The driving regulators of the connectivity protein network of brain malignancies

Tahmassebi, Amirhessam, Pinker-Domenig, Katja, Wengert, Georg, Lobbes, Marc, Stadlbauer, Andreas, et al.

Amirhessam Tahmassebi, Katja Pinker-Domenig, Georg Wengert, Marc Lobbes, Andreas Stadlbauer, Norelle C. Wildburger, Francisco J. Romero, Diego P. Morales, Encarnacion Castillo, Antonio Garcia, Guillermo Botella, Anke Meyer-Bäse, "The driving regulators of the connectivity protein network of brain malignancies," Proc. SPIE 10216, Smart Biomedical and Physiological Sensor Technology XIV, 1021605 (16 May 2017); doi: 10.1117/12.2263554

SPIE.

Event: SPIE Commercial + Scientific Sensing and Imaging, 2017, Anaheim, California, United States

The Driving Regulators of the Connectivity Protein Network of Brain Malignancies

Amirhessam Tahmassebi[a], Katja Pinker-Domenig[a,b,c], Georg Wengert[b], Marc Lobbes[d], Andreas Stadlbauer [e], Norelle C. Wildburger[f], Francisco J. Romero[g], Diego P. Morales[g], Encarnacion Castillo[g], Antonio Garcia[g], Guillermo Botella[h], and Anke Meyer-Bäse[a,d]

[a] Department of Scientific Computing,
Florida State University, Tallahassee, Florida 32310-4120, USA
[b] Medical University of Vienna, Vienna, Austria
[c] Memorial Sloan-Kettering Cancer Center, New York, USA
[d] Department of Radiology and Nuclear Medicine, Maastricht University Medical Center, Maastricht, The Netherlands
[e] Department of Neurosurgery, University of Erlangen-Nürnberg, Germany
[f] Department of Neurology, Washington University, St. Louis, Missouri, USA
[g] Department of Electronics and Computer Technologies, Facultad de Ciencias, Univ. of Granada, Granada, Spain
[h] Department of Computer Architecture, Universidad Complutense de Madrid, Madrid, Spain

ABSTRACT

An important problem in modern therapeutics at the proteomic level remains to identify therapeutic targets in a plentitude of high-throughput data from experiments relevant to a variety of diseases. This paper presents the application of novel modern control concepts, such as pinning controllability and observability applied to the glioma cancer stem cells (GSCs) protein graph network with known and novel association to glioblastoma (GBM). The theoretical frameworks provides us with the minimal number of "driver nodes", which are necessary, and their location to determine the full control over the obtained graph network in order to provide a change in the network's dynamics from an initial state (disease) to a desired state (non-disease). The achieved results will provide biochemists with techniques to identify more metabolic regions and biological pathways for complex diseases, to design and test novel therapeutic solutions.

Keywords: Graph theory, nonlinear dynamics, dynamic graph, pathway analysis and control, glioma cancer stem cells

1. INTRODUCTION

Tumor-initiation and maintenance in several cancers including glioblastoma may be driven by a small subset of cells, the so-called cancer stem cells (CSC). These cells may be refractory to radiation and chemotherapy, and thus have important implications for tumor biology and therapeutics. Therefore, knowledge of CSC signaling pathways responsible for maintenance of tumor stem cells may lead to novel therapeutic targets for successful cancer treatment.

Metabolomic, transcriptomic and phosphoproteomic measurements provide a wealth of information about the biochemical state of cells and play an increasing role to elucidate the mechanism and importance of novel therapeutics in cancer research and other lethal diseases. However, this detailed information does not automatically provide a deep insight into how to use this acquired knowledge in reverse engineering and drug design. Therefore, it is very useful to determine the nodes in a network that offer us full access to control the dynam-ics of the network and represents "key points" of external signals, which are applicable to networks, in order to change the dynamic status from an initial (disease) state into a final (disease-free) state. Other successful applied methods are based on novel exploratory combinatorial optimization algorithms, as an analytical tool for the interpretation of such large data sets in structural biology.[1–3]

Smart Biomedical and Physiological Sensor Technology XIV, edited by
Brian M. Cullum, Douglas Kiehl, Eric S. McLamore, Proc. of SPIE Vol. 10216,
1021605 · © 2017 SPIE · CCC code: 0277-786X/17/$18 · doi: 10.1117/12.2263554

