2015 11th International Conference on Signal-Image Technology & Internet-Based Systems

Topological Evaluation of Methods for Reconstruction of Genetic Regulatory Networks

Jakub Olczak¶, Narsis A. Kiani¶, Hector Zenil∓ and Jesper Tegnér∓

***Abstract*—Network inference is advancing rapidly, and new methods are proposed on a regular basis. Understanding the advantages and limitations of different network inference methods is key to their effective application in different circumstances. The common structural properties shared by diverse networks naturally pose a challenge when it comes to devising accurate inference methods, but surprisingly, there is a paucity of comparison and evaluation methods. Historically, every new methodology has only been tested against *gold standard* purpose-designed synthetic and real-world (validated) biological networks. In this paper we aim to assess the impact of taking into consideration topological aspects on the evaluation of the final accuracy of an inference procedure. Specifically, we will compare the best inference methods, in statistical terms, for preserving the topological properties of synthetic and biological networks. A new method for performance comparison is introduced by borrowing ideas from gene set enrichment analysis. Experimental results show that no individual algorithm stands out among the three inference tasks assessed, and the challenging nature of network inference is evident in the struggle of some of the algorithms to turn in a performance that's better than random guesswork. Therefore care should be taken to suit the method used to the specific purpose.**

***Keywords*—Network reconstruction; topological properties; inference methods evaluation; evaluating measure**

Many real-world networks, such as complex technological and social networks, belong in the category of so called 'complex networks', and have a number of the properties that govern the formation and evolution of complex networks [20], [1], [2]. The accurate inference of networks from biological data is an open challenge. The modelling and inference of genetic regulatory networks has developed into a broad field of study in the past few years, with the application of ever more sophisticated techniques. The recent Dialogue for Reverse Engineering Assessment and Methods (DREAM) challenge [29], [30] has resulted in significant progress. The DREAM challenge aims to fairly compare the strengths and weaknesses of inference methods. Network inference methods have complementary pros and cons under different conditions. Ideally, the validation and interpretation of GRN models must keep pace with new knowledge and experimental data available for modelling, and thus it is important to illustrate all aspects and capacities of a network inference method. For two major reasons, the assessment of inferred networks is not trivial. First, our understanding of gene regulatory networks is still only partial. Second, networks are structured objects and we cannot simply evaluate them on a local scale, but also on intermediate levels, on the level of the whole network, and any combination thereof [4], [19]. Generally, researchers produce artificial networks and simulated data for method assessment. Synthetic data do not usually reflect the complexity of a real biological system if no biological prior information is introduced. Although the exact details may differ, most methods of evaluation of network reconstruction consider the sensitivity, specificity, precision, and in some cases, a receiver operating characteristic (ROC) curve to illustrate the performance of an approach. The analysis of biological networks has led to the realisation that the architecture of these networks shares many features with other complex networks. They show non-trivial topological properties such as modular structure and long-tail degree distribution [26]. The common structural properties shared by diverse networks naturally pose a challenge when it comes to devising more accurate inference methods capable of preserving them. Surprisingly there has been no evaluation or comparison of different models from this point of view. Understanding the advantages and limitations of different network inference methods is necessary for their effective application in specific circumstances. In this paper we address this question, evaluating the similarity between the structural features of a true network and an inferred network. We have chosen 6 different inference algorithms from among the best-performing algorithms in past DREAM challenges. These methods have been studied using statistical performance measures such as the F-score [19] or area under the receiver operator curve (AUROC) [17], [14]. Some other attempts have been made to consider aspects related to the overall properties of the inferred network rather than the specific number of false and true positive/negative edge inference cases [28]. In this article, we analyse network inference methods employing topological measures and indices, in combination with ensemble data, in order to assess their performance. An effective similarity metric is needed for scoring network inference methods, one which, given two complex networks, evaluates the degree of similarity between their structural features, beyond just looking at individual numbers. We borrow ideas from gene set enrichment analysis (GSEA) [31], [18], [5] to formulate an intelligent method which we offer as a new way to measure the topological similarity of two complex networks. We benchmark them using synthetic transcriptional networks proposed by Mendes et al. [25]. These networks consist of 100 genes and are organised either in an Erdös-Rényi (random network), a small world or a scale-free topology [32], [7], [26]. Mendes et al. have used these networks with well-defined topologies to run *in-silico* experiments simulating real laboratory micro-

Unit of Computational Medicine, Science For Life Laboratory (SciLifeLab), Centre for Molecular Medicine, Department of Medicine Solna, Karolinska Institute, Stockholm, Sweden.

