# Gene expression biclustering with FIrefly Algorithm

Denilson Oliveira Melo[1], Paulo Eduardo Ambrósio[1],

*1 Universidade Estadual de Santa Cruz*

## Abstract

Gene expression profiling is the measurement and study of expression levels of various genes at once, with the intent to discover and understand gene function. Gene expression data is usually presented in an expression matrix, which can be analyzed with computational and statistical methods. One way to analyze this expression data is through clustering techniques, which aim to group genes of similar expression tendencies together, suggesting that these genes are subject to a common pattern of regulation. Clustering techniques however are limited, they are only able to create groups considering the entire set of conditions, but genes are not necessarily related to every condition, and they also doesn't allow coupling, thus ignoring the possibility that an individual gene may be related to multiple groups in different subsets of conditions. This limitation can be surpassed with biclustering techniques, which aim to group both genes and conditions together. With biclustering groups are formed considering each subset of condition, thus allowing more grouping possibilities that better represent the relationships between genes. Biclustering however is a NP-hard problem. Problems of this class have no exact solution and are usually solved using heuristics. Several metaheuristics are already being used to find biclusters and in this preliminary study we propose to explore and utilize one of nature-inspired metaheuristic proposed by Xin-She Yang - the Firefly Algorithm (FA) - a swarm based approach inspired from the behaviour of fireflies revolving around their blinking patterns. This metaheuristic considers the attraction between fireflies, where given two individuals, the less bright one will be attracted to the brightest. With little effort one could see that with these parameters groups of fireflies would be formed naturally, and that's the basic idea. In this preliminary study we aim to employ the grouping characteristics of the algorithm to find biclusters of gene expression data.