The goal of our paper is to determine the driver nodes in graph networks obtained from experimental data in connection with glioma cancer stem cells (GSCs). We apply the algorithms developed in[4] that are robust and efficient for large-scale experiments and the design of novel therapeutic solutions for GSCs. We anticipate that these algorithms will become a standard tool for biochemists to identify more metabolic regions and biological pathways for complex diseases, such as brain cancer. The cellular mechanisms and experimental designs that were the basis of the graph networks developed in[5–7] are shown in Figure 1. This represents the most detailed systems biology study of GSCs differentiation known so far.[8,9]

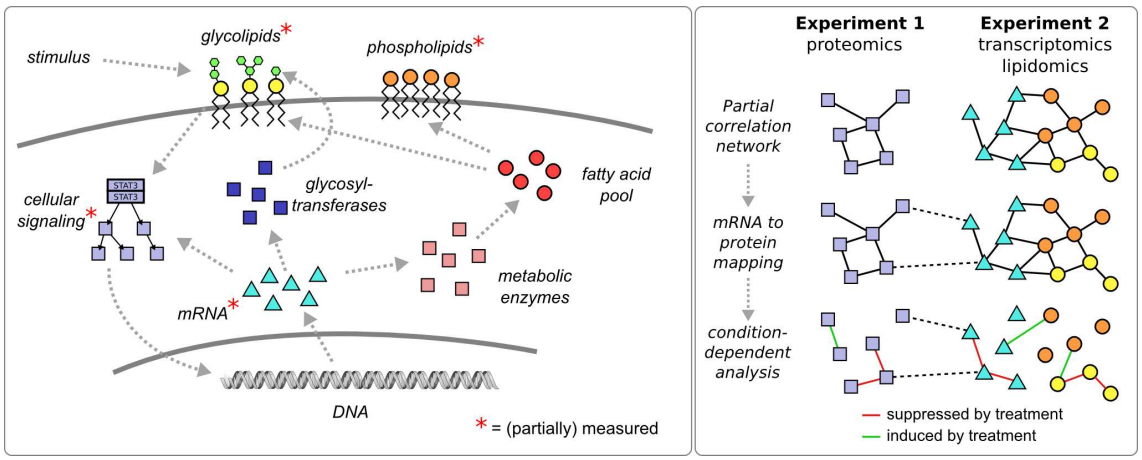
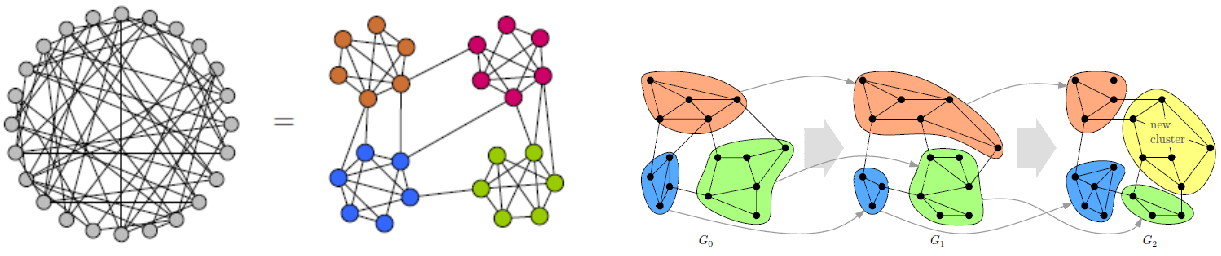


Figure 1. Illustration of the cellular mechanisms and experimental designs of the GSCs.[10]

New control paradigms become imperative when analyzing and interpreting vast experimental data sets with respect to developing novel therapeutics. Graph theory represents a powerful method to visualize and target in combination with modern control theory relevant nodes in the resulting graph. Previously we applied it in the form of static graph theory to show the differences in treatments in GSCs.[5–7] Mathematically, graph networks are defined as relations among a bounded set of nodes with the the typical data model being a graph $G = (V, E)$ with vertices V and edges E representing relations between the nodes. In addition to graph theory, the empirical nature of the field imposes statistical approaches as a complementary tool. While static graphs give a snapshot of a single representation, dynamic graphs describe the temporal evolution of relations among nodes as shown in Figure 2.

