

One Feature Doesn't Fit All: Characterizing Topological Features of Targets in Signaling Networks

Huey Eng Chua
School of Computer
Engineering
Nanyang Technological
University, Singapore
CHUA0530@ntu.edu.sg

Sourav S. Bhowmick
School of Computer
Engineering
Nanyang Technological
University, Singapore
assourav@ntu.edu.sg

Lisa Tucker-Kellogg
Duke-NUS Graduate Medical
School
National University of
Singapore, Singapore
lisa.tucker-
kellogg@duke-
nus.edu.sg

ABSTRACT

A key challenge facing drug discovery is the identification of *target(s)* in a signaling network whose perturbation results in a desired therapeutic outcome. Recent studies have shown that analysis of biological networks based on topology can facilitate target identification by providing valuable information on characteristics of targets. In this paper, we present an algorithm called DIFFER that discovers the *discriminative topological features* (DTF) from a signaling network to distinguish the targets from the non-targets. Our empirical study on five signaling networks reveals that the majority of DTFs are able to identify most of the known targets in these networks. Furthermore, they are *distinct* for different networks. That is, no single topological feature can characterise targets in all signaling networks. This is in contrast to the findings in [28] where *bridging nodes* are considered to be good targets with low lethality across several PPI networks.

Categories and Subject Descriptors

J.3 [Life and Medical Sciences]: Biology and genetics

General Terms

Algorithms, Measurement

Keywords

drug target characterization, signaling networks, discriminative topological features

1. INTRODUCTION

Cells use sophisticated communication between proteins in order to perform a variety of cellular functions such as growth, survival, proliferation and development. As signaling proteins rarely operate in isolation through linear pathways, cell signaling can be viewed as a large and complex network. Specifically, the network view emerges due to 'cross-talks' between signaling pathways. Such network

contains numerous features such as feedback and feedforward loops, which render it virtually impossible to manually comprehend how signals are integrated in these pathways. Understanding signal flow in the network is paramount as alterations of cellular signaling events, such as those that arise by gene mutations or epigenetic changes, can result in various diseases. For example, alterations to the genes that encode key signaling proteins, such as *RAS* and *PI3K*, are commonly observed in many types of cancers.

A key challenge towards drug discovery for various complex diseases is the identification of *target(s)* in a signaling network whose perturbation results in a desired therapeutic outcome [26]. Informally, a *target* in the signaling network is a node that, when perturbed, modulates the activity of a specific node, referred to as *output node*. An *output node* is a protein that is either involved in some biological processes which may be deregulated, resulting in manifestation of a disease, or be of interest due to its potential role in the disease (*e.g.*, phosphorylated *ERK* in the *MAPK-PI3K* network [25]).

An intriguing possibility is to explore whether analysis of topology of the signaling network themselves may provide valuable information on characteristics of targets to facilitate their identification. This is more so as recent studies have strengthened the hypothesis that network topology is an essential feature in the emergent system function of the protein when it is perturbed [28]. For instance, Hwang et al. [28] have suggested *bridging nodes* (nodes with high *bridging centrality*) in a protein-protein interaction (PPI) network as potential drug targets, although modulation of the bridging targets themselves may still be indirect. An initial network analysis of the current drug targets of approved drugs indicated that drug targets are commonly highly connected but not essential nodes [52, 92].

Unfortunately, the aforementioned techniques for target characterisation have primarily focused on analyzing topology of PPI networks. For instance, Hwang et al. [28] analyzed the topology of the yeast PPI network, *C21-steroid* hormone metabolism network, *steroid* biosynthesis network and a protein-interaction network of candidate sudden cardiac death susceptibility genes. Unfortunately, edges in PPI networks are undirected; there is neither flow of information nor mass between nodes - an edge simply indicates that two proteins bind [64]. Hence, they may not effectively provide insights into the dynamics of the interacting molecular play-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
BCB'14, September 20–23, 2014, Newport Beach, CA, USA.
Copyright 2014 ACM 978-1-4503-2894-4/14/09 ...\$15.00.
<http://dx.doi.org/10.1145/2649387.2649424>.

the largest shortest path distance [88]. In contrast to *closeness centrality* which uses the set of nodes that a node u can reach (influence range), *proximity prestige* assesses importance based on the set of nodes that can reach u (influence domain).

