that has topological figure of the power law degree distribution. Therefore, we adjusted network topology fine-tuning for a supposed network size at informational entropy of network degree distribution.

We hypothesize that it is important to match the comparing networks in terms of network size and topology. A simple way is to shrink the larger network to a comparable size to the relative smaller network, which is to be compared with. This can be achieved by setting the power value on the adjacency matrix of a gene-gene correlation data as a penalty term. The larger the difference is between the networks, a larger power value may be introduced. We tested a series of data sets to investigate whether a general power value exists and tested if the normalized networks retained the biological features that are well known in the literature. As a basis network, we constructed a *normal* network using data from healthy peoples, which was compared to gene co-expression networks of the four colon cancer subtypes, CMS1, CMS2, CMS3 and CMS4.

As a result, we were able to observe that the normalization of the network was valid by evaluating the genes retained in the normalized, or shrunken, networks in terms of the well known biological mechanisms of the COAD subtypes. We are now expanding the study to analyze the networks in units of pathways, which we term as pathway unit correlational network (PUCN).

Keywords: *Pathway analysis; Network analysis; PUCN; Batch effect; Network topology; Entropy;*

Acknowledgements: This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2020M3C9A5085604).