Corresponding authors: {Narsis.Kiani, Hector.Zenil}@ki.se

¶, ∓: These authors contributed equally to this paper.

978-1-4673-9721-6/15 $31.00 © 2015 IEEE
DOI 10.1109/SITIS.2015.107

array experiments. We have compared ARACNe [22], CLR (Context Likelihood of Relatedness [8], GENIE3 [15], INFERELATOR [10], TIGRESS [11] and Correlation on the basis of diameter, average shortest path length, clustering and centrality scores.

I. METHODS

Six different network inference algorithms are considered in this study and will be discussed in the following section. Table I-A summarises the differences between the models used.

A. Network inference methods

Several methods have been proposed for inferring gene regulatory interactions from measured gene expression levels. Approaches employed include Bayesian networks, Boolean models, auto-regressive models, correlation-based models, clustering techniques and differential equation models, among others [9], [23], [6], [17], [21]. Some of them are static, while others take into account the dynamic aspects of the dependencies. Mutual information network inference methods are a class of network inference methods which infer regulatory interactions between genes based on pairwise mutual information. The low computational complexity and the low number of required samples are the main advantages of mutual information based inference methods. We have examined two commonly used state-of-the-art network inference methods based on pairwise mutual information: Algorithm for the Reconstruction of Accurate Cellular Networks, ARACNe [22] and Context Likelihood of Relatedness, CLR [8]. ARACNe is based on an information theoretical approach that uses the concept of Mutual Information, MI, a measure of entropy, to determine the pairwise interaction between nodes by assessing the MI between them. It then applies a data processing inequality (DPI), to eliminate indirect interactions. The CLR algorithm is an extension of the network relevance approach. It is another information theoretic approach and computes the MI between two nodes, comparing it to the empirical background distributions of MI. Regression based network inference methods comprise one of the largest network inference sub-categories, and we have studied 3 of the best regression based methods: GENIE3, TIGRESS and INFERELATOR. GENIE [15] decomposes the prediction of a regulatory network between p genes into p different regression problems such that in each, the expression pattern of one of the genes may be predicted from the expression patterns of all the other genes, using tree-based ensemble methods– Random Forests or Extra-Trees. TIGRESS [11] formulates the inference problem as a sparse linear regression problem. It uses least angle regression (LARS) and adds an additional *stability selection* criterion to assess the significance of nodes in the regression. INFERELATOR uses regression and variable selection to identify transcriptional influences on genes based on the integration of genome annotation and expression data. In addition to these methods, we have used Correlation to reconstruct networks. For Correlation, CLR, GENIE, INFERELATOR and TIGRESS, we used the implementations at the Michigan Institute of Technology's Broad Institute [27], [12].

TABLE I. SUMMARY OF ASSESSED NETWORK INFERENCE MODELS.

Method	Category	Features
ARACNe	Information-theoretic	MI estimated using a copula based approach Use of the DPI to break up fully connected triplets
CLR	Information-theoretic	MI dependencies(Gaussian assumption) Normalisation of MI
GENIE3	Tree-Based Methods	Decomposes into p different regression problems Prediction using tree-based ensemble methods
INFERELATOR	ordinary differential Equations	Hybrid method involving differential equations and Regression
TIGRESS	Regression Methods	least angle regression (LARS) combined with stability selection
Basic Correlation	Statistic	-

We used an ARACNe implementation in GE Workbench 2.5. 1 [13]. Parameters are always default, set by GenePattern 2.0 or GE Workbench, unless otherwise stated.