2. DYNAMICAL CONTROL STRATEGIES

Network controllability is becoming an important area in molecular therapeutics. The current methods applicable to manipulating signaling pathways are (1) partial mixed-valued control[12] and (2) nonlinear dynamical graph



(a) (b)

Figure 2. Graph structures. (a) Graphs excel at hiding their structure. Graph clustering aims at revealing their structure. (b) Time-dependent graph clustering. Three time steps of a dynamic graph: smooth dynamic clustering and cluster tracking over time (gray arrows).[11]

theory to determine driver nodes in networks or reach a consensus.[4, 13]

Traditional quantitative or semi-quantitative studies are not sufficient to understand the dynamic properties of biological networks and to quantitatively predict suitable targets that would alter the responses of the graph networks for therapeutic purposes. To control these dynamic models to achieve desired therapeutic responses, novel concepts from modern control theory can be employed.

2.1 Consensus Dynamics, Pinning Control, and Driver Nodes in Complex Networks

The most intriguing question when analyzing a dynamic graph network is the role of each node. To reach therapeutic efficacy we need to "drive" a regulatory network from an existing disease-state to an optimal disease-free state. The complexity of the networks poses many limitations to traditional analysis tools:[14] (1) most graph networks are directed, (2) the size of the network does not allow testing of several combinations to determine driver nodes, and (3) the weights between nodes are not equal and time-dependent. Modern control theory[4, 13] provides many tools to control such a network and thus successfully implement a therapeutic strategy. In the parlance of control theory, tools are described that are able to identify the set of driver nodes and thus guide the network's entire dynamics.

We introduce a weighted directed graph $G = (V, E, A)$ of order N that has a set of nodes $V = \{v_1, \cdots, v_N\}$, a set of directed edges $E \subseteq V \times V$, and a weighted adjacent matrix $A = (a_{ij})_{N\times N}$. The Laplacian matrix $L = (L_{ij})_{N\times N}$ of the graph is defined as $L_{ij} = -a_{ij}$ for $i \neq j$, with $i, j \in \{1, \cdots, N\}$ and $L_{ii} = k_i^{in}$ for $i \in \{1, \cdots, N\}$, and $k_i^{in} = \sum_{j=1, i\neq j}^{N} G_{ij}$, represents the sum of all afferent edges. It's evident that $\sum_{j=1}^{N} L_{ij} = 0$ for all $i = 1, 2, \cdots, N$.

We define the consensus problem as a modality to reach an agreement between a group of autonomous agents, in our case the nodes, when these change dynamically.

Mathematically, the consensus protocol in a multi-node system is defined as:

$$\dot{x_i}(t) = \sum_{j\neq i} a_{ij}(x_j(t) - x_i(t)) = -\sum_{j=1}^{N} L_{ij}x_j(t) \tag{1}$$

where $x_i(t) \in R^n$ is the state of the node. $L = L(t)$ is a time-varying matrix when the graph network topology changes over time.

Assuming that the dynamics of the node is nonlinear,[15] then the state equation becomes

$$\dot{x_i}(t) = f(x_i(t)) - c\sum_{j=1}^{N} L_{ij}\Gamma x_j(t) \tag{2}$$

with $f() \in R^n$ representing the nonlinearity, c the coupling strength, and $\Gamma = diag(\gamma_1, \cdots, \gamma_n) \in R^{n\times n}$ being a semi-positive definite diagonal matrix with $\gamma_j > 0$. If $\gamma_j \neq 0$ means that the nodes can communicate through their jth state.

A desired trajectory to be reached by the system, corresponding to a therapeutical solution, is defined as

$$\dot{s}(t) = f(s(t)) \tag{3}$$

where $s(t)$ is an isolated equilibrium point. To achieve this equilibrium point, the new evolving equation becomes

$$\dot{y_i}(t) = f(x_i(t)) - f(s(t)) - c\sum_{j=1}^{N} L_{ij}\Gamma y_j(t) \tag{4}$$

where $y_i = x_i - s_i$. The pinning control strategy is to guide the network to the desired state $s(t)$. The controllability of the system is evaluated based on the algebraic connectivity. Measures derived from the smallest

and largest eigenvalue of the connecting matrix are essential to determine the success of controllability. The number of controlling nodes is smaller than the number of total nodes in the network and a direct control is possible only at these nodes and then propagated to the rest of network through vertices.

