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

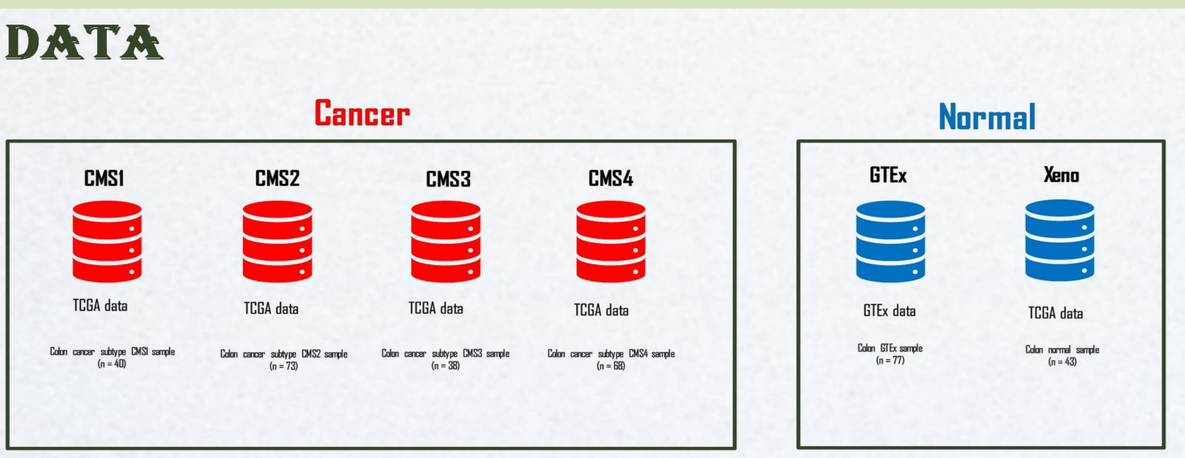
Abstract

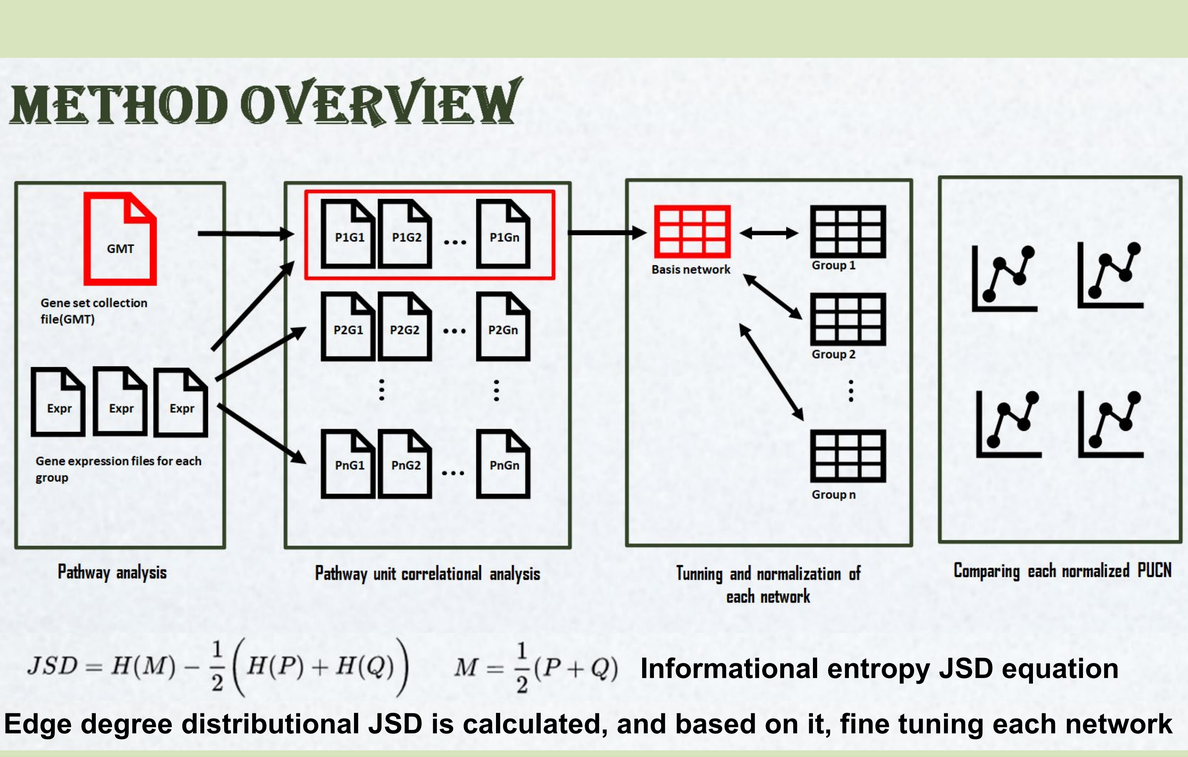
Biological pathway analysis is one of the most important approaches to explain a disease or an interesting phenotype and this method is authorized by GSVA that is broadly utilized for that way [1]. This pathway methodology is based on proteins are related at biological manner and it could be represented as network analysis. At this perspective of relation, another well known method such as WGCNA also exists [2]. We tried to obtain the network of target features. However, it was often too large to explain or too biased to make an fair comparison. Therefore, 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, so we adjusted network topology fine-tuning for a supposed network size at informational entropy of network degree distribution. 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 [3][4]. 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).