B. The datasets

Benchmarking is important in order to be able to understand the reliability of the reconstructed network. Traditionally, the assessment proceeds by collecting all curated interactions and considering them as true positives, while treating as false positives all predicted interactions between two genes that are not documented in the curated database. Such a method tends to overestimate the false-positive prediction rate while ignoring all new interactions. As a result, methods that merely reproduce current knowledge outperform those that do well at finding new results. To compensate for this, the gold standard networks were selected from among synthetic transcriptional networks proposed by Mendes et al. [24]. These networks with well-defined topologies have been used by them to run *in-silico* experiments simulating real laboratory micro-array experiments. They consist of 100 genes and are organised either in an Erdös-Rényi (ER), Smallworld (SW) or a scale-free (SF) topology. We have chosen 10 networks from each topology, RND001 to 010, SW001 to 010 and SF001 to 010. The simulated data from these networks have been used as input for network inference methods.

C. Assessment method

The performance of network inference methods has traditionally been evaluated using a confusion matrix with respect to the gold standard network, GSN, providing the number of true positives TP, true negatives TN, false positives FP and false negatives FN. The measures in this confusion matrix have the following meaning in the context of this paper: TN refers to an edge that belongs neither to the predicted network nor to the gold standard network; FP is the number of predicted edges that do not belong to the gold standard network; FN is the number of edges in the gold standard that are missing from the predicted network; and TP is the number of correct predictions of an edge in the gold standard network. To quantify network reconstruction performance, we first used receiver operator characteristic (ROC) and precision recall (PR) analysis. We have used a threshold δ for discretisation of edge values, where the weight $W_{i,j}$ for a particular edge is compared with δ. If $|W_{i,j}| \leq \delta$, the edge $e_{i,j}$ is assumed to be present in the network $e_{i,j} \in E$, and absent otherwise. The resulting network with edge set E is then compared against the gold standard network, and sensitivity, specificity and precision

are computed for given δ. This is then repeated by varying δ, and sensitivity is plotted over specificity for different δ in a ROC plot. Finally, the ROC curve can be summarised by computing the area under the curve [17]. As our second approach, we have used the Jaccard coefficient [16]. This commonly used similarity metric measures the probability that the two networks, the gold standard and the inferred network, have common edges, focusing on randomly selected edges in either of the networks.

$$\text{JaccardCoefficient(GSN,IN)}:= |E(GSN) \bigcap E(IN)|/|E(GSN) \bigcup E(IN)|.$$

1) New assessment procedure:Topological Indices Enrichment Analysis : We have compared the topological indexes by borrowing ideas from gene set enrichment analysis (GSEA). We call our procedure Topological Indices Enrichment Analysis, TIEA. GSEA is one of the most widely used methods for detecting differentially expressed gene sets. GSEA [18] is a discreet version of the weighted Kolmogorov-Smirnov test, which is applied to a running sum statistic over ranked lists, counting how often elements are or are not in the list of interest. Unlike in the analysis of gene expression data, the sets here were defined not by genes but by nodes from networks, and ranked not based on expression but on the topological index of interest. For TIEA, nodes are first ranked by topological score. Then a "running sum" statistic is calculated for each network, based on the ranks of subsets of nodes in the network, relative to those of non-members. An enrichment score (ES) is defined to be the maximum of the running sum across all nodes, which corresponds to a weighted Kolomogorov-Smirnov statistic. The equation for the calculation of ES for the sorted list was processed from top to bottom, and two running sums, RS_{N_k} and $\overline{RS}_{N_k}$, were computed. RS_{N_k} was increased by one each time a node belonged to N_k, and $\overline{RS}_{N_k}$ each time a node belonged to the complementary set N_k:

$$RS_{N_k}(i) = \Sigma_{\substack{n_j \in N_k \\ j \leq i}} \frac{1}{\|N_k\|}$$

$$\overline{RS}_{N_k}(i) = \Sigma_{\substack{n_j \notin N_k \\ j \leq i}} \frac{1}{\|N - N_k\|}$$

$$TIES_{N_k} = max(|RS_{N_k} - \overline{RS}_{N_k}|)$$

2) Topological indices: A network, graph G, consists of a set of nodes representing biological entities $V(G)$, while the edges $E(G)$ denote relationships between node pairs. Its topological structure is the most basic and direct information available about a network. The architectural features of biological networks can be roughly categorised into three classes: individual, local and global features. Individual features are topological properties associated with only one node, including degree and centrality measures; global features involve all the vertices in networks, while local features are those behaviours that involve part of the network rather than the whole network containing motifs [3] and communities. This paper confines itself to individual features. We focus on the preservation of diameter, average path length, clustering, centrality and degree distributions.

