

SICOP: identifying significant co-interaction patterns

Andreas Spitz¹, Katharina A. Zweig^{2,*} and Emőke-Ágnes Horvát^{1,2,*}

¹Graph Theory and Network Analysis Group, Interdisciplinary Center for Scientific Computing (IWR), University of Heidelberg, 69115 Heidelberg, Germany and ²Department of Computer Science, University of Science and Technology Kaiserslautern, 67663 Kaiserslautern, Germany

Associate Editor: Gunnar Ratsch

ABSTRACT

Summary: Interactions between various types of molecules that regulate crucial cellular processes are extensively investigated by high-throughput experiments and require dedicated computational methods for the analysis of the resulting data. In many cases, these data can be represented as a bipartite graph because it describes interactions between elements of two different types such as the influence of different experimental conditions on cellular variables or the direct interaction between receptors and their activators/inhibitors. One of the major challenges in the analysis of such noisy datasets is the statistical evaluation of the relationship between any two elements of the same type. Here, we present *SICOP* (significant co-interaction patterns), an implementation of a method that provides such an evaluation based on the number of their common interaction partners, their so-called co-interaction. This general network analytic method, proved successful in diverse fields, provides a framework for assessing the significance of this relationship by comparison with the expected co-interaction in a suitable null model of the same bipartite graph. *SICOP* takes into consideration up to two distinct types of interactions such as up- or downregulation. The tool is written in Java and accepts several common input formats and supports different output formats, facilitating further analysis and visualization. Its key features include a user-friendly interface, easy installation and platform independence.

Availability: The software is open source and available at cna.cs.uni-kl.de/SICOP under the terms of the GNU General Public Licence (version 3 or later).

Contact: agnes.horvat@iwr.uni-heidelberg.de or zweig@cs.uni-kl.de

Received on March 1, 2013; revised on May 9, 2013; accepted on July 8, 2013

1 INTRODUCTION

High-throughput experiments resulting in interaction data require computational methods that enable the identification of their key interaction patterns. If the interaction is observed between elements of two different types, the data can be modelled as a bipartite graph in which the elements are represented by nodes and their interaction by edges. More than one type of interaction, e.g. up- and downregulation, can be represented this way by the edges of the graph.

Existing tools for the analysis of bipartite graphs like the R packages *bipartite* (Dormann *et al.*, 2009) or *networksis* (Admiraal and Handcock, 2008) are mainly tailored to the

purpose of understanding principles in community ecology. Systems biology applications pose three important challenges that these tools cannot cope with: (i) large-scale experiments that are often prone to noise, (ii) mild interaction effects that are difficult to detect and (iii) simultaneously observed distinct types of interactions.

A recently suggested method deals with these problems by using a network analytic approach (Uhlmann *et al.*, 2012; Malumbres, 2012). In this procedure, the number of common interaction partners of two elements is compared with its expected value in a randomized null model that maintains the number of interaction partners of all elements. The approach adopts and extends the general procedure for assessing the statistical significance of the so-called network motifs (Milo *et al.*, 2002). The underlying assumption is that two elements with a statistically significant number of common interaction partners share the same functional role according to the so-called guilt-by-association principle (Quackenbush, 2003). The method sketched in Figure 1 is based on the theoretical insights formulated in Zweig and Kaufmann (2011) and uses a more realistic null model than other existing approaches. The open source software *SICOP* (significant co-interaction patterns) implements the algorithm as an easy-to-use client-side tool. It is able to detect significant mild co-interaction effects by taking into account up to two different types of interactions.

2 SOFTWARE FEATURES

From interaction data between elements of two distinct types, *SICOP* computes the significance of the number of common interaction partners for all pairs of elements of one of the two types. The most important functionalities of *SICOP* are as follows.

Data import from diverse input file formats. The tool accepts a list of the observed interactions stored in a text file, a matrix containing the measured level of all interactions in a csv file (comma separated value) or a graph representation of the network in a gml file (graph mark-up language) as produced by graph editing programs such as yEd.

Simplex and duplex network support and precomputational edge filtering. Given the experimental data, *SICOP* first constructs a bipartite graph model (Fig. 1a and b). If the observed interactions are assigned weights, the user may add a threshold to filter them. If there are two types of interactions, the user can treat them as a single type (*simplex* network data) or use both of them (*duplex* network data).

*To whom correspondence should be addressed.

