Biological Network Visualization for Targeted Proteomics based on Mean First-Passage Time in Semi-Lazy Random Walks

Tomasz Arodz

Department of Computer Science, Virginia Commonwealth University,
Richmond, VA 23284, USA
tarodz@vcu.edu

Abstract. Experimental data from protein microarrays or other targeted assays are often analyzed using network-based visualization and modeling approaches. Reference networks, such as a graph of known protein-protein interactions, can be used to place experimental data in the context of biological pathways, making the results more interpretable. The first step in network-based visualization and modeling involves mapping the measured experimental endpoints to network nodes, but in targeted assays many network nodes have no corresponding measured endpoints. This leads to a novel problem – given full network structure and a subset of vertices that correspond to measured protein endpoints, infer connectivity between those vertices. We solve the problem by defining a semi-lazy random walk in directed graphs, and quantifying the mean first-passage time for graph nodes. Using simulated and real networks and data, we show that the graph connectivity structure inferred by the proposed method has higher agreement with underlying biology than two alternative strategies.

Keywords: Biological networks · Random walks · Node influence.

1 Introduction

Profiling experiments involving gene or protein microarrays or assays based on next-generation sequencing have become a standard approach for gaining new knowledge about biological processes and pathologies. Mining the resulting data for patterns of interest, for example differences between phenotypes, can be done with purely data-driven statistical and machine learning methods [11] that perform the discovery *de novo*. But approaches that make use of existing knowledge about biological networks in analyzing profiling data are, in principle, better suited to deal with the complexity of biological systems.

Extensive knowledge has been gathered about physical or functional interactions between biological entities of many types. For example, it may be known that a certain kinase phosphorylates a specific protein, or that a particular microRNA interacts physically with mRNA transcript of a gene, and in effect the protein encoded by the gene is not expressed. All known interactions of a specific

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50420-5_40

type, taken together, form reference networks, such as a protein-protein interaction network or a gene regulatory network. A reference network describes the interaction potential of a given species. In a specific phenotype, that is, a specific tissue in a specific condition, only some of the interactions actually take place.

Based on the network topology and measurements from profiling experiments, dynamic behavior of the system can be modeled using stochastic Petri nets [12], Boolean networks [14], Bayesian networks [31], or systems of differential equations [5]. Pathways that can discriminating between phenotypes can be discovered by mapping expression data on reference networks and performing a bottom-up [13, 7] or top-down [6, 2] search. Depicting experimental results visually by mapping the up-regulated or down-regulated genes or proteins can also make the data more interpretable to biologists.

2 Motivation and Problem Statement

Network-based analysis involves mapping the measured experimental endpoints to nodes in the reference network. Often, many nodes will have no corresponding endpoints. This is particularly true in studies that involve targeted assays. For example, only a limited number of antibodies are available and validated for use with the reverse phase protein array (RPPA) immunoassay [22]. Thus, for many nodes in a reference protein-protein interaction network, experimental protein levels will not be measured.

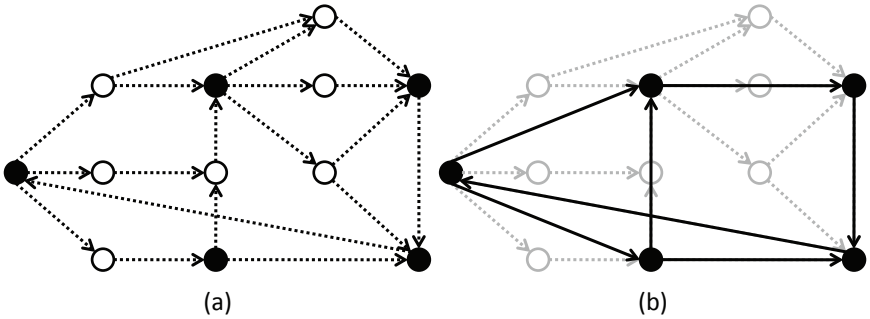

Fig. 1. Illustration of the problem. (*a*) Input: a reference biological network and a subset of nodes (filled black nodes) for which measurements from a targeted assay are available; measurements for the other nodes are not available. Typically the measured nodes are a small fraction of all nodes in the network. (*b*) Output: an informative and interpretable network (black edges) connecting the measured nodes.