Definition I.1. The diameter of a network is the largest distance between any two nodes in the network

Definition I.2. The average path length is the average distance between any two nodes in the network.

Average path length is bounded from above by the diameter; in some cases it can be much shorter than the diameter. If the network is not connected, one often checks the diameter and the average path length in the largest component.

Definition I.3. The overall clustering coefficient Cl(G) is given by

$$Cl(g) = \frac{3 * number\,of\,triangles\,in\,the\,network}{number\,of\,connected\,triples\,of\,nodes}$$

where a "connected triple" refers to a node with edges linked to an unordered pair of nodes.

Definition I.4. The individual clustering for a node i is

$$Cl_i(g) = \frac{number\,of\,triangles\,connected\,to\,vertex\,i}{number\,of\,triples\,centered\,at\,i}$$

II. Results

ARACNe, Basic Correlation and GENIE3 successfully inferred networks from inputs. CRL returned empty networks for all inputs; INFERELATOR broke down due to "zero variance" for the subset of SF networks, but worked for the other sets; TIGRESS returned results only for seven networks in the ER subset. The AUCROC values for each algorithm and dataset can be seen in 1A. All models turn in performances significantly better than random guesswork, except ARACNe. Despite the significant difference, the magnitude of difference compared to random guessing is not, in the best case, more than 10%. Relative to the datasets ER and SF, the best performer is GENIE, while for SW the best performer is Correlation, but there is no significant difference between GENIE and Correlation for SW networks. When we examine the area under the precision recall curve (AUCPR) of the models for the same set of networks, we find that GENIE significantly outperforms Correlation (see 1B). This shows the higher number of false positives in the network predicted by Correlation. It has been shown that most biological networks are scale free networks, yet all methods perform significantly better for SW topology. All in all, ARACNe turns in the worst performance on all network categories. This may be due to the number of arbitrary parameters and the effect of the cut of parameters on network architecture. Surprisingly, Basic Correlation turns in a performance comparable to other methods in all categories when statistical measures are the parameters being compared.

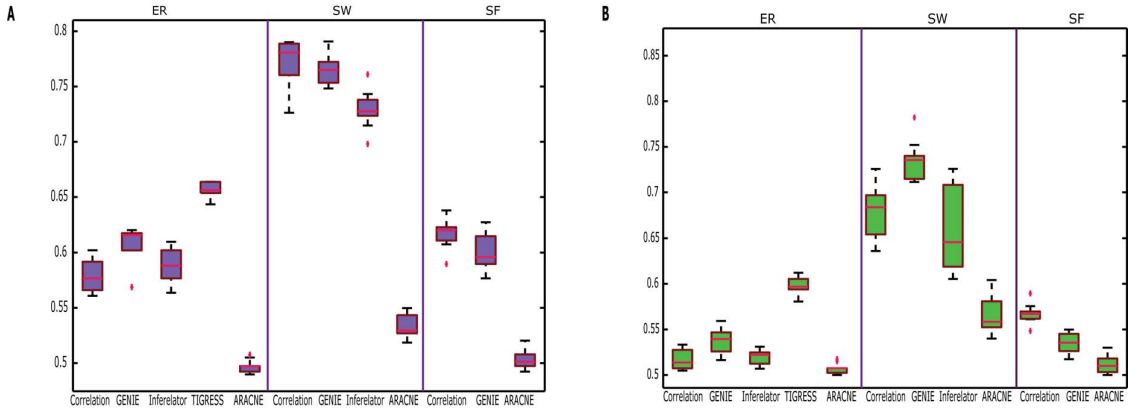

Fig. 1. Evaluation of simulated data. Network reconstruction was performed for all networks in all 3 categories. Shown are **A:** the distribution of the area under the ROC curve (AUCROC), and **B:** the area under the precision recall curve (AUCPR) of 30 sets of simulated data, over the different topologies. From left to right: Correlation, GENIE, Inferelator, TIGRESS and ARACNe. AUC $= 0.5$ is the expected result for random guessing.

