



Original software publication

diffeRenTES: An R package for computing cell differentiation trees from Boolean networks

Michele Braccini

Department of Computer Science and Engineering, Università di Bologna, via dell'Università, 50, Cesena, I-47521, Italy



ARTICLE INFO

Keywords:

Cell differentiation
 Differentiation trees
 Boolean networks
 R package
 Gene regulatory networks
 Dynamical systems

ABSTRACT

The analysis of possible cell differentiation scenarios of gene regulatory network models can represent an invaluable tool for testing existing hypotheses and developing new biological insights. For this reason, *diffeRenTES*, an R package that implements the abstractions of a prominent mathematical model of cell differentiation, is presented. *diffeRenTES* computes tree-like structures describing the progression of the cell differentiation process, starting from Boolean models of gene regulatory networks. Additionally, it offers the possibility to perform robustness analyses of Boolean networks.

Code metadata

Current code version	v0.3.2
Permanent link to code/repository used for this code version	https://github.com/SoftwareImpacts/SIMPAC-2023-8
Permanent link to Reproducible Capsule	https://codeocean.com/capsule/0946852/tree/v1
Legal Code License	GPLv3
Code versioning system used	Git
Software code languages, tools, and services used	R
Compilation requirements, operating environments & dependencies	R ≥ 3.3, BoolNet
If available Link to developer documentation/manual	https://mbraccini.github.io/diffeRenTES/
Support email for questions	braccini.michele@gmail.com

1. Introduction

Dynamical systems have been successfully employed to investigate and reproduce biological phenomena. Among these, Boolean networks (BNs) [1,2], thanks to their simplicity combined with the rich bouquet of complex behaviors they can exhibit, have proved capable of reproducing relevant properties and dynamics of cell differentiation [2–8]. In these works, the focus of the analysis is on the identification of the *attractors* – portions of the state space of a dynamical system that attract distinct trajectories – which can express gene expression patterns similar to those found in real cells. However, the description of the relations and constraints among the attractors [9–12], which ultimately determines the whole differentiation process, is often neglected or, if present, tailored to the real case at hand [13,14], and thus not generalizable.

Recently, a powerful abstract mathematical model able to reproduce the differentiation process of a cell modeled as a noisy Boolean

network subject to progressively decreasing noise has been presented [15–17]. This model, based on the Threshold Ergodic Set (TES) concept, proved to be able to replicate the main properties of biological differentiation, namely: *different degrees of differentiation*, *stochastic* and *deterministic differentiation* and *induced pluripotency*, just to cite a few.

Here *diffeRenTES* is presented, an R package that implements the abstractions of the above-mentioned differentiation model. It offers the possibility to compute the typical tree-like structures describing the global picture of the all possible differentiation pathways that a single cell can experience, starting from the Boolean networks and the relative synchronous attractors objects obtained through the well-known “BoolNet” R package [18]. The returned differentiation tree describes the process that, in principle, goes from the totipotent cells to the fully differentiated ones, where the switch among the different cell types is the result of the change in the intracellular noise level.

The code (and data) in this article has been certified as Reproducible by Code Ocean: (<https://codeocean.com/>). More information on the Reproducibility Badge Initiative is available at <https://www.elsevier.com/physical-sciences-and-engineering/computer-science/journals>.

E-mail address: m.braccini@unibo.it.

<https://doi.org/10.1016/j.simpa.2023.100470>

Received 9 January 2023; Received in revised form 17 January 2023; Accepted 17 January 2023

2. Software description

The *diffeRenTES* package is written using the R statistical software environment [19]. It is freely available on the Comprehensive R Archive Network (CRAN) at <https://CRAN.R-project.org/package=diffeRenTES> and can be installed using the following commands:

- `install.packages("diffeRenTES")`
- `library(diffeRenTES)`

The functions provided by the package together with the model abstractions they implement are described below.

