

This is a repository copy of *PAFway: pairwise associations between functional annotations in biological networks and pathways*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/154294/>

Version: Submitted Version

Monograph:

Mahjoub, Mahair and Ezer, Daphne (2020) *PAFway: pairwise associations between functional annotations in biological networks and pathways*. Working Paper.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Systems Biology

PAFway: pairwise associations between functional annotations in biological networks and pathways

Mahiar Mahjoub^{1, 2, 3} and Daphne Ezer^{2,4,5*}

¹Department of Mathematics, University of Cambridge, Cambridge, UK,

²The Alan Turing Institute, London, UK,

³Royal Prince Alfred Hospital, Central Clinical School, University of Sydney, Sydney, Australia,

⁴Department of Statistics, University of Warwick, Coventry, UK and

⁵Department of Biology, University of York, York, UK.

*To whom correspondence should be addressed: daphne.ezer@york.ac.uk.

Associate Editor: XXXXXXXX

Received on XXXXXX; revised on XXXXXX; accepted on XXXXXX

Abstract

Motivation: Large gene networks can be dense and difficult to interpret in a biologically meaningful way.

Results: Here, we introduce PAFway, which estimates pairwise associations between functional annotations in biological networks and pathways. It answers the biological question: do genes that have a specific function tend to regulate genes that have a different specific function? The results can be visualised as a heatmap or a network of biological functions. We apply this package to reveal associations between functional annotations in an *Arabidopsis thaliana* gene network.

Availability: PAFway is available in CRAN.

Contact: daphne.ezer@york.ac.uk

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Biological networks can be very large, dense and difficult to visualise and interpret. PAFway is a tool to interpret large, dense biological networks in the context of functional annotations, such as gene ontology (GO). Some methods that analyse GO enrichment within networks, such as BiNGO [Maere *et al.*, 2005], operate by partitioning the network into clusters and then finding functional enrichment within each cluster.

Another family of methods, called topological network enrichment methods, utilise the structure of the network to find gene ontology terms that are enriched in a network or subnetwork [Mitrete *et al.*, 2013]. The output of these algorithms is generally a ranked list of annotations, ordered by how much they are enriched in the network.

In contrast, PAFway finds *pairwise associations of functional annotations* in biological networks and pathways, which is calculated efficiently using the Fast Fourier Transform (FFT). The results can be illustrated either in the form of a heat map or as a network where the nodes in the graph are functional annotations. We apply this method to AraNet [Lee *et al.*, 2015], a gene network for *Arabidopsis thaliana*.

2 Methods

The PAFway function takes as input a directed network, with or without edge weights, and a list all the functional annotations associated with each node. We refer to each *edge type* as an ordered pair of functional annotations, representing the scenario where a gene with the first functional annotation regulates a gene with the second functional annotation. The output of PAFway is the probability of observing at least the observed number (or sum of edge weights) of each edge type, under a null model in which the functional annotations are randomly distributed in the network (after correcting for multiple hypothesis testing).

2.1 P-value of edge counts

Let us say that the frequency of the first functional annotation in the network is p_a and the second is p_b . The probability of observing an edge between p_a and p_b is $p_{a,b} = p_a p_b$ if they are randomly distributed in the network. The probability of observing n edges between the first and second functional annotation in a network with N edges is determined by a binomial distribution:

$$n \sim B(N, p_{a,b}) \quad (1)$$

This means that it is possible to determine the probability of observing at least n edges of a certain type by using the binomial test.

OXFORD