Introduction

Pathway unit correlational network(PUCN) analysis needs gene set collection. For that reason, we can choose gene set collection subset to evaluate target clinical features properly. However, first obtained correlational network to evaluate with each other group was often too large and there were many meaningless edges. 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. On top of that, recent research focus on network topology or dynamics to manufacturing their target network through various methods such as centrality measurement, and informational entropy. Among them, we adjusted an approach for perspective of informational entropy, and it clearly lead to normalization of networks and a fair comparison was possible.

Material and Methods

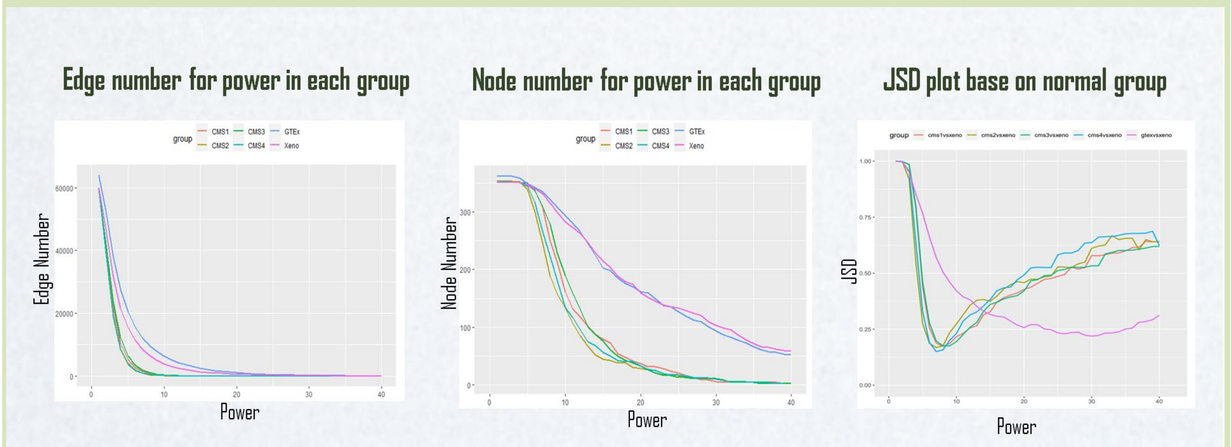


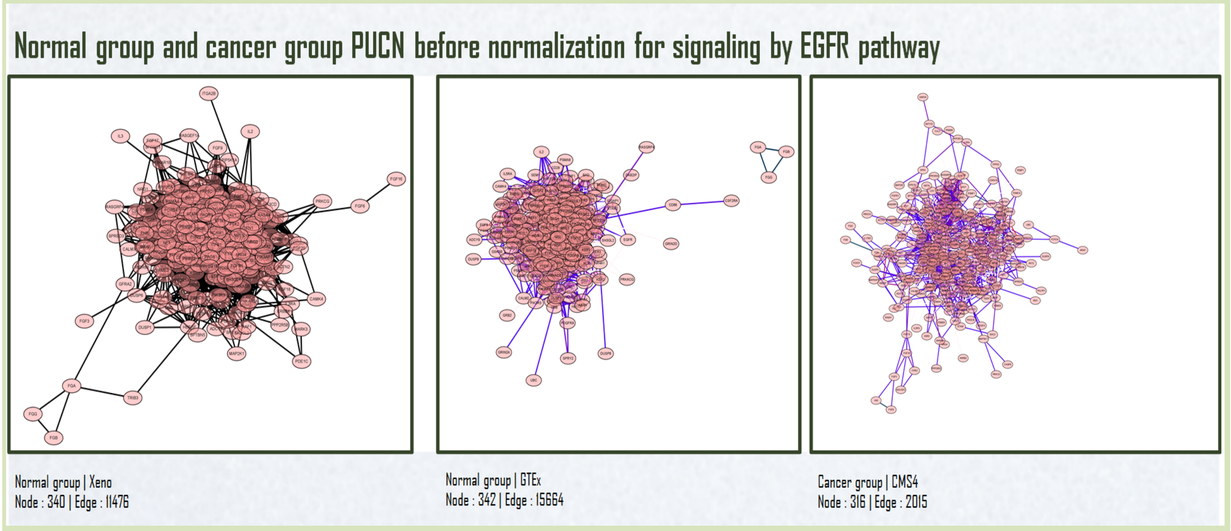


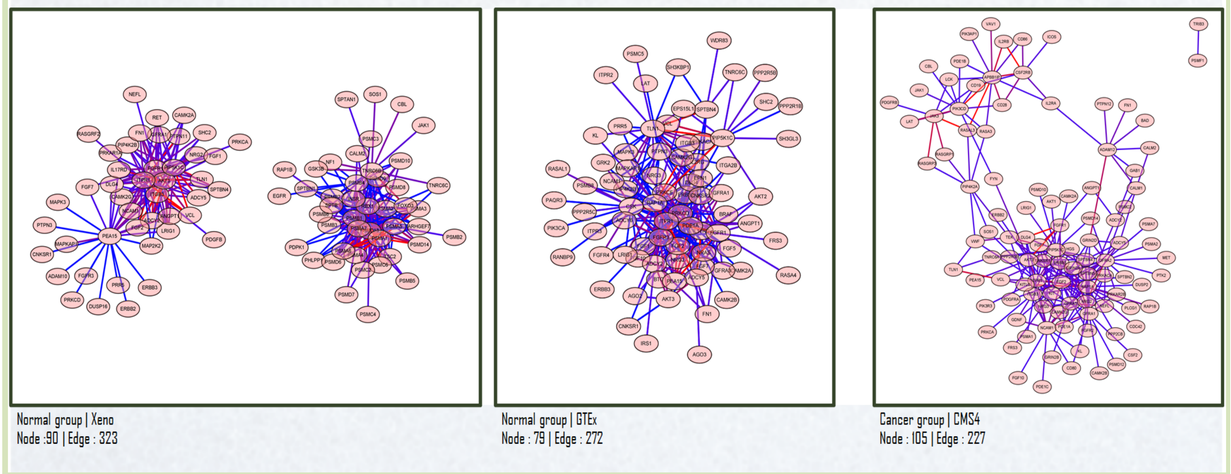
$$JSD = H(M) - \frac{1}{2}\left(H(P) + H(Q)\right) \qquad M = \frac{1}{2}(P + Q)$$ **Informational entropy JSD equation**

Edge degree distributional JSD is calculated, and based on it, fine tuning each network

Results







Conclusion

In pathway unit correlational network study, network informational bias that contain batch effect can interrupt fair comparison. Base on J-S divergence, shrinking network through reasonable penalty can remove batch effect of each different group. This normalized networks seems to overcome biased sample issue, and be utilized to explain specific features of target disease. We are going to expand network analysis in terms of PUCN for various types of diseases.

References

[1] Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics* **14,** 7 (2013). https://doi.org/10.1186/1471-2105-14-7

[2] Langfelder, P., Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9,** 559 (2008). https://doi.org/10.1186/1471-2105-9-559

[3] Cancer Genome Atlas Research Network, Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R., Ozenberger, B. A., Ellrott, K., Shmulevich, I., Sander, C., & Stuart, J. M. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. *Nature genetics*, *45*(10),1113–1120.

[4] Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young,N.,et al.: The genotype-tissue expression (GTEx) project. Nature genetics45(6), 580–585 (2013)

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).