2.1. Attractor transition matrix

As stated in the introduction, the cell differentiation model implemented by *diffeRenTES* is based on the asymptotic dynamics of BNs subject to noise. The kind of intracellular noise introduced in this model is a transient logic negation of a node's state, after which the dynamics of the Boolean network returns to follow the synchronous and deterministic updating scheme. In this way, we can observe jumps from one attractor to another. The *Attractor Transition Matrix* (ATM) gives us an estimate of the probabilities of the observed transitions between attractors, in the form of a square matrix of dimension n , where n is the number of attractors. It can also represent a robustness measure of the system with respect to a random flip of an arbitrary state [8].

To compute the ATM, the *diffeRenTES* package requires the previously computed set of attractors of a Boolean network. So, after having generated a Boolean network and having computed its synchronous attractors by means of the *BoolNet* package, the ATM can be computed by using the *getATM()* function.

getATM() accepts the following arguments:

- *net* — a Boolean network loaded or generated by using the “*BoolNet*” package.
- *synchronous_attractors* — the attractors of the Boolean network, obtained with the “*BoolNet*” package by simulating its dynamics with the synchronous update scheme.
- *MAX_STEPS_TO_FIND_ATTRACTORS* — maximum number of steps after which the search for the attractors stops. Its default value is 1000.

In the named list obtained as a result of calling the *getATM()* function we can find the computed ATM structure and the number of the lost flips, i.e., the number of perturbations that have not reached another attractor within the maximum number of steps.

2.2. Threshold ergodic set

At the top of the ATM it is possible to calculate the set of *Threshold Ergodic Sets* (TESs), that is, the abstractions that model the different cell types. A TES is a set of attractors that traps the long-term dynamics of the BN, under the hypothesis that attractor transitions with a probability less than a given threshold are not possible. Formally, a TES is a strongly connected component with no outgoing arcs of the resulting directed graph composed of nodes representing the attractors and arcs summarizing the transitions among them.

So, with low threshold values only the transitions having low probabilities are removed. This case models the condition of pluripotent cells since the BN dynamics can freely wander among the attractors. Instead, high threshold values cause the removal of even the most likely transitions, thus representing the case of mature differentiated cells, ideally trapped in one or few attractors. In short, by varying the threshold, which scales as the reciprocal of the noise, the TESs are computed and the differentiation process unfolds.

diffeRenTES computes the complete set of TESs of a BN by following this procedure: first, the probabilities reported in the ATM are arranged in ascending order, then they are picked progressively from lowest to

highest and used as thresholds values for computing the TESs, thresholds progression also determines the direction of the differentiation process.

The function for computing the cell types (i.e., the TESs) is *getTESs()*, which accepts the following argument:

- *ATM* — the Attractor Transition Matrix structure as returned from the *getATM()* function.

From the named list obtained as output, we can extract the computed cell types, which eventually determine the different differentiation stages, and the noise thresholds at which they emerged.

2.3. Cell differentiation process

The software, *diffeRenTES*, provides the ability to produce a visual representation of the main characteristics of the differentiation process that a Boolean network is capable of expressing according to the cell differentiation model mentioned above. It arranges the landscape of cell types (i.e., TESs) that emerge as a result of the application of the noise thresholds to the ATM structure into a tree structure. The levels of the tree structure describe the different degrees of cell differentiation; while their progression, from top to bottom, represents the (pseudo-)temporal evolution of cell types imposed by the progressive decrease in noise.

The resulting differentiation tree (a forest if multiple TESs are present at the root level) can be saved to the file system as an SVG image using the function *saveDifferentiationTreeToFile()*.

saveDifferentiationTreeToFile() accepts the following arguments:

- *TESs* — the object returned by the *getTESs()* function.
- *filename* — the filename used for the image representing the differentiation tree.

Fig. 1 shows an example of a differentiation process computed using the presented package.

3. Illustrative example

In this section, a complete and working usage example is reported. For a complete user's guide, see the library documentation.