The theoretical results in[4,13] have shown that: (a) nodes with low degrees should be pinned first and not hubs, which are usually of high degree, and (b) the minimum number of nodes to be selected for control can be theoretically determined. In large real-world networks, however, the detection of controlling regions becomes a constrained optimization problem.[16] These results are valid for both directed and undirected graphs.

2.2 Controllability of Complex Networks

In[4] a different approach was proposed to study the controllability of complex networks.

The networks were modeled as a linear system

$$\dot{x}t = Ax(t) + Bu(t) \tag{5}$$

where $x(t) \in R^N$ is the state of the system, $u(t) \in R^M$ is the input, A is an $N \times N$ state matrix, and B is an $M \times N$ input matrix.

The system described in equation 5 is said to be controllable if the controllability $N \times NM$ controllability matrix C

$$C = (B, AB, A^2B, \cdots, A^{N-1}B) \tag{6}$$

has full rank, that is $rank(C) = N$.

However, to apply the above theory to a complex network means complete knowledge of the network's weights, which is for most real-world networks almost impossible. To overcome this problem, the number of driver nodes was determined in[4] based on the cavity method. This method computes the number of driver nodes over all network realizations compatible with the input degree distribution. The main hypothesis is that the control of the hubs is essential for the controllability of the network. Hubs are known in static graph theory as nodes of high degree, which are important for structural integrity of network against failures.

The main result is that the number of driver nodes n_D is determined mainly by the in/out node-degree distribution. By employing the cavity method, a tool from statistical physics, the n_D could be predicted from the incoming and outgoing degree of distribution of a given node $P(k_{in}, k_{out})$ analytically. The determined number is shown for the most common networks in Figure 3. It has been shown that in several real systems (regulatory and neuronal) driver nodes tend to avoid hubs.

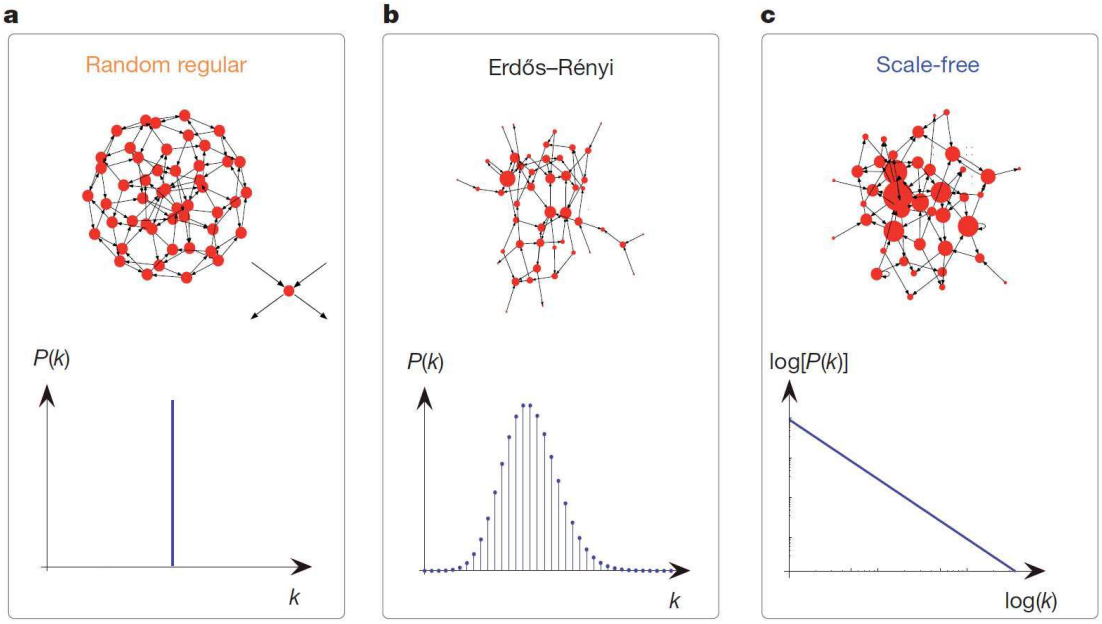
Summarizing, the most important findings are: (a) the denser a network, the fewer driver nodes are necessary to control it, (b) sparse and heterogeneous networks require the most driver nodes, and (c) not every network is controllable.