DEFINITION 3. Given a signaling network $G = (V, E)$, let $I_u \subseteq V$ be the set of nodes having at least one path leading to node u and l_{uv} be the shortest path length between nodes u and v , where $u, v \in V$. Then, the **closeness centrality** β_u , **eccentricity centrality** γ_u and **proximity prestige** μ_u of node u are defined as $\beta_u = \frac{|V|}{\sum_{v \in V} l_{uv}}$, $\gamma_u = \frac{1}{\max\{l_{uv}\}}$, and $\mu_u = \frac{|I_u|}{\sum_{v \in V} l_{vu}}$, respectively.

In a signaling network, the above measures of a node can be used to determine how central it is to the regulation of other nodes in the network [70]. For instance, **SHGs** which lies near the center of the network is well connected to many other nodes in the network. Hence, it has higher *closeness centrality* compared to other nodes (e.g., **MKP3**) that lie near the boundary of the network. Also, nodes with high *eccentricity centrality* are likely to be influential signal transmitters, regulating many other nodes [70]. For instance, **PI3K*** which lies near the center of the network has higher eccentricity centrality compared to other fringe nodes such as **ERK** since the fringe nodes tend to be further away from other nodes in the network. Note that although *proximity prestige* has rarely been considered in the past for biological networks, this does not preclude its importance in this domain. Indeed, as we shall see later in Section 5.1, *proximity prestige* is able to distinguish known targets from non-targets better than commonly-used features such as *clustering coefficient*.

Betweenness Centrality. This feature assigns node centrality value based on the ease in which a node can reach other nodes in the network [9].

DEFINITION 4. Given a signaling network $G = (V, E)$, let $d_{st}(v)$ be the number of shortest paths from nodes s to t passing through v where $s, t, v \in V$. Then, the **betweenness centrality** of v is defined as $\delta_v = \sum_{s \neq v \neq t \in V} \frac{d_{st}(v)}{d_{st}}$.

In a signaling network, these nodes can be considered efficient and crucial signal transmitters as they tend to lie on a majority of the shortest paths between node pairs in the network. For instance, **AktPI3K**, a hub node, has high betweenness centrality in the network as it is well connected to many other central nodes, hence providing fast access to other nodes in the network. Comparatively, nodes (e.g., **MKP3**) that lie on the fringe of the network has low betweenness centrality.

Bridging Centrality and Bridging Coefficient. The *bridging centrality* identifies *bridging nodes* (nodes with high *bridging centrality* value) which are located between functional modules in the signaling network and mediate signal flow between the modules [28]. The *bridging coefficient* measures the average probability of a node transmitting signals to its direct neighbourhood.

DEFINITION 5. Given a signaling network $G = (V, E)$, let $\theta_{total}(v)$ be the total degree of node $v \in V$, N_v be the set of neighbors of v , and η_i be the number of outgoing edges of node i , where $i \in N_v$. Then, the **bridging coefficient** of a node v is defined as $\pi_v = \frac{1}{\theta_{total}(v)} \sum_{i \in N_v, \theta_{total}(i) > 1} \frac{\eta_i}{\theta_{total}(i) - 1}$.

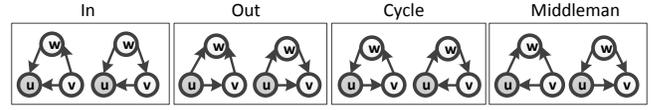


Figure 2: Directed triangle graph [21].

DEFINITION 6. Given the inverses of betweenness centrality rank and bridging coefficient rank of node v denoted as $\psi_{\frac{1}{\delta:v}}$ and $\psi_{\frac{1}{\pi:v}}$, respectively, the **bridging centrality** is defined as $\zeta_v = \psi_{\frac{1}{\delta:v}} \times \psi_{\frac{1}{\pi:v}}$.