In the code snippet shown in listing 1, we can see how first a random Boolean network is created using the *BoolNet* package. The network model, as will be explained later in Section 4, does not have to be random, but can be read from a file using the *loadSBML()* function of the *BoolNet* package. Next, the network attractors are computed. At this point, the proposed software package comes into play, which through the *getATM()* function calculates the Attractor Transition Matrix. Then, passing the latter as the actual argument of the *getTESs()* function, we obtain the abstractions that in the differentiation model represent the various cell types expressed by the gene regulatory network model of interest. Finally, with the function *saveDifferentiationTreeToFile()*, the cell types (i.e., TES abstractions) are visually organized into a typical differentiation tree, respecting their mutual relationships.

Listing 1: Code example

```
library(diffeRenTES)
library(BoolNet)

# Boolean network generation via BoolNet package
net <- BoolNet::generateRandomNKNetwork(10, 2)

# Attractors computation via BoolNet package
synch_attractors <- BoolNet::getAttractors(net)

# Attractors Transition Matrix computation
ATM <- getATM(net, synch_attractors)
```

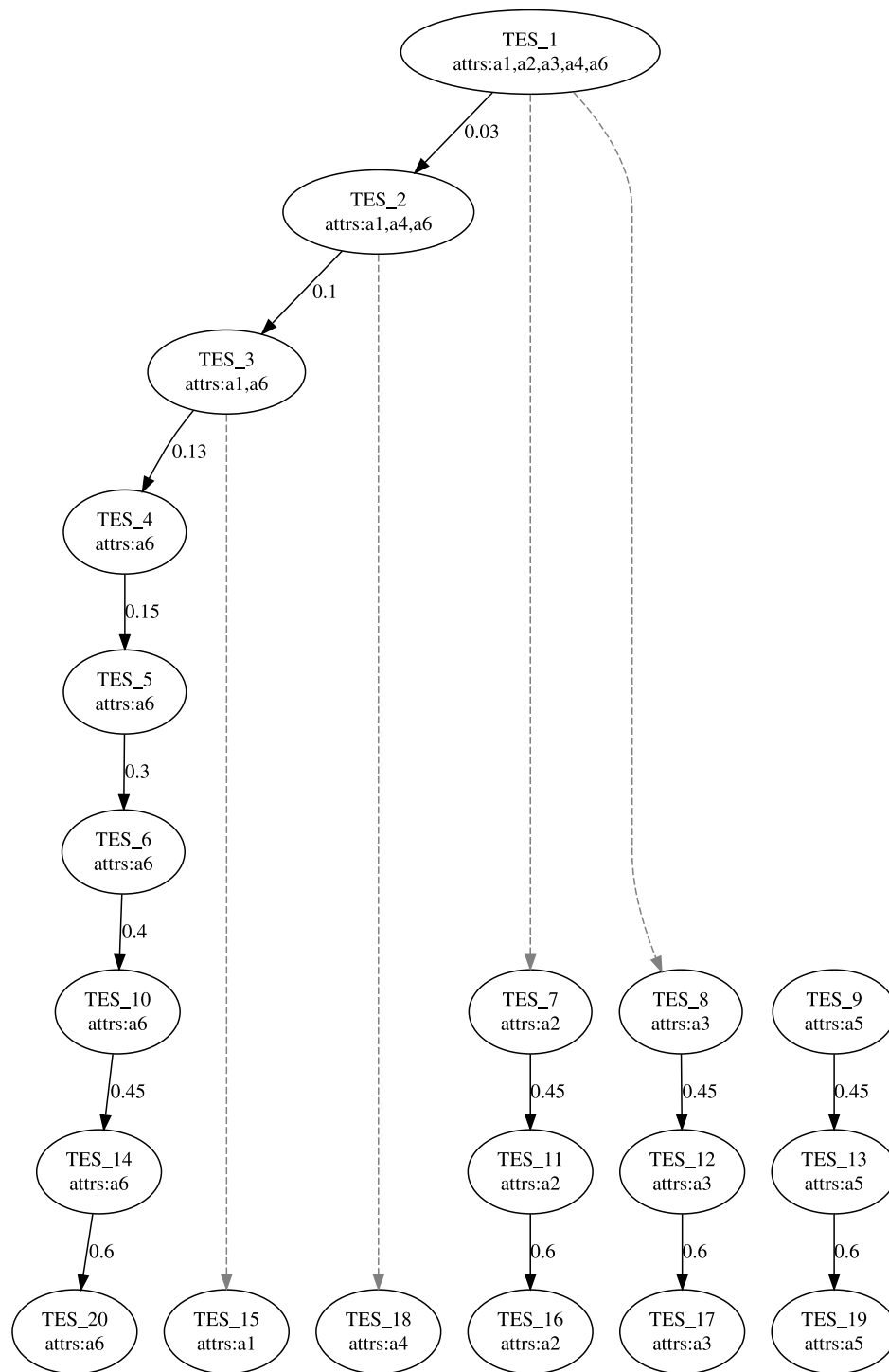


Fig. 1. The image shows an example of a differentiation tree that can be obtained using the *diffeRenTES* package. Each node represents a TES with the additional detail of the attractors that compose it. Instead, labels on the arcs indicate the threshold value at which TESs at the previous level split or reduce to produce the ones in the subsequent level.

```
# Cell types (TESs) computation
TESs <- getTESs(AIM)

# Saving the differentiation tree in "example.
  svg"
saveDifferentiationTreeToFile(TESs, "example.svg
  ")
```

4. Impact

Given the increasing availability of real gene expression data, such as single-cell data, it becomes of paramount importance to formulate models that can accommodate the dynamics these data reveal of the biological phenomenon they represent. Software tools can enable and accelerate the process from model construction to the stage of hypothe-

sis formulation and subsequent testing in wet lab, and finally to model revision, if necessary.

The proposed software package, *diffeRenTES*, addresses the issue of modeling cell differentiation by allowing computational biologists the possibility of obtaining the differentiation dynamics that a Boolean network is capable of expressing. A far from exhaustive list of possible applications and scientific research that this software can enable includes: (i) the simulation of the differentiation scenarios of existing Boolean models of genetic regulatory networks taken from online datasets, such as *Cell Collective* [4]; (ii) the possibility of verifying models predictions against real data, e.g. using it in combination with *Bioconductor* [20,21] and (iii), following the “ensemble approach” proposed by Kauffman [3,22], the identification of the generic properties of gene regulatory networks able to match the statistical features of the specific differentiation process under consideration, similar to what is done in [23–25].