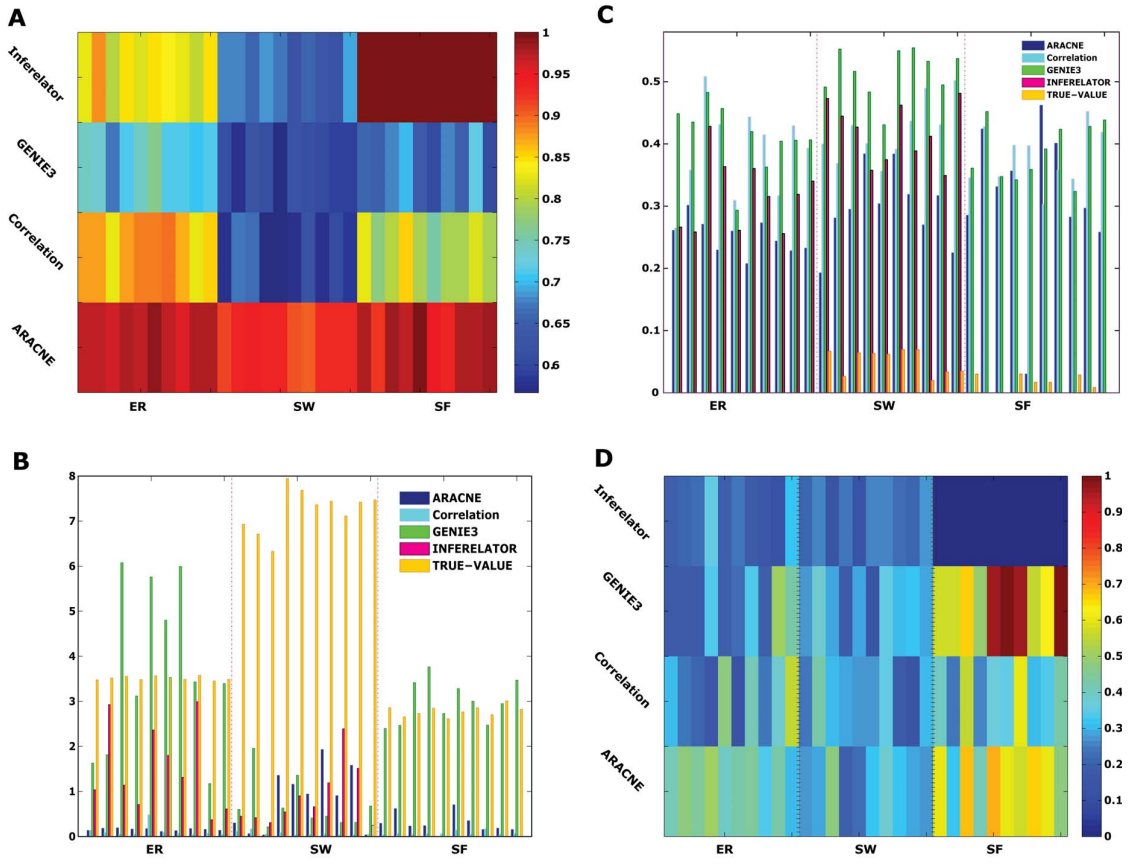

Fig. 2. The figure shows the performance of all models based on different measures and assessment approaches. Assessment of models' performance using Jaccard distance (panel A), Average shortest path (panel B) Clustering coefficient (panelC). The hub enrichment score is shown in panel D.