As remarked in Section 1, Hwang et al. [28] reported that compared to hub nodes (nodes with high degree), bridging nodes are more effective drug targets with fewer off-target effects in several PPI networks. For instance, **PI3K** has high *bridging coefficient* and *bridging centrality* since it is positioned at the boundary of a strongly connected component (SCC) within the network and helps to transmit signal between nodes outside the SCC and those within it.

Clustering Coefficient. This feature determines how well the neighbourhood of a node is connected [84] by considering how close the neighbourhood is to being a clique where every node within the clique is connected to every other node in it [1]. The original definition was designed for undirected graph. A variety of definition exists [21] when edge directions are considered (Figure 2).

DEFINITION 7. Given a signaling network $G = (V, E)$, let $e_{ij} \in E$ denote an edge connecting nodes i to j where $i, j \in V$ and $A = \{a_{ij}\}$ be the adjacency matrix where $a_{ij} = 1$ if and only if $\exists e \in \{e_{ij}, e_{ji}\} \subseteq E$ and zero otherwise. Then, the **undirected-, in-, out-, cycle- and middleman-clustering coefficient** of a node $u \in V$ denoted as $\kappa_{undir(u)}$, $\kappa_{in(u)}$, $\kappa_{out(u)}$, $\kappa_{cyc(u)}$ and $\kappa_{mid(u)}$, respectively, are defined as

$$\begin{aligned} \kappa_{undir(u)} &= \frac{(A^3)_{ii}}{\theta_{total(u)}(\theta_{total(u)} - 1)}, \kappa_{in(u)} = \frac{(A^T A^2)_{ii}}{\theta_{in(u)}(\theta_{in(u)} - 1)}, \\ \kappa_{out(u)} &= \frac{(A^2 A^T)_{ii}}{\theta_{out(u)}(\theta_{out(u)} - 1)}, \kappa_{cyc(u)} = \frac{(A^3)_{ii}}{\theta_{in(u)}\theta_{out(u)} - A_{ii}^2}, \\ \kappa_{mid(u)} &= \frac{(A A^T A)_{ii}}{\theta_{in(u)}\theta_{out(u)} - A_{ii}^2}, \end{aligned}$$

where $\theta_{in(u)}$, $\theta_{out(u)}$ and $\theta_{total(u)}$ are the in, out and total degree of u , respectively; A^T is the transpose of A ; A^n is the matrix product of n copies of A ; and A_{ii} denotes the i^{th} element of the main diagonal of A .

Note that in the above definition, the neighbourhood size must be greater than one. For smaller neighbourhood sizes ($N_u = 0$ and $N_u = 1$), the coefficients are set to zero.

Target Downstream Effect (TDE). TDE assesses the potential impact on the network when a node is perturbed based on the probability of perturbing a downstream node² w and the likelihood of w causing off-target effect [13].

DEFINITION 8. Given a signaling network $G = (V, E)$, let W be the set of downstream nodes of $v \in V \setminus W$. Let $\rho_{v,w}$ be the probability of perturbing $w \in W$ when target node v is perturbed and $\theta_{total(w)}$ be the total degree of w . The **target downstream effect** of v is defined as $\omega_v = \sum_{w \in W} (\rho_{v,w} \times \theta_{total(w)})$.

²Node w is downstream of v if there exists a path from v to w .

3. REPRESENTATIVE SIGNALING NETWORKS AND TARGETS

We examine the targets of five signaling networks in this paper, namely, **Ras** activation [20], **MAPK-PI3K** [25], **glucose-stimulated insulin secretion** [32], **endomesoderm gene regulatory** [43] and **glucose metabolism** [42] networks. A reference standard for judging which are the “true” biological targets of these networks must be assembled. To this end, manual curation of literature generates substantially lower error rates than text mining-based approaches [75]. However, manual curation is tedious and time-intensive. Hence, we restrict the number of networks studied in this paper to five. In particular, we selected these networks because they have been well-studied with sufficient literature to facilitate the curation effort. We shall now briefly introduce these networks and describe the curation process. The networks are obtained from the *BioModels.Net* repository [45] and are summarized in Table 2. Note that details of the curation results are reported in [12].