Some network analysis methods can deal with lack of measurements at a subset of nodes. Prize-collecting Steiner tree and related approaches [2] can involve unmeasured nodes in finding discriminative pathways, but these approaches limit

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50420-5_40

the results to a tree or forest structure. They also limit the computational methods used for differential analysis at individual nodes to univariate statistical tests, and are not applicable to directed graphs, such as kinase-substrate network that describes protein signaling. Graphical models such as Bayesian networks [31] can in principle deal with unmeasured nodes, but their applicability is limited by their computational complexity.

Other algorithms for network-oriented analysis typically assume that all nodes in the reference network correspond to measured endpoints or, conversely, that the reference network provides direct edges between the measured nodes and does not leave nodes only connected to unmeasured nodes. Simply eliminating all the unmeasured nodes from consideration is a poor option as it fragments the reference network and leaves many nodes unconnected. The alternative simple approach of adding a direct edge between all pairs of measured nodes that are connected by a path in the reference network will result in a dense graph where connections lose their specificity. Data-driven network inference algorithms commonly used to predict regulatory networks based on gene expression [19] can provide a graph linking the measured nodes, but then the network is grounded in data and not in existing biological knowledge, which prevents its use as an additional source of information to complement the experimental results.

Finding a network of connections between measured nodes based only on a given topology of a larger network consisting of nodes with and without experimental measurements is thus a non-trivial new problem. It can be stated in the following way. Given a reference directed network $G = (V, E)$ and a subset of measured nodes $W \subset V$ that is often much smaller than V, find a new network $G_W = (W, E')$ that captures best the biological processes, such as regulation or signaling, described by G. The new graph G_W should not be based on experimental data, but only on the set of known interactions represented by edges of the original graph G. A graphical illustration of the problem is shown in Fig. 1.

Since the signal in molecular networks can spread through many paths, we want to take them into account in a way that avoids making the graph too dense. In a simple strategy for addressing the problem, we could place a direct edge from measured node $i \in W$ to measured node $j \in W$ in the output network G_W if and only if the underlying network G has a path connecting i to j that does not pass through any other measured node from W. Another strategy could place an edge from i to j in G_W when that unmeasured path is a shortest path. Here, we propose a method that produces networks that are easier to visualize and more interpretable than networks produced by these simple strategies, and at the same time have higher agreement with the underlying biology.

3 Proposed Method

In the proposed method, we treat each measured node in the network as a source and aim to find other measured vertices that would, in the new network being constructed, serve as targets of direct edges from that source measured node. Specifically, for a given source measured node, our goal is to identify a group of

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50420-5_40

measured nodes that are hit first as the signal from the source spreads in the reference network. Those nodes will be connected directly to the source. On the other hand, measured nodes that are reachable from the source but most of the signal passing to them traverses first through other measured nodes will not be connected to the source directly. This intuition leads to a solution that is based on mean first-passage times in a semi-lazy random walk on a directed graph.

3.1 Mean First-Passage Time in Directed Graphs

The mean first-passage time $H(i,j)$, known also as the expected hitting time, from node i to j in a strongly connected, directed graph is defined as the expected number of steps it takes for a random walker starting from node i to reach node j for the first time, where the walk is Markov chain defined by transition probabilities resulting from the graphs connectivity. The average is taken over the number of transitions, that is, lengths L of all paths $s_{(i\to j)}$ from i to j that do not contain a cycle involving j, with respect to probabilities P of the paths:

$$H(i,j) = \sum_{s_{(i\to j)}} P(s_{(i\to j)})L(s_{(i\to j)}). \tag{1}$$

Compared to the shortest path from i to j, the mean first-passage time includes multiple paths and node degrees into consideration. For example, paths through hub nodes increase H, since the walker has high probability of moving to nodes connected to the hub that are not on the shortest path to the target.