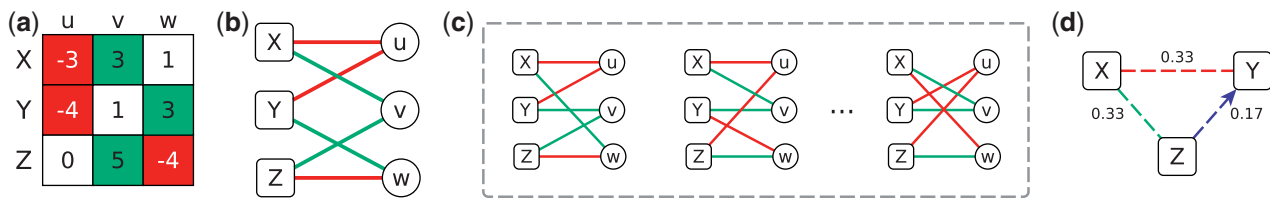


Fig. 1. Illustration of the method behind *SICOP*: (a) Heatmap showing the results of a hypothetical high-throughput experiment. Two types of interactions are distinguished based on the sign of their strength. (b) Bipartite graph representation of the thresholded interaction data. The two types of interactions are shaded differently. (c) Null model consisting of randomized graphs created from the original bipartite graph by permuting edges of the same type. (d) Resulting co-interaction network that contains three types of edges: two uni-type co-interaction patterns (undirected edges) and one mixed co-interaction pattern (directed edge). The edge labels show empirical P -values computed as the fraction of randomized graphs in which the given co-interaction appeared at least as often as in the original graph

Statistical significance assessment of co-interactions. *SICOP* detects patterns in the bipartite graph that are non-random when compared with a null model, which consists of a set of random graphs generated through a series of permutations (Gionis *et al.*, 2007). During permutation, the following key structural properties of the graph are maintained: the number of elements and the number of different types of interactions per element (Fig. 1c). In duplex networks, it is possible to differentiate between common interaction partners regarding all combinations of the two interaction types. On the example of up- and downregulation, the number of co-up-, co-down- and antagonistically regulated partners is evaluated for all pairs of elements. The statistical significance of the individual co-regulations is then quantified by a z -score or a P -value (Fig. 1d).

Multiple data export formats. The same edge selection options are available when exporting the data as when importing it. Thus, the user may create and store multiple networks with different threshold values corresponding to different P -values or z -scores. The obtained co-regulation networks may be exported in any of the input file formats or alternatively as Graphml.

High configurability. Besides the key functionalities described earlier, *SICOP* allows more confident users to modify the parameter values. The default values are based on theoretical and empirical considerations and are automatically adjusted to the size of the input data. Increasing the preset values enhances accuracy but comes at the cost of additional computational time. We refer interested readers to the manual provided with *SICOP*.

3 DISCUSSION

The algorithm was applied to the data from high-throughput screening experiments that show the effect of various microRNAs on a set of proteins that are highly relevant in breast cancer (Uhlmann *et al.*, 2012). MicroRNAs have recently been recognized to act as tumour suppressors and have an oncogene role. Uncovering their targeting patterns is difficult because they are known to only have a moderate effect on the protein

expression level. The extensive null model approach behind *SICOP* enabled the identification of key microRNAs that were shown in subsequent experiments to inhibit cell-cycle progression and proliferation. This specific result confirms that the algorithm behind *SICOP* is a powerful tool for the detection of statistically significant protein co-regulations. However, as described earlier, *SICOP* can be applied to a wide range of datasets, for instance, transcription factors binding to DNAs, gene-coding RNAs interacting with co-transcriptional non-coding RNAs, genes in relation with diseases or diseases and their symptoms. Our flexible tool offers an effortless way to explore such data to its full potential, considerably reducing the required amount of further experiments.

Funding: E.Á.H. is supported by the Heidelberg Graduate School of Mathematical and Computational Methods for the Sciences (HGS MathComp), University of Heidelberg, Germany, which is funded by the German Excellence Initiative (GSC 220).

Conflict of Interest: none declared.

REFERENCES

- Admiraal,R. and Handcock,M.S. (2008) networksis: a package to simulate bipartite graphs with fixed marginals through sequential importance sampling. *J. Stat. Softw.*, **24**, 1–21.
- Dormann,C.F. *et al.* (2009) Indices, graphs and null models: analyzing bipartite ecological networks. *Open Ecol. J.*, **2**, 7–24.
- Gionis,A. *et al.* (2007) Assessing data mining results via swap randomization. *ACM Trans. Knowl. Discov. Data*, **1**, 14.
- Malumbres,M. (2012) miRNAs versus oncogenes: the power of social networking. *Mol. Syst. Biol.*, **8**, 569.
- Milo,R. *et al.* (2002) Network motifs: simple building blocks of complex networks. *Science*, **298**, 824–827.
- Quackenbush,J. (2003) Microarrays—guilt by association. *Science*, **302**, 240–241.
- Uhlmann,S. *et al.* (2012) Global microRNA level regulation of EGFR-driven cell-cycle protein network in breast cancer. *Mol. Syst. Biol.*, **8**, 570.
- Zweig,K.A. and Kaufmann,M. (2011) A systematic approach to the one-mode projection of bipartite graphs. *Soc. Netw. Anal. Min.*, **1**, 187–218.