2.3 Observability of Complex Networks

Pinning observability in a neural network pertains to observing very few neurons such that the states of the other neurons can be recovered. Differently from the widely known concept of pinning controllability, the dynamics of the neurons can be heterogeneous. A connectivity network is a very complex large-scale network and therefore impossible to observe all nodes for recovering the network by reaching synchronization between an original and a newly constructed network. To achieve synchronization only a small number of nodes need to be observed in order to recover the other states.

We consider in the following the general linear connectivity network equation describing the temporal evolution of the nodal states for the ith node of a N–node network is:

$$\dot{x}_i = -A_i x_i + \sum_{j=1}^{n} d_{ji} x_j \qquad i = 1, ..., N \tag{7}$$



a
Random Regular:
degree homogeneous network

b
Erdos-Renyi:
heterogenous,
Poisson degree distribution
$n_D \approx e^{-\bar{k}/2}$

c
Scale free:
heterogenous,
power law degree distribution
$n_D \approx e^{\left[-\frac{1}{2}\left(1-\frac{1}{\gamma-1}\right)\bar{k}\right]}$

Figure 3. Network structure and number of driver nodes.[4] n_D represents the number of driver nodes, $\bar{k}$ the average node degree and γ is the degree exponent. Reprinted from[4] with permission from Nature Publishing Group.

where $x_i = (x_{i1}(t), \cdots, x_{in}(t))^T \in R^n$ is the state vector of the node i, $A_i \in R^n$ is a matrix. d_{ji} represents an edge connection parameter between the jth node and the ith node and is defined as the matrix $\bar{D} = (d_{ij})$.

Pinning observability is applied only to a small number of nodes l with $0 < l < 1$, thus we obtained a pinning observable network supposing these first l nodes are selected which is formulated mathematically as:

$$\dot{\tilde{x}}_i = -a_i \tilde{x}_i + \sum_{j=1}^{n} \tilde{d}_{ji} \tilde{x}_j \qquad i = 1, ..., n \tag{8}$$

where

$$u_i = -d_i(\tilde{x}_i(t) - x_i(t)) \qquad 1, \cdots, l \tag{9}$$

are n-dimensional linear feedback controllers with the control gains $d_i > 0, i = 1, \cdots, l$ and $d_i = 0$ for the other control gains, $i = l + 1, \cdots, N$.

The theorem for pinning observability was given in[17] and states:

THEOREM 2.1. *Let $A = A^T$ and $L = \tilde{D} = \tilde{D}^T$ and we also assume that Γ is an N-dimensional identity matrix. Assuming a scale-free network topology, the two networks (7) and (8) are globally synchronized if the following condition is satisfied*

$$A - D \otimes I_n + K \cdot \tilde{D} \otimes \Gamma < 0 \tag{10}$$

where $A = \begin{pmatrix} A_1 & \cdots & 0 \\ 0 & \cdots & A_N \end{pmatrix}$, $D = \begin{pmatrix} d_1 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & d_l & \cdots & 0 \\ 0 & \cdots & 0 & \cdots & 0 \end{pmatrix}$, $L = (l_{ij})_{n\times n}$ *and* I_n *being an* n*-dimensional identity matrix.*

As shown in,[17] it is possible to show the number of nodes that can be observed. The following corollary gives a condition to check for each node, whether it can be controlled or not, without involving the other nodes.

A very simple condition with a fixed constant c is given: if $d_i = 0$ and the condition given in the below equation is satisfied for a node i, then the node may not be controlled. Otherwise, if this condition is not satisfied for node i, then the node can be controlled.

$$\begin{aligned} \Xi_i(c) &= \begin{pmatrix} A_i - d_i I_n + \tilde{d}_{ii} I_n + \frac{c}{2} \sum_{j=1, j\neq i}^{N} \tilde{d}_{ij} I_n & \sqrt{\sum_{j=1, i\neq j}^{N} |\tilde{d}_{ij}|} \\ \sqrt{\sum_{j=1, i\neq j}^{N} |\tilde{d}_{ij}|} & -2c \end{pmatrix} < 0, \\ i &= 1, 2, \cdots, N \end{aligned}$$

The above equation shows a simple modality to find the few nodes that have to be controlled such that synchronization is reached.