The study of mean first-passage time on various domains has long history in physics [24]. It has also been well-characterized for undirected graphs [10]. Recently, it has been shown that for directed graphs, mean first-passage time $H(i,j)$ can be obtained analytically in close form given the Laplacian matrix and node stationary probabilities in a random walk in the graph [4]. More specifically, let A be the, possibly weighted, adjacency matrix of the input strongly connected, directed graph, D a diagonal matrix of node out-degrees, and I an identity matrix. Then, the expected hitting time can be calculated as [4]:

$$H(i,j) = M(j,j) - M(i,j) + \sum_{k\in V}(M(i,k) - M(j,k))\pi(k). \tag{2}$$

where $\Pi = Diag(\pi)$ is the matrix of node stationary probabilities, $P = D^{-1}A$ captures node transition probabilities, and $M = L^{+}$ is defined as the Moore-Penrose pseudo-inverse of the assymetric graph Laplacian $L = \Pi(I - P)$.

3.2 Semi-Lazy Random Walk and Mean First-Passage Time

Assume we have an unweighted strongly connected directed graph $G = (V, E)$ with two types of nodes, $V = U + W$. Nodes in U are regular nodes, which do not affect the behavior of a random walker in the graph. On the other hand, upon arriving at a node from W, the random walker is trapped. In each subsequent

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50420-5_40

step the walker remains at the node with probability γ. That is, in each time step, the walker has probability $1-\gamma$ of escaping the trap and continuing with the walk through other nodes[1]. This bears resemblance to a lazy random walk, in which the random walker stays at a node with probability $\frac{1}{2}$ or, more generally, with some fixed probability. Here, we call the walk semi-lazy, since the random walker is lazy only at nodes from W.

In this setting, mean first-passage time no longer depends only on the topology of the graph, but also on whether the paths contain nodes from W or not. We can define the mean first-passage time for a semi-lazy random walk induced by imperfect traps as:

$$\begin{aligned} H_{IT}(i,j) = & \sum_{s_{(i\to U\to j)}} P(s_{(i\to U\to j)})L(s_{(i\to U\to j)}) \\ & + \sum_{s_{(i\to M\to j)}} P(s_{(i\to W\to j)})[L(s_{(i\to W\to j)}) \\ & + \Delta(s_{(i\to W\to j)})], \end{aligned} \tag{3}$$

where $s_{(i\to U\to j)}$ is any path from i to j that goes only through regular nodes from U and $s_{(i\to W\to j)}$ is any path that includes at least one trap from set W, and Δ is a stochastic penalty function depending on the number of nodes from the set W on the path. By convention, if $i \in W$ then $H_{IT}(i,j)$ is defined as a walk that starts at the point when random walker escapes the trap i, that is, the first step is always a step to some other node.

We calculate H_{IT} for a directed graph with transition probabilities P separately for each starting node i. We create a new transition probability matrix $P'_i = \gamma I_{W,i} + (I - \gamma I_{W,i})P$, where $I_{W,i}$ is a diagonal matrix that has ones for rows and columns corresponding to $W \setminus \{i\}$ and zeros elsewhere. The Markov chain specified by P'_i is irreducible and aperiodic for a strongly connected graph G. Based on P'_i, we calculate node stationary probabilities and the Moore-Penrose pseudo-inverse of the graph Laplacian, and then use Equation 2 to obtain $H_{IT}(i,j)$ for each j.

3.3 Connectivity between Measured Endpoints in Biological Networks

Given a reference biological network and experimental data, we equate the set of traps W with the nodes for which we have experimental measurements and the set U with all other nodes. In this way, if most of the paths from i to j lead through other measured nodes, the mean first-passage time will be much higher than if the paths lead only through non-measured nodes.

First, for every measured node $i \in W$, we calculate $H_{IT}(i,j)$ to all measured nodes $j \in W$. We ignore hitting times from or to non-measured nodes in U. Prior to the calculation of the values $H_{IT}(i,\cdot)$ for starting node i, we eliminate all nodes from U that do not lie on any path from i to any node in W, because

[1] We set the default value of γ to 0.99

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50420-5_40

either they cannot be reached from i, or they do not have a path to any node in W. Also, if there are dangling nodes, that is, nodes with null out-degree, we add a connection from those nodes to i, allowing the walker to continue the walk.

