MONET2: Development of HPC-based genome-scale regulatory network inference system for investigating Alzheimer's disease genes

Younghoon Kim
Korea Institute of Science and Technology Information (KISTI), Daejeon, South Korea

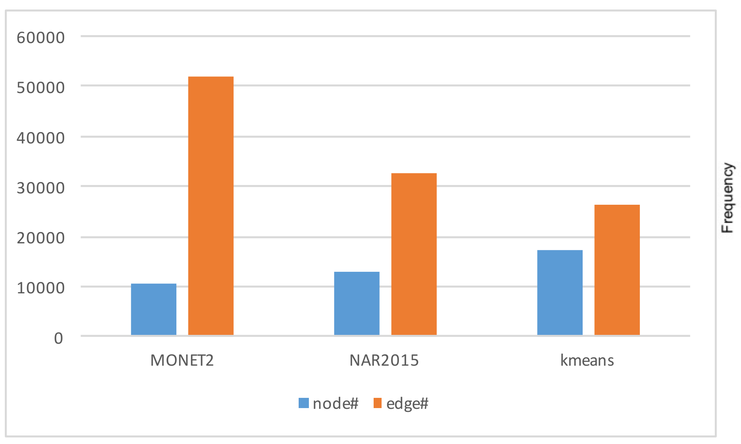
Abstract

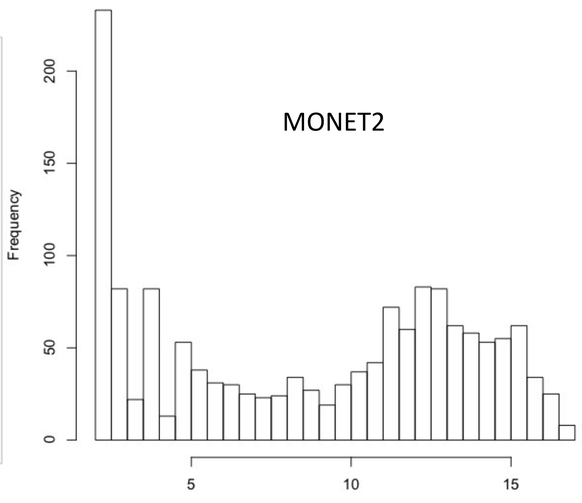
Inferring gene regulatory networks are important in investigating disease genes and to understand overall perspectives and mechanisms for diseases, global gene regulatory networks are needed. However, inferring global networks require tremendous computations and huge information content, and the inference without these requirements results in infeasible execution in reasonable time and inaccurate networks. To solve this problem, we had developed MONET technique, MOdularized NETwork learning, in which computational burden can be overcome by divide-and-conquer strategy with High Performance Computing(HPC)-resources and insufficiency of information is tackled by integrating the existing biological information. In this study, we have developed MONET2, which extends to Homo sapiens and enables more accurate resulting networks by integrating multiple biological information from genomics, epigenomics, interactome, disease gene data, and so on. We used MONET2 to construct the global networks of Alzheimer's disease by incorporating more than eleven biological data sources in thirteen brain cell types. MONET2 enables the inference of accurate global networks in a feasible time by using parallel processing techniques based on HPC-resources of KISTI and we plan to open this method to the public like the first version of MONET which was developed as a plugin of Cytoscape.