The intermediate results that can be obtained from the functions in this package are also noteworthy. Indeed, as already stated in Section 2.1, the ATM structure computed with the *getATMO* function can be used for performing the robustness analysis of gene regulatory network models, as done in the works [8,26,27].

To date, an initial implementation of the mathematical differentiation model on which *diffeRenTES* is based can be found in the CABERNET software [28], as a Cytoscape application [29]. *diffeRenTES* makes interoperability one of its main strengths. Indeed, since it is developed with R, a programming language widely used in multidisciplinary research fields and particularly in the biology community, it can be easily integrated with existing pipelines developed in both R and python, by taking advantage of one of the many software bridges that allow the combined use of these two programming languages in mixed work environments, e.g., *reticulate* [30] and *rpy2*. This helps to make the world of biological modeling increasingly accessible even to scientists without a background in computer science.

5. Conclusion and future development

Understanding cell differentiation and the implications of its dysregulations on biological organisms are very challenging tasks. Modeling approaches can mitigate these difficulties. The proposed *diffeRenTES* package implements the abstractions of a prominent mathematical model of cell differentiation. In particular, it allows to compute the typical differentiation tree describing the whole set of differentiation scenarios that a cell, modeled as a noisy Boolean network, is able to express. Moreover, it provides the possibility to analyze the robustness of the attractors of a Boolean network.

The capabilities offered by the proposed software, together with their ease of use and widespread adoption of the R language, can help advance the understanding of the processes underpinning biological cell differentiation.

Future development of this package will be devoted primarily to the integration of automatic design techniques for the production of Boolean networks capable of giving rise to the differentiation dynamics of the real biological system under investigation; examples of these techniques can be found in [31]. In addition, another direction of development will involve the implementation of other models of cell differentiation, again based on Boolean models of gene regulatory networks.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

I gratefully acknowledge the useful discussions with Andrea Roli (Università di Bologna).