Once $H_{IT}(i,j)$ is calculated for every $i,j \in W$, we treat H_{IT} as a weighted adjacency matrix and calculate shortest paths $\sigma(i,j)$ for $i,j \in W$. Finally, we create the output graph G_W by keeping edges $i \to j$ for which there is no shorter path in H_{IT} than the direct edge:

$$\forall i,j \in W: \quad G_W(i,j) = 1 \quad \text{iff} \quad H_{IT}(i,j) = \sigma(i,j). \tag{4}$$

In effect, we place a direct edge from i to j if the random walker starting from i tends to avoid other nodes from W on its way towards hitting j. If some other node k is often encountered during the $i \to j$ walk, then $H_{IT}(i,k) + H_{IT}(k,j) < H_{IT}(i,j)$ since the trap at k is not considered when quantifying hitting times $H_{IT}(i,k)$ and $H_{IT}(k,j)$, but it is considered when estimating $H_{IT}(i,j)$. In this way, the new graph G_W will contain only edges between measured nodes, and the edge structure will be based on connectivity in the original reference network in a way that keeps connections through measured nodes explicit and avoids indirect connections.

The computational complexity of the proposed method is $\mathcal{O}(|W||V|^3)$. For each node $i \in W$, calculating $H_{IT}(i,\cdot)$ involves finding the pseudoinverse of the Laplacian and estimating the stationary node probabilities, which both are $\mathcal{O}(|V|^3)$. Calculations for different i can be done independently in parallel, and need to be followed by all-pairs shortest path involving W nodes, which requires $\mathcal{O}(|W|^3)$. In effect, the method can be successfully applied to biological networks, which have on the order of 10^4 nodes or less.

4 Experimental Validation

We evaluated our method by comparing it with two alternative strategies. In the connectivity-based strategy, we place an edge from measured node i to measured node j in the output network if and only if the underlying network has a path connecting i to j that does not pass through any other measured node. In the shortest-paths-based strategy, we place an edge from i to j in the output network when a shortest path from i to j in the underlying network does not pass through any other measured node.

To compare the quality of the network of measured nodes resulting from the proposed method and a network returned by an alternative strategy, we used expression data measured over a set of samples. In both networks each node is associated with vector of expression values of a corresponding gene or protein. For each edge $i \to j$ in both networks, we calculated the p-value of the correlation between expression vector associated with i and the expression vector associated with j. We treat the correlation as an imperfect but easy to obtain surrogate measure of edge quality. We assume that when comparing two graphs inferred from the same reference network G without looking at expression data, the one

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50420-5_40

Table 1. Comparison of the proposed approach with two alternative strategies for 5 simulated and 1 real-world dataset. Columns are: $\overline{\mathbf{R}}$: mean -log(p-value) of correlation between endpoints connected by edges from set **R**, that is, present in results of an alternative strategy and retained by our method; $\overline{\mathbf{F}}$: mean -log(p-value) of correlation between endpoints connected by edges from set **F**, that is, present in results of an alternative strategy but filtered out by our method; p-value for a test if the means of the negated log-transformed p-values in **R** are higher than in **F**, that is, if the expression profiles for nodes linked by retained edges are more highly correlated than for nodes linked by filtered out edges; #**R**: number of edges in **R**; #**F**: number of edges in **F**.

Comparison with connectivity-based strategy					
Dataset	$\overline{\mathbf{R}}$	$\overline{\mathbf{F}}$	p-value	#**R**	#**F**
DREAM 4 I	2.89	1.28	3.47e-4	1861	165
DREAM 4 II	2.93	0.98	3.78e-6	2071	184
DREAM 4 III	4.52	2.35	5.11e-18	3734	1442
DREAM 4 IV	3.36	1.52	7.10e-39	3602	2345
DREAM 4 V	4.87	2.84	1.69e-8	2482	386
TCGA BRCA	13.52	10.55	0.0110	156	533
Comparison with shortest-path-based strategy					
Dataset	$\overline{\mathbf{R}}$	$\overline{\mathbf{F}}$	p-value	#**R**	#**F**
DREAM 4 I	2.89	1.22	4.57e-4	1858	147
DREAM 4 II	2.93	0.98	1.47e-5	2068	161
DREAM 4 III	4.54	2.37	1.44e-14	3706	1058
DREAM 4 IV	3.37	1.46	1.86e-31	3580	1689
DREAM 4 V	4.87	2.88	1.69e-6	2473	283
TCGA BRCA	13.52	10.67	0.0364	156	240

that has higher correlation between expression of genes or proteins linked by the graph edges represents the underlying signaling or regulation better.

