**MONET**: a toolbox integrating top-performing methods for network modularisation

Mattia Tomasoni¹,²,* Sergio Gómez³, Jake Crawford⁴,⁵, Weijia Zhang⁶, Sarvenaz Choobdar¹,², Daniel Marbach⁷ and Sven Bergmann¹,²,⁸*  

¹Department of Computational Biology, University of Lausanne, Lausanne, Switzerland  
²Swiss Institute of Bioinformatics, Lausanne, Switzerland  
³Departament d’Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, Tarragona, Catalonia, Spain  
⁴Department of Computer Science, Tufts University, MA, USA  
⁵Graduate Group in Genomics and Computational Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA  
⁶School of Information Technology and Mathematical Sciences, University of South Australia, Adelaide, Australia  
⁷Roche Pharma Research and Early Development, Pharmaceutical Sciences, Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd, 4070 Basel, Switzerland  
⁸Department of Integrative Biomedical Sciences, University of Cape Town, Cape Town, South Africa

* To whom correspondence should be addressed.

**Abstract**

**Summary:** We define a disease module as a partition of a molecular network whose components are jointly associated with one or several diseases or risk factors thereof. Identification of such modules, across different types of networks, has great potential for elucidating disease mechanisms and establishing new powerful bio-markers. To this end, we launched the "Disease Module Identification (DMI) DREAM Challenge", a community effort to build and evaluate unsupervised molecular network modularisation algorithms (Choobdar et al., 2018). Here we present **MONET**, a toolbox providing easy and unified access to the three top methods from the DMI DREAM Challenge for the bioinformatics community.

**Availability and Implementation:** **MONET** is a command line tool for Linux, based on Docker and Singularity containers; the core algorithms were written in R, Python, Ada and C++. It is freely available for download at https://github.com/BergmannLab/MONET.git

**Contact:** mattia.tomasoni@unil.ch (MT); sven.bergmann@unil.ch (SB)

**1 Introduction**

Gene networks, such as protein interaction, signalling, gene co-expression and homology networks, provide scaffolds of linked genes. Sub-networks, or modules, include genes normally acting in concert but whose joint function may be disrupted, if any of its members is missing, or dis-regulated. For Disease Modules this disruption can lead to a disease phenotype. The identification of such modules is therefore useful for elucidating disease mechanisms and establishing new bio-markers and potential therapeutic targets. Yet, which methods work best to extract such modules from different types of networks is not well understood. This prompted us to initiate the "Disease Module Identification (DMI) DREAM Challenge" (Choobdar et al., 2018), providing an unbiased and critical assessment of 75 contributed module identification methods. Our method evaluation used summary statistics from more than 200 disease relevant Genome-wide Association Studies (GWAS) in conjunction with our Pascal tool (Lamparter et al., 2016), avoiding the bias of using annotated molecular pathways.

The top-performing methods implemented novel algorithms that advanced the state of the art, clearly outperforming off-the-shelf tools. We therefore decided to make the top three methods available for the bioinformatics community in a single user-friendly package: **MONET** is a command line tool based on Docker and Singularity virtualization technologies, automatically installing the tool with all its dependencies inside a container, avoiding time-consuming and error-prone manual installations of computing environments and libraries. All computations then take place in this sandbox environment and once the output is ready, all resources can be fully released bringing the user’s machine back to its original state.
Markov Cluster Algorithm (Newman, 2004), Tabu search (Arenas and Arenas, 2005), Spectral optimization (Newman, 2006), Fast algorithm optimized using an ensemble of algorithms: Extremal optimization (Duch, 2008), and fine-tuning by (Satuluri et al., 2018).

2.1 K1: Top method using kernel clustering
K1 is based on the “Diffusion State Distance” (DSD), a novel graph metric which is built on the premise that paths through low-degree nodes are stronger indications of functional similarity than paths that traverse high-degree nodes by Cao et al. (2014). The DSD metric is used to define a pairwise distance matrix between all nodes, on which a spectral clustering algorithm is applied. In parallel, dense bipartite sub-graphs are identified using standard graph techniques. Finally, results are merged into a single set of non-overlapping clusters.

BLOG: https://www.synapage.org/#/Synapage:sys7349492/wiki/407359

2.2 M1: Top method using modularity optimization
M1 employs an original technique named Multiresolution introduced by Arenas et al., 2008 to explore all topological scales at which modules may be found. The novelty of this approach relies on the introduction of a parameter, called resistance, which controls the aversion of nodes to form modules. Modularity (Newman and Girvan, 2004; Arenas et al., 2008; Girvan and Newman, 2002). Analysis of the structure of complex networks at different resolution levels. New Journal of Physics, 6(5), 053039.

Cao, M., Pietras, C. M., Feng, X., Doroshchuk, K. J., Schuller, T., Park, J., Zhang, H., Coven, L. J., and Hescott, B. J. (2014). New directions for diffusion-based network prediction of protein function: incorporating pathways with confidence. Bioinformatics, 30(12), 2219-2227.

Choudhary, S., Ahlen, M. E., Crawford, J., Tomasoni, M., Lamparter, D., Lin, J., Hescott, B., Hu, X., Mercer, J., Natoli, T., et al. (2018). Open community challenge reveals molecular network modules with key roles in diseases. bioRxiv, page 265353 (ACCEPTED for publication in Nature Methods).

Duch, J. and Arenas, A. (2005). Community detection in complex networks using extremal optimization. Physical Review E, 72(2), 027104.

Kurtzer, G. M., Sochat, V., and Bauer, M. W. (2017). Singularity: Scientific containers for mobility of compute. PLOS ONE, 12(5), 1-20.

Lamparter, D., Marbach, D., Raeder, R., Katalík, Z., and Brglez, S. (2016). Fast and rigorous computation of gene and pathway scores from omic summary statistics. PLOS Computational Biology, 12(1), 1-20.

Lancichinetti, A., Fortunato, S., and Radicchi, F. (2008). Benchmark graphs for testing community detection algorithms. Phys. Rev. E, 76(4), 066110.

Merkel, D. (2014). Docker: Lightweight linux containers for consistent development and deployment. Linus J., 2014.239.

Newman, M. E. (2004). Fast algorithm for detecting community structure in networks. Physical Review E, 69(6), 066133.

Newman, M. E. (2006). Modularity and community structure in networks. Proceedings of the National Academy of Sciences USA, 103(23), 8577-8582.

Newman, M. E. and Girvan, M. (2004). Finding and evaluating community structure in networks. Physical Review E, 69(2), 026113.

Satuluri, V., Parthasarathy, S., and Ucar, D. (2010). Markov clustering of protein interaction networks with improved balance and scalability. In Proceedings of the First ACM International Conference on Bioinformatics and Computational Biology, BCB ’10, pages 247-256, New York, NY, USA. ACM.