References

- [1] S. Kauffman, Homeostasis and differentiation in random genetic control networks, *Nature* 224 (5215) (1969) 177–178.
- [2] S. Kauffman, Metabolic stability and epigenesis in randomly constructed genetic nets, *J. Theoret. Biol.* 22 (3) (1969) 437–467.
- [3] S. Bornholdt, S. Kauffman, Ensembles, dynamics, and cell types: Revisiting the statistical mechanics perspective on cellular regulation, *J. Theoret. Biol.* 467 (2019) 15–22.
- [4] T. Helikar, B. Kowal, S. McClenathan, M. Bruckner, T. Rowley, A. Madrahimov, B. Wicks, M. Shrestha, K. Limbu, J.A. Rogers, The cell collective: Toward an open and collaborative approach to systems biology, *BMC Syst. Biol.* 6 (1) (2012) 96.
- [5] S. Huang, G. Eichler, Y. Bar-Yam, D. Ingber, Cell fates as high-dimensional attractor states of a complex gene regulatory network, *Phys. Rev. Lett.* 94 (2005) 128701:1–4.
- [6] S. Huang, I. Ernberg, S. Kauffman, Cancer attractors: A systems view of tumors from a gene network dynamics and developmental perspective, in: *Seminars in Cell & Developmental Biology*, vol. 20, (no. 7) Elsevier, 2009, pp. 869–876.
- [7] S. Huang, D.E. Ingber, Shape-dependent control of cell growth, differentiation, and apoptosis: Switching between attractors in cell regulatory networks, *Exp. Cell Res.* 261 (1) (2000) 91–103.
- [8] S. Montagna, M. Braccini, A. Roli, The impact of self-loops on boolean networks attractor landscape and implications for cell differentiation modelling, *IEEE/ACM Trans. Comput. Biol. Bioinform.* 18 (6) (2021) 2702–2713, <http://dx.doi.org/10.1109/TCBB.2020.2968310>.
- [9] S. Huang, The molecular and mathematical basis of Waddington’s epigenetic landscape: A framework for post-Darwinian biology? *Bioessays* 34 (2) (2012) 149–157.
- [10] A. Roli, M. Braccini, Attractor landscape: A bridge between robotics and synthetic biology, *Complex Syst.* 27 (2018) 229–248, <http://dx.doi.org/10.25088/ComplexSystems.27.3.229>.
- [11] C.H. Waddington, *The Epigenotype*, Endeavour, 1942.
- [12] C.H. Waddington, *The Strategy of the Genes*, Allen and Unwin, London, 1957.
- [13] M.E. Martinez-Sanchez, L. Mendoza, C. Villarreal, E.R. Alvarez-Buylla, A minimal regulatory network of extrinsic and intrinsic factors recovers observed patterns of CD4+ T cell differentiation and plasticity, *PLoS Comput. Biol.* 11 (6) (2015) 1–23, <http://dx.doi.org/10.1371/journal.pcbi.1004324>.
- [14] J.X. Zhou, X. Qiu, A. Fouquier d’Hérouël, S. Huang, Discrete gene network models for understanding multicellularity and cell reprogramming: From network structure to attractor landscape, *Comput. Syst. Biol.* (2014) 241–276.
- [15] R. Serra, M. Villani, A. Barbieri, S. Kauffman, A. Colacci, On the dynamics of random Boolean networks subject to noise: Attractors, ergodic sets and cell types, *J. Theoret. Biol.* 265 (2) (2010) 185–193.
- [16] M. Villani, A. Barbieri, R. Serra, A dynamical model of genetic networks for cell differentiation, *PLoS One* 6 (3) (2011) e17703.
- [17] M. Villani, R. Serra, On the dynamical properties of a model of cell differentiation, *EURASIP J. Bioinform. Syst. Biol.* 2013 (1) (2013) 4, <http://dx.doi.org/10.1186/1687-4153-2013-4>.
- [18] C. Müsself, M. Hopfensitz, H.A. Kestler, BoolNet – an R package for generation, reconstruction and analysis of boolean networks, *Bioinformatics* 26 (10) (2010) 1378–1380.
- [19] R. Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2022.
- [20] W. Huber, V.J. Carey, R. Gentleman, S. Anders, M. Carlson, B.S. Carvalho, H.C. Bravo, S. Davis, L. Gatto, T. Girke, R. Gottardo, F. Hahne, K.D. Hansen, R.A. Irizarry, M. Lawrence, M.I. Love, J. MacDonald, V. Obenchain, A.K. Oleś, H. Pagès, A. Reyes, P. Shannon, G.K. Smyth, D. Tenenbaum, L. Waldron, M. Morgan, Orchestrating high-throughput genomic analysis with bioconductor, *Nature Methods* 12 (2) (2015) 115–121, <http://dx.doi.org/10.1038/nmeth.3252>.
- [21] R.C. Gentleman, V.J. Carey, D.M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A.J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J.Y. Yang, J. Zhang, Bioconductor: Open software development for computational biology and bioinformatics, *Genome Biol.* 5 (10) (2004) R80, <http://dx.doi.org/10.1186/gb-2004-5-10-r80>.
- [22] S. Kauffman, A proposal for using the ensemble approach to understand genetic regulatory networks, *J. Theoret. Biol.* 230 (2004) 581–590.
- [23] I. Shmulevich, S. Kauffman, M. Aldana, Eukaryotic cells are dynamically ordered or critical but not chaotic, *Proc. Natl. Acad. Sci.* 102 (38) (2005) 13439–13444.
- [24] M. Braccini, A. Roli, M. Villani, S. Montagna, R. Serra, A simplified model of chromatin dynamics drives differentiation process in boolean models of GRN, in: *ALIFE 2022: The 2022 Conference on Artificial Life*, 2019, pp. 211–217, http://dx.doi.org/10.1162/isal_a_00163.
- [25] M. Aldana, E. Balleza, S. Kauffman, O. Resendiz, Robustness and evolvability in genetic regulatory networks, *J. Theoret. Biol.* 245 (3) (2007) 433–448, <http://dx.doi.org/10.1016/j.jtbi.2006.10.027>.