Since edges detected by our method are a subset of edges detected by the alternative strategies, we partitioned the edge p-values into two groups. In the retained edges group, **R**, we put the p-values of edges that are found both by the proposed method and by the alternative strategy used for comparison. In the filtered-out edges group, **F**, we put the p-values of edges detected only by the alternative method but not by the proposed method. Then, we tested if the mean of p-values in the **R** group is lower than mean in the **F** group, that is, if the proposed method is effective at filtering out low p-value edges.

4.1 Simulated Data

In our evaluation, we used simulated expression data, for which the expression profiles come from known, pre-specified networks connecting genes, and are simulated using a system of differential equations. We used networks and data from GeneNetWeaver [25] available as part of the DREAM 4 In Silico Network Challenge [18]. We used the five multifactorial directed networks from the challenge, each with 100 nodes. Each network is accompanied by simulated expression data

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50420-5_40

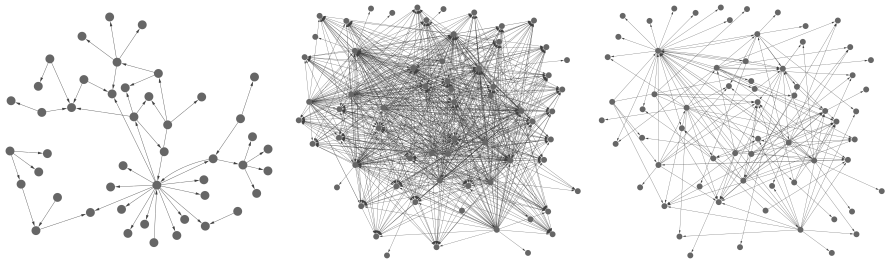

Fig. 2. Visualization of human kinase phospohrylation network of 1191 proteins for a dataset of 69 proteins measured using reverse phase protein array. The approach of using direct edges between measured proteins results in a network that only presents 46 out of 69 measured proteins and thus leaves 33% of measured proteins out of the picture (left). The alternative approach of placing an edge between all directly or indirectly connected measured proteins as long as there is a path between them that does not pass through any other measured proteins results in an uninterpretable, dense network with 689 edges (center). The proposed algorithm results in a sparse, interpretable network (right) that connects all 69 measured nodes through a set of 156 direct or indirect connections chosen based on mean first-passage time criterion.

for the 100 endpoints in 100 samples. To test our method, we randomly picked 20 nodes, kept their expression data, and ignored the expression data for the remaining 80 nodes. The task for our method is to connect those 20 nodes based on the known network of all 100 nodes, without looking at the data. We repeated the experiment 100 times with different random samples of measured nodes, and grouped the p-values for the discovered edges together.

We carried out the above procedure independently for each of the 5 networks available in DREAM 4. The results are presented in Table 1. In each of the simulated networks, the p-values of the edges retained by our methods are significantly lower on average than those that are filtered out, compared to edges picked by the two alternative strategies.

4.2 Real-world Data

We validated the proposed methods using protein expression data measured using RPPA assays for 410 samples from breast cancer patients gathered from the Cancer Genome Atlas (TCGA) [15]. As the underlying reference network, we used a recently published directed human phosphorylation network that captures protein signaling [20]. The network has 1191 nodes, of which 69 have corresponding protein measured with reverse phase protein array in the TCGA samples. The task for our method is to connect those 69 nodes based on the known network of all 1191 nodes.

We used the same approach as above to compare the connectivity between the 69 nodes resulting from our methods with the connectivity from the alternative strategies. As shown in Table 1, our method performs significantly better. The

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50420-5_40

number of edges returned by the proposed method is only 156, whereas the connectivity-based strategy returns a dense structure of 689 edges for a 69 node graph, and the shortest-path-based strategy returns 396 edges. As seen in Figure 2, the network returned by the method is much more interpretable than the network resulting from the strategy of placing an edge between all connected nodes. The pairs of proteomic endpoints connected by the edges retained by the proposed method are on average more highly correlated than those connected by the edges from alternative strategies we filtered out.