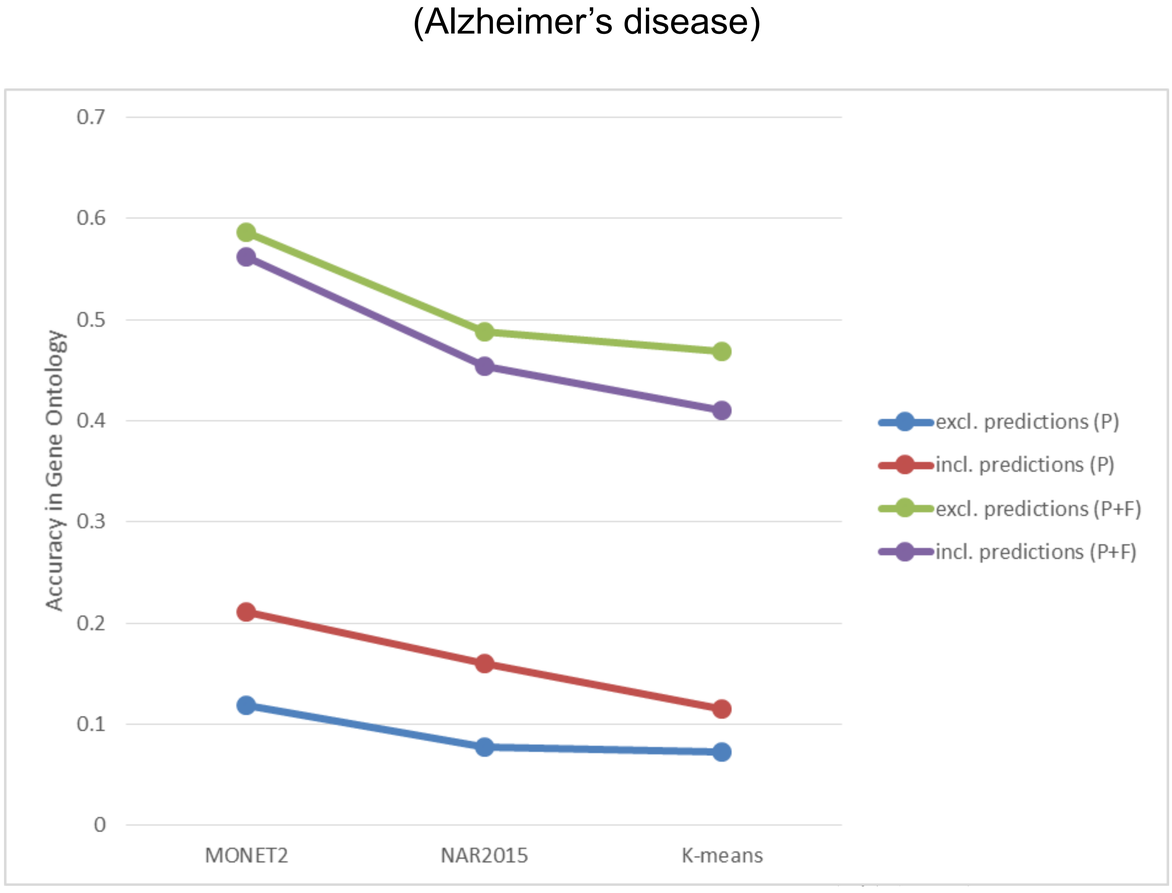
Methods & Results

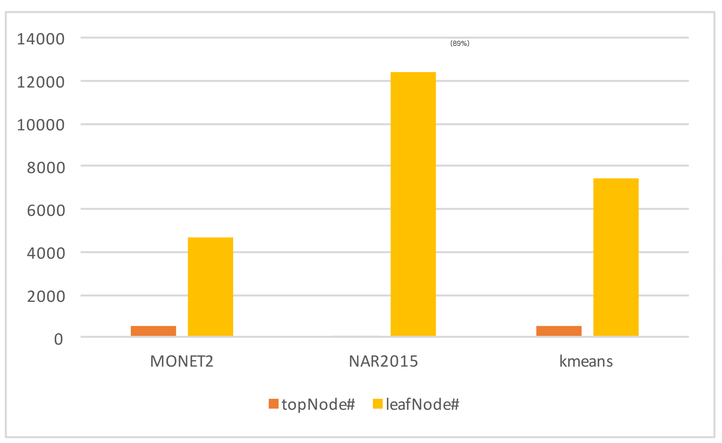
MONET2 algorithm consists of four steps; 1) seed extraction 2) seed expansion 3) network inference for each cluster 4) combine each cluster into single global network. First, it selects seed genes which will be basic genes for each cluster by using disease-related genesets for Alzheimer's from various databases (GWAS catalog, OMIM, AlzGene, COSMIC, DEG list). Second, the algorithm expands the seed gene sets by adding the close gene pairs. The closeness measure is based on various dimensions including TF-target relationship, functional gene-gene relationship, and expression-based gene-gene relationship. TF-target relationship information came from ENCODE, Chromatin Interaction information, and TF motif information. For functional gene-gene relationship HumanNet data was used, and for expression-based gene-gene relationship microarray data of Alzheimer's patients were used. Then, it performs network inference for each cluster which is the developed gene sets from the previous steps. The network inference is done by Bayesian network inference with Markov chain Monte Carlo(MCMC) method. Finally, the algorithm merges all sub-networks into single global network through intermediate genes among clusters.

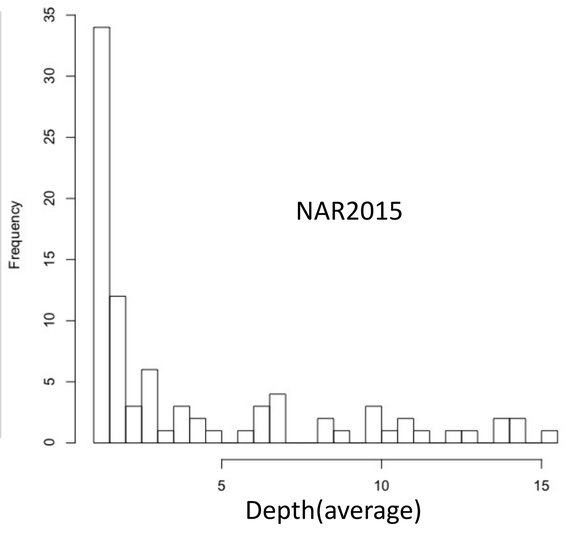
The resulting network showed good characteristics as a global gene regulatory network. Whereas the previous method (NAR2015) had too many leaf nodes and relatively small number of edges, MONET2 network had appropriate number of leaf nodes and relatively large number of edges, showing MONET2's ability to generate a global regulatory network properly. Also, in the validation experiment with Gene Ontology, MONET2 showed better accuracies than the previous method.











<Test if the gene pairs have the same GO terms>

Reference

1. Global transcription network incorporating distal regulator binding reveals selective cooperation of cancer drivers and risk genes, Nucleic Acids Research, 2015
2. Modularized learning of genetic interaction networks from biological annotations and mRNA expression data. Bioinformatics, 2005