3. UNDERSTANDING THE REGULATORY MECHANISM OF GSC AT THE PROTEIN LEVEL

We apply the theoretical controllability and observability aspects to the network showing known and novel proteins in a recently published GSC proteomics dataset.[18] Ingenuity Pathway Analysis (IPA) was employed to determine predicted upstream regulators. They identified symplekin (SYMPK) as a known protein, and in addition, they identified novel putative upstream regulators interleukin-5 (IL5) and synoviolin (SYVN1).

The theoretical analysis of the IPA established protein network revealed in addition to this the following insight: the controllability criterion states that there are at least two driver nodes and the observability theorem depicts three nodes that can be controlled. These nodes are: SYVN1, SYMPK and galectin-1 (LGALS1). The experimental results described in[18] are here analyzed and the driver nodes are shown in Figure 4.

SYVN1 protects neurons from apoptosis and serves as a negative regulator of TP53 by sequestering it in the cytoplasm, enabling its degradation. SYMPK is considered to be a useful target in the design of a novel chemotherapy-based treatment strategy for GBM. LGALS1 is known to suppress the immune system based on inducing the apoptosis of T-cells. At the same time, increased levels of LGALS1 at the tumor margin are linked to invasivity.[18] Thus being able to control those inputs of the protein connectivity network, a desired trajectory in cancer therapeutics can be achieved.

4. CONCLUSION AND DISCUSSION

We have shown that many aspects of pinning controllability and observability of dynamical systems can be applied analytically to the connectivity protein networks in glioma cancer stem cells (GSCs). Specifically, our goals were to develop and implement control theory for networks as an alternative to traditional models, to identify the nodes in a graph network that are relevant for controlling the dynamics of the network in order to achieve a desired "therapeutic trajectory". We determined theoretically three important driver nodes that are crucial in influencing apoptosis and resistance to chemotherapy. In summary, we have shown that by combining graph network theory with control theory, we open new avenues to enhance our understanding of complex GSCs systems.

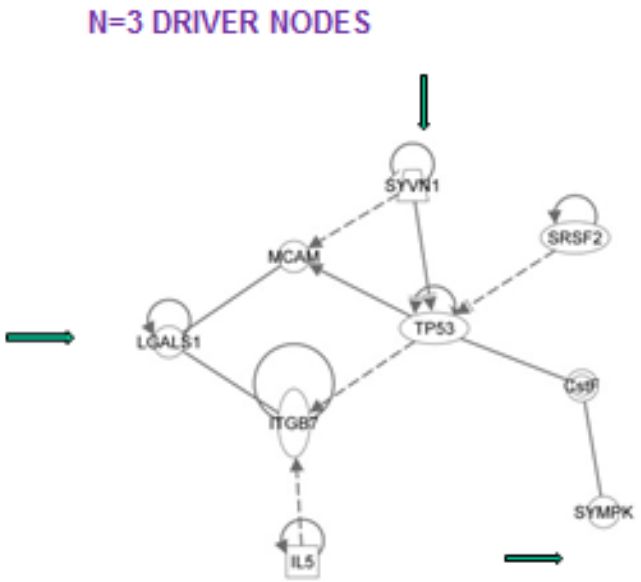


Figure 4. Illustration of the protein connectivity network between known and novel regulators.[18] The number of theoretically determined driver nodes is 3. These are SYVN1, SYMPK, and galectin-1 (LGALS1). Figure adapted from.[18]

REFERENCES

[1] Althaus, E., Canzar, S., Ehrler, C., Emmett, M. R., Karrenbauer, A., Marshall, A. G., Meyer-Baese, A., Tipton, J., and Zhang, H., "Computing h/d-exchange rates of single residues from data of proteolytic fragments," *BMC Bioinformatics* **11**, 11:424 (5 2010).

[2] Althaus, E., Canzar, S., Ehrler, C., Emmett, M. R., Karrenbauer, A., Marshall, A. G., Meyer-Baese, A., Tipton, J., and Zhang, H., "Computing h/d-exchange speeds of single residues from data of peptic fragment," *Proceedings of the 2008 ACM symposium on Applied computing* **3**, 1273–1277 (5 2008).