- [26] S. Montagna, M. Braccini, A. Roli, The impact of self-loops in random boolean network dynamics: A simulation analysis, in: M. Pelillo, I. Poli, A. Roli, R. Serra, D. Slanzi, M. Villani (Eds.), *Artificial Life and Evolutionary Computation – 12th Italian Workshop, WIVACE 2017, Revised Selected Papers*, in: CCIS, vol. 830, Springer, 2018, pp. 104–115.
- [27] M. Braccini, S. Montagna, A. Roli, Self-loops favour diversification and asymmetric transitions between attractors in Boolean network models, in: S. Cagnoni, M. Mordonini, R. Pecori, A. Roli, M. Villani (Eds.), *Artificial Life and Evolutionary Computation*, Springer International Publishing, Cham, 2019, pp. 30–41.
- [28] A. Paroni, A. Graudenzi, G. Caravagna, C. Damiani, G. Mauri, M. Antoniotti, CABERNET: A cytoscape app for augmented Boolean models of gene regulatory networks, *BMC Bioinformatics* 17 (1) (2016) 64.
- [29] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: A software environment for integrated models of biomolecular interaction networks, *Genome Res.* 13 (11) (2003) 2498–2504, <http://dx.doi.org/10.1101/gr.1239303>.
- [30] K. Ushey, J. Allaire, Y. Tang, Reticulate: Interface to 'Python', 2022, <https://rstudio.github.io/reticulate/>, <https://github.com/rstudio/reticulate>.
- [31] M. Braccini, A. Roli, M. Villani, R. Serra, Automatic design of Boolean networks for cell differentiation, in: *Advances in Artificial Life, Evolutionary Computation, and Systems Chemistry: 11th Italian Workshop, WIVACE 2016, Fisciano, Italy, October 4-6, 2016, Revised Selected Papers*, Springer International Publishing, Cham, 2017, pp. 91–102, http://dx.doi.org/10.1007/978-3-319-57711-1_8.