	Jacc. dist	ΔCC	ΔSP	$\Delta Diam$
ARACNe[a]	0.9550	-0.1547	4.088	3.50
TIGRESS[c]	0.6225	-0.1562	-0.844	-6.14
Basic Correlation[a]	0.7567	-0.2634	4.428	4.37
GENIE3[a]	0. 6607	-0.3056	2.045	-3.93
INFERELATOR[b]	0.7477	-0.3350	4.183	0.95

TABLE II. THE JACCARD DISTANCE, CLUSTERING COEFFICIENT DIFFERENCE(ΔCC), SHORTEST PATH DIFFERENCE (ΔSP) AND DIAMETER DIFFERENCE ($\Delta Diam$) BETWEEN ALL GSN AND INFERRED NETWORKS HAVE BEEN CALCULATED FOR ALL MODELS USING EUCLIDEAN DISTANCE. AVERAGES OVER ALL 30 NETWORKS ARE SHOWN. ALL NETWORKS, [a]: BASED ON 30 NETWORKS, [b]: BASED ON 20 NETWORKS, [c]: BASED ON 7 RND NETWORKS.

	HubES	CCES	SPES	DiamES
ARACNe[a]	0.4336	0.4847	0.3429	0.4873
GENIE3[a]	0.4277	0.4305	0.3587	0.4902
INFERELATOR[b]	0.2187	0.4132	0.5362	0.1393
Basic Correlation[a]	0.3252	0.1342	0.2248	0.2042
TIGRESS[c]	0.1623	0.1230	0.1181	0.2362

TABLE III. THE HUB ENRICHMENT SCORE (HUBES), CLUSTERING COEFFICIENT ENRICHMENT SCORE (CCES), SHORTEST PATH ENRICHMENT SCORE (SPES) AND DIAMETER ENRICHMENT SCORE (DIAMES) ARE SHOWN. CALCULATION WAS DONE BY COMPARING THE TOP 10% OF GSN WITH THE PREDICTED NETWORKS, WITH DATA WITH DIFFERENT LEVELS OF NOISE AND MISSING VALUES, WITH THE SN2 DATA, INCLUDING A NEGATIVE FEEDBACK LOOP, AND THE SN3 (KEGG) DATA SET. VALUES SHOWN ARE AVERAGE VALUES OVER 30 NETWORKS. ALL NETWORKS, [a]: BASED ON 30 NETWORKS, [b]: BASED ON 20 NETWORKS, [c]: BASED ON 7 RND NETWORKS.

Then we assessed the Jaccard distance between each GSN and its corresponding inferred network. Figure 2A shows the result. As can be seen, the overall picture remains the same as AUCROC, the topology of the underlying network significantly affects the performance of all methods, and GENIE3 outperforms other methods. Then we have used topological indices to assess all inference methods. We have done this using both enrichment scores and Euclidean distance. An overview of the results has been shown in II and in Figures 2B and C. We found discrepancies in the rankings obtained using the 3 approaches. As has been mentioned, ARACNe was the worst performer, but when assessed using a topological index, it was one of the two most effective methods. As another example, prediction by GENIE3 outperforms ARACNe, and it is closest to the GSN if we focus on the shortest path and measure difference using euclidean distance, but ARACNe is better when the clustering coefficient is the parameter being compared.

We then compared all models using our new approach– TIEA. First, we considered hub enrichment scores for all models. For this purpose we selected 10% of the top hub nodes in the GSN and calculated the enrichment score of this set in all prediction models. Of all the methods, we found ARACNe to have the highest median hub enrichment score. GENE3 got the highest score for SF networks; see 2D. We used the same procedure for the clustering coefficient and diameter of a network. ARACNe's ranking improves dramatically when we use TIEA for evaluation. This shows the power of ARACNe when it comes to capturing the central section of a network. On the other hand, GENEI fares better at predicting the totality of a network. TIEA measures the ability of a model to predict the most important features of a network, such as euclidean distance, by giving equal weight to all nodes. The result has been reported in tables II and II. A comparison of Tables II and II brings home the importance of picking a model suited to the specific task at hand.