5 Discussion

We have proposed an approach for visualizing, in a compact way, biological networks in scenarios when only some subset of nodes has measurements available. Our approach is based on theory of random walks [3]. Random walks have been previously used for estimating influence between nodes in biological networks [27, 32, 30, 29, 16, 1, 28, 9]. The influence has been defined in terms of a diffusion kernel [17], diffusion with loss [23] or a heat kernel [8], but these kernels are defined for undirected graphs, which reduces their use for directed networks such as kinase-substrate protein signaling network or gene regulatory networks. These measures of influence also ignore the time progression associated with the spread of signal in the network, since they are based on the stationary state of the random walk.

The progression of the random walk can be quantified using mean first-passage times for individual nodes. In computational biology, it has been used previously for analyzing state transition graphs in probabilistic Boolean networks to identify genes perturbations that lead quickly to a desired state of the system [26]. Here, we proposed to use it to decide if signaling from one measured node to another measured node typically passes through other measured nodes. This task bears similarities to the problem in physical chemistry of finding reaction paths from a reactant to a product. Mean first-passage time has been used as one way of solving that problem for reactions with continuous or discrete reaction coordinates [21], for example to uncover the path of excitation migration after photon absorption in photosynthetic complex. Our approach could be viewed as an exploration of reaction paths on a cellular scale where the reaction coordinates are nodes in a directed graph.

The analogy between the graph problem explored here and the chemical reaction path detection problem indicates that the mean first-passage time could be an effective way of representing paths in the underlying network by direct edges between measured nodes. Experimental validation using simulated and real networks and data show that this is indeed the case. The uncovered connectivity structures better approximate the underlying biology that other strategies we used for comparison. With the proposed approach, for a specific experimental study, one can obtain a dedicated network that includes only nodes for which experimental data is measured in the study, linked by edges representing causal interactions based on known connections from the reference network. The net-

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50420-5_40

work can then be used as input for network-oriented analyses, or for compact, interpretable visualization of the relationships between measured nodes.

Acknowledgements

TA is supported by NSF grant IIS-1453658.

References

1. Arodz, T., Bonchev, D.: Identifying influential nodes in a wound healing-related network of biological processes using mean first-passage time. New Journal of Physics **17**(2), 025002 (2015)
2. Bailly-Bechet, M., Borgs, C., Braunstein, A., Chayes, J., Dagkessamanskaia, A., François, J.M., Zecchina, R.: Finding undetected protein associations in cell signaling by belief propagation. Proceedings of the National Academy of Sciences **108**, 882–887 (2011)
3. Berg, H.C.: Random walks in biology. Princeton University Press (1993)
4. Boley, D., Ranjan, G., Zhang, Z.L.: Commute times for a directed graph using an asymmetric Laplacian. Linear Algebra and its Applications **435**, 224–242 (2011)
5. Chen, T., He, H., Church, G.: Modeling gene expression with differential equations. In: Pacific Symposium on Biocomputing. pp. 29–40 (1999)
6. Chowdhury, S., Nibbe, R., Chance, M., Koyutürk, M.: Subnetwork state functions define dysregulated subnetworks in cancer. In: Research in Computational Molecular Biology. pp. 80–95 (2010)
7. Chuang, H., Lee, E., Liu, Y., Lee, D., Ideker, T.: Network-based classification of breast cancer metastasis. Molecular Systems Biology **3**, 140 (2007)
8. Chung, F.: The heat kernel as the pagerank of a graph. Proceedings of the National Academy of Sciences **104**, 19735–19740 (2007)
9. Cowen, L., Ideker, T., Raphael, B.J., Sharan, R.: Network propagation: a universal amplifier of genetic associations. Nature Reviews Genetics **18**(9), 551 (2017)
10. Göbel, F., Jagers, A.: Random walks on graphs. Stochastic Processes and Their Applications **2**, 311–336 (1974)
11. Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., et al.: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science **286**, 531–537 (1999)
12. Goss, P., Peccoud, J.: Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets. Proceedings of the National Academy of Sciences **95**, 6750–6755 (1998)
13. Ideker, T., Ozier, O., Schwikowski, B., Siegel, A.F.: Discovering regulatory and signalling circuits in molecular interaction networks. Bioinformatics **18**(S1), S233–S240 (2002)
14. Kauffman, S.: Metabolic stability and epigenesis in randomly constructed genetic nets. Journal of Theoretical Biology **22**, 437–467 (1969)
15. Koboldt, D., Fulton, R., McLellan, M., Schmidt, H., Kalicki-Veizer, J., McMichael, J.: Comprehensive molecular portraits of human breast tumours. Nature **490**, 61–70 (2012)
16. Köhler, S., Bauer, S., Horn, D., Robinson, P.N.: Walking the interactome for prioritization of candidate disease genes. American Journal of Human Genetics **82**, 949–958 (2008)