[3] Althaus, E., Canzar, S., Ehrler, C., Emmett, M., Karrenbauer, A., Marshall, A., Meyer-Bäse, A., Tippton, J., and Zhang, H., "Discrete fitting of hydrogen-deuterium-exchange-data of overlapping fragments," *International Conference on Bioinformatics and Computational Biology 2009* , 496–502 (4 2009).

[4] Liu, Y., Slotine, J., and Barabasi, A., "Controllability of complex networks," *Nature* **473**, 167–173 (12 2011).

[5] Meyer-Baese, A., Goerke, R., Wagner, D., He, H., Emmett, M., Marshall, A., and Conrad, C., "Determining and interpreting correlations in lipidomic networks found in glioblastoma cells," *BMC Systems Biology* **4**, 126–136 (7 2010).

[6] Meyer-Baese, A., Goerke, R., Wagner, D., He, H., Emmett, M., Marshall, A., and Conrad, C., "Determining and interpreting correlations in lipidomic networks found in glioblastoma cells," *SPIE Symposium Computational Intelligence* **7704**, 770406 (7 2010).

[7] Saalbach, A., Lange, O., Nattkemper, T., and Meyer-Bäse, A., "On the application of (topographic) independent and tree-dependent component analysis for the examination of dce-mri data," *Biomedical Signal Processing and Control* **4**, 247–253 (1 2009).

[8] He, H., Nilsson, C., Emmmett, M., Marshall, A., Kroes, R., Moskal, J., Ji, Y., Coleman, H., Priebe, W., Lang, F., and Conrad, C., "Glycomic and transcriptomic response of gsc11 glioblastoma stem cells to stat3 phosphorylation inhibition and serum-induced differentiation," *Journal of Proteome Research* **27**, 2098–2108 (11 2010).

[9] Nilsson, C., Dillon, R., Devakumar, A., Shi, S., Greig, M., Rogers, J., Krastins, B., Rosenblatt, M., Kilmer, G., Major, M., Kaboord, B., Sarracino, D., Rezai, T., Prakash, A., Lopez, M., Ji, Y., Priebe, W., Coleman, H., Lang, F., and Conrad, C., "Quantitative phosphoproteomic analysis of the stat3/il-6/hif1 alpha signaling network: An initial study in gsc11 glioblastoma stem cells," *Journal of Proteome Research* **9**, 430–443 (11 2010).

[10] Nilsson, C. L., "Advances in quantitative phosphoproteomics," *Analytical Chemistry* **9**, 735–746 (11 2012).

[11] Goerke, R., *An Algorithmic Walk from Static to Dynamic Graph Clustering*, PhD thesis, PhD thesis Karlsruhe Institute of Technology (2010).

[12] Cheng, D. and Qi, H., "State-space analysis of boolean network," *IEEE Transactions on Neural Networks* **21**, 584–594 (5 2010).

[13] Yu, W., Chen, G., and Cao, M., "Consensus in directed newtorks of agents with nonlinear dynamics," *IEEE Transactions on Automatic Control* **56**, 1436–1441 (5 2011).

[14] Mesbahi, M. and Egerstedt, M., [*Graph Theoretic Methods in Multiagent Networks*], Princeton University Press (2010).

[15] Theis, F. and Meyer-Bäse, A., [*Biomedical Signal Analysis: Contemporary Methods and Applications*], MIT Press (2010).

[16] Tang, Y., Wang, Z., Gao, H., Swift, S., and Kurths, J., "A constrained evolutionary compuattion method for detecting controlling regions of cortical networks," *IEEE/ACM Transactions on computational biology and bioinformatics* **9**, 1569–1581 (1 2012).

[17] Yu, W., Wen, G., Lü, L., and Yu, X., "Pinning observability in complex networks," *IET Control Theory and Applications* **8**, 2136–2144 (12 2014).

[18] Lichti, C., Wildburger, N., Liu, H., Shavkunov, A., Mostovenko, E., Sulman, E., and Nilsson, C., "The proteomic landscape of glioma stem-like cells," *EuPA Open Proteomics* **13**, 85–93 (11 2015).