III. CONCLUSIONS

Current efforts aim to understand the individual strengths and weaknesses of various network reconstruction methods by applying them to equal and different data sets. Generally, sensitivity, specificity, precision and the (ROC) curve are calculated to illustrate the performance of a particular approach. However, there are more aspects of network construction that should be taken into consideration. Here we suggest using topological indices to evaluate network inferences, and furthermore we introduce a concept similar to gene set enrichment for network inference evolution. Depending on the ultimate goal, one can use Euclidean distances or TIEA to compare models. This introduces a graph-theoretical perspective on the problem that allows one to study particular substructures of networks and not be limited to studying networks in their entirety. In this paper, several commonly used computational approaches for constructing gene regulatory networks are compared, using both topological and statistical indices. We have used data produced by synthetic networks to address questions such as the following: Which topologies are best suited for a specific network inference method. Which models work best for predicting specific properties of a network? Obviously, synthetic data cannot reflect the complexity of a real biological system. However, standards are still unavailable for evaluating different inference methods using real biological data. Results obtained using three different datasets show that the overall performance of the models assessed is poor. There are significant differences in the results obtained with the datasets (ER, SW and SF). In general, we observe that when based on statistical measures, network inference methods could be said to perform better with small world networks, while when based on topological measures, they perform better with SF networks. Our new assessment process revealed a new feature of the models: we observed that the performance of a model depended on the topological index. We have measured the performance of various gene regulatory network construction methodologies against various sizes of simulated data with different numbers of samples. A few conclusions can be drawn from this exercise. First, GENIE and INFERELATOR performed well in constructing the global network, while ARACNe did well in identifying a few connections with high specificity. Surprisingly, Correlation performed well in constructing the global network, doing better than ARACNe and approaching GENIE. GENIE performed well in both respects, but it is not suitable for identifying the hub nodes which can often be of biological interest. ARACNe performed well in identifying the hub genes. Since there is no single method that outperforms other methods in all respects, care should be taken to choose an appropriate method based on the purpose of the study.

ACKNOWLEDGMENTS

This work was supported in part by the Foundational Questions Institute (HZ), the VINNOVA (VINNMER) Marie-Curie fellowship, (NK), StratNeuro (JT,NK), AFA Insurance(JT), Torsten Söderberg Foundation (JT), STATegra (JT, HZ), the Stockholm County Council and the Swedish Research Council. The funders played no role in the design of the study, in data collection and analysis, in the decision to publish, or in the preparation of the manuscript.

REFERENCES

[1] Albert, R. & Barabási, A.-L. (2002) Statistical mechanics of complex networks. *Rev. Mod. Phys.*, **74**, 47.

[2] Albert, R., Jeong, H. & Barabasi, A.-L. (2000) Error and attack tolerance of complex networks. *Nature*, **406**(6794), 378–382.

[3] Alon, U. (2006) *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall/CRC, Boca Raton, FL, USA.

[4] Altay, G. & Emmert-Streib, F. (2010) Revealing differences in gene network inference algorithms on the network level by ensemble methods. *Bioinformatics*, **26**(14), 1738–1744.

[5] Amberkar, S., Kiani, N. A., Bartenschlager, R., Alvisi, G. & Kaderali, L. (2013) High-throughput RNA interference screens integrative analysis: Towards a comprehensive understanding of the virus-host interplay. *World J Virol*, **2**(2), 18–31. 24175227[pmid].

[6] Bansal, M., Belcastro, V., Ambesi-Impiombato, A. & di Bernardo, D. (2007) How to infer gene networks from expression profiles. *Mol. Syst. Biol.*, **3**, 78.

[7] Erdős, P. & Rényi, A. (1959) On Random Graphs I. *Publicationes Mathematicae (Debrecen)*, **6**, 290–297.

[8] Faith, J. J., Hayete, B., Thaden, J. T., Mogno, I., Wierzbowski, J., Cottarel, G., Kasif, S., Collins, J. J. & Gardner, T. S. (2007) Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles.. *PLoS BIOLOGY*, **5**(1), e8.journal.pbio.0050008.pdf.

[9] Gardner, T. S. & Faith, J. J. (2005) Reverse-engineering transcription control networks. *Phys Life Rev*, **2**(1), 65–88.

[10] Greenfield, A., Madar, A., Ostrer, H. & Bonneau, R. (2010) DREAM4: Combining Genetic and Dynamic Information to Identify Biological Networks and Dynamical Models. *PLoS ONE*, **5**(10), e13397.journal.pone.0013397.pdf.