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50420-5_40

17. Kondor, R.I., Lafferty, J.: Diffusion kernels on graphs and other discrete input spaces. In: International Conference on Machine Learning. pp. 315–322 (2002)
18. Marbach, D., Costello, J., Küffner, R., Vega, N., Prill, R., et al.: Wisdom of crowds for robust gene network inference. Nature Methods **9**, 797–804 (2012)
19. Margolin, A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Favera, R., Califano, A.: ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics **7**(S1), S7 (2006)
20. Newman, R.H., Hu, J., Rho, H.S., Xie, Z., Woodard, C., et al.: Construction of human activity-based phosphorylation networks. Molecular Systems Biology **9**, 655 (2013)
21. Park, S., Sener, M.K., Lu, D., Schulten, K.: Reaction paths based on mean first-passage times. Journal of Chemical Physics **119**, 1313–1319 (2003)
22. Paweletz, C.P., Charboneau, L., Bichsel, V.E., Simone, N.L., Chen, T., et al.: Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. Oncogene **20**, 1981–1989 (2001)
23. Qi, Y., Suhail, Y., Lin, Y.y., Boeke, J.D., Bader, J.S.: Finding friends and enemies in an enemies-only network: a graph diffusion kernel for predicting novel genetic interactions and co-complex membership from yeast genetic interactions. Genome Research **18**, 1991–2004 (2008)
24. Redner, S.: A guide to first-passage processes. Cambridge University Press (2001)
25. Schaffter, T., Marbach, D., Floreano, D.: GeneNetWeaver: In silico benchmark generation and performance profiling of network inference methods. Bioinformatics **27**, 2263–2270 (2011)
26. Shmulevich, I., Dougherty, E.R., Zhang, W.: Gene perturbation and intervention in probabilistic Boolean networks. Bioinformatics **18**, 1319–1331 (2002)
27. Tsuda, K., Noble, W.S.: Learning kernels from biological networks by maximizing entropy. Bioinformatics **20**(S1), i326–i333 (2004)
28. Valdeolivas, A., Tichit, L., Navarro, C., Perrin, S., Odelin, G., Levy, N., Cau, P., Remy, E., Baudot, A.: Random walk with restart on multiplex and heterogeneous biological networks. Bioinformatics **35**(3), 497–505 (2018)
29. Vandin, F., Clay, P., Upfal, E., Raphael, B.J.: Discovery of mutated subnetworks associated with clinical data in cancer. In: Pacific Symposium on Biocomputing. pp. 55–66 (2012)
30. Vandin, F., Upfal, E., Raphael, B.J.: Algorithms for detecting significantly mutated pathways in cancer. In: Research in Computational Molecular Biology. pp. 506–521 (2010)
31. Yu, J., Smith, V.A., Wang, P.P., Hartemink, A.J., Jarvis, E.D.: Advances to Bayesian network inference for generating causal networks from observational biological data. Bioinformatics **20**, 3594–3603 (2004)
32. Zhang, W., Lei, X.: Two-step random walk algorithm to identify cancer genes based on various biological data. In: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 1296–1301 (2018)

ICCS Camera Ready Version 2020
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-50420-5_40