[11] Haury, A.-C., Mordelet, F., Vera-Licona, P. & Vert, J.-P. TIGRESS: Trustful Inference of Gene REgulation using Stability Selection. *arXiv:1205.1181 [stat.ML]*.1205.1181v1.pdf.

[12] http://dream.broadinstitute.org/ (2014) .

[13] http://wiki.c2b2.columbia.edu/workbench/index.php/ARACNe (2014) .

[14] Husmeier, D. (2003) Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks. *Bioinformatics*, **19**(17), 2271–2282.

[15] Huynh-Thu, V. A., Irrthum, A., Wehenkel, L. & Geurts, P. (2010) Inferring Regulatory Networks from Expression Data Using Tree-Based Methods. *PLoS ONE*, **5**(9), e12776 doi:10.1371/journal.pone.0012776.journal.pone.0012776.pdf.

[16] Jaccard, P. (1912) The distribution of the flora in the alpine zone. *New Phytologist*, **11**, 37–50.j.1469-8137.1912.tb05611.x.pdf.

[17] Kiani, N. & Kaderali, L. (2014) Dynamic probabilistic threshold networks to infer signaling pathways from time-course perturbation data. *BMC Bioinformatics*, **15**(1), 250.

[18] Knapp, B., Rebhan, I., Kumar, A., Matula, P., Kiani, N., Binder, M., Erfle, H., Rohr, K., Eils, R., Bartenschlager, R. & Kaderali, L. (2011) Normalizing for individual cell population context in the analysis of high-content cellular screens. *BMC Bioinformatics*, **12**(1), 485.

[19] Lopes, M. & Bontempi, G. (2013) Experimental assessment of static and dynamic algorithms for gene regulation inference from time series expression data. *Frontiers in Genetics*, **4**(303).

[20] Madar, D., Dekel, E., Bren, A. & Alon, U. (2011) Negative auto-regulation increases the input dynamic-range of the arabinose system of *Escherichia coli*. *BMC Systems Biology*, **5**, 111.

[21] Marbach, D., Prill, R. J., Schaffter, T., Mattiussi, C., Floreano, D. & Stolovitzky, G. (2010) Revealing strengths and weaknesses of methods for gene network inference. *Proc Natl Acad Sci U S A*, **107**(14), 6286–91.

[22] Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzkya, G., Favera, R. D. & Califano, A. (2004) ARACNe: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. *BMC Bioinformatics*, **7**(Suppl I), S7.1471-2105-7-S1-S7.pdf.

[23] Markowetz, F. & Spang, R. (2007) Inferring cellular networks – a review. *BMC Bioinformatics*, **8**, S5.

[24] Mendes, P., Sha, W. & Ye, K. (2003a) Artificial gene netwokrs for objective comparison of analysis algorithms. *Bioinformatics*, **19**, ii122–ii129.

[25] Mendes, P., Sha, W. & Ye, K. (2003b) Artificial gene networks for objective comparison of analysis algorithms. *Bioinformatics*, **19**(Suppl II), ii 122–ii 129. Bioinformatics-2003-Mendes-ii122-9.pdf.

[26] Newman, M. E. J. (2003) The Structure and Function of Complex Networks. *SIAM Review*, **45**(2), 167–256.

[27] Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P. & Mesirov, J. P. (2006) GenePattern 2.0. *Nature Genetics*, **38**, 500 – 501.

[28] Siegenthaler C, Gunawan R (2014) Assessment of Network Inference Methods: How to Cope with an Underdetermined Problem. PLoS ONE 9(3): e90481.

[29] Stolovitzky, G., Monroe, D. & Califano, A. (2007) Dialogue on Reverse Engineering Assessment and Methods:. *Ann. N.Y. Acad. Sci.*, **11158**, 1.

[30] Stolovitzky, G., Prill, R. J. & Califano, A. (2009) Lessons from the DREAM2 Challenge. *Ann. N.Y. Acad. Sci.*, **1158**, 159–195.

[31] Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S. & Mesirov, J. P. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, **102**(43), 15545–50.pnas-0506580102.pdf.

[32] Watts, D. J. & Strogatz, S. H. (1998) Collective dynamics of 'small-world' networks. *Nature*, **393**, 440–442.393440a0.pdf.