Human Disease-Related Networks. Amongst the five networks we study, three are associated with human diseases (**MAPK-PI3K** [25], **Ras** activation [20], and the **glucose-stimulated insulin secretion** networks [32]). The curation process for these networks is as follows:

1. Obtain a list of unique drugs and compounds relevant to the human disease from clinical trial database [56].
2. Obtain the targets of these drugs and compounds via drug related databases [87] and literature survey.
3. Identify the targets that are in the scope of the signaling network.

The **Ras** activation network describes the **phospholipase C(PLC)- ϵ** -driven compartment switching of **Ras** activation at the plasma membrane and the Golgi. **Ras**-activation mutation is found to be present in various cancers, including ovarian cancer [68]. The **MAPK-PI3K** network, on the other hand, describes the **heregulin (HRG)**-induced **ErbB** receptor signaling network in Chinese hamster ovary (CHO) cells. In particular, it consists of two interacting pathways, the **MAPK** and **PI3K-Akt** signaling cascades. Similar to the **Ras** activation network, the **MAPK-PI3K** network is involved in human cancers due to its roles in cell survival signaling [48]. Specifically, in this paper, we associate both the **Ras** activation network and the **MAPK-PI3K** network with ovarian cancer.

Hence, we use the keywords “ovarian cancer drug” to search the clinical trial repository [56] which yielded 2054 studies. However, only 989 studies involve drugs specific for the purpose of ovarian cancer study while the rest involve questionnaires, radiation treatment, or general terms such as chemotherapy. The same drug can be used in multiple studies. Hence, duplicated drugs are removed from the 989 studies, resulting in a total of 458 unique drugs. Amongst these 458 drugs, some of the drugs target other networks. The targets of the 458 drugs were identified using literature survey. Amongst these, only those that are relevant to these two networks are considered as reference drug targets and are reported in Table 3 (denoted by I_0 and I_1).

The **glucose-stimulated insulin secretion** network describes **glucose-stimulated insulin secretion** in pancreatic β cells which consists of five subsystems, namely, glycolysis, the **tricarboxylic acid cycle (TCA)**, the respiratory chain, **nicotin-amide adenine dinucleotide (NADH)** shuttles and the **pyruvate** cycle. In particular, this network is associated

with τ_{2DM} . Hence, we use the keywords “type 2 diabetes mellitus drug” to search the clinical trial repository [56] which yielded 5858 studies. However, only 3880 studies involve the study of effects of drugs or food constituents on type 2 diabetes and these studies implicated a total of 617 unique compounds. Amongst these compounds, some may target other networks. The targets of the drugs and food constituents were identified using literature survey. Amongst these, only those that are relevant to the network are considered as reference drug targets as reported in Table 3 (denoted by I_2).

Biological Process-Related Networks. The remaining networks we studied describe specific biological processes of particular organisms. The curation process for these networks is as follows:

1. Obtain a list of unique molecules (genes or proteins) relevant to the biological process of the specific organism from *PubMed* using specific keywords.
2. Identify the molecules that are in the scope of the signaling network.

The **endomesoderm gene regulatory** network describes **endomesoderm gene regulation** in sea urchin (*Strongylocentrotus purpuratus*), a model organism for embryonic development. In particular, it describes the key steps in **endomesoderm** development, namely, the initiation of the **endomesoderm** specification signal, the maintenance of the specification signal, the activation of the **Delta/Notch** signaling pathway, and the specification of **veg1** endoderm. We performed manual curation (detailed in [14]) by searching the *PubMed* repository using “sea urchin endomesoderm” as keywords. A regulatory pathway pertaining to the regulation of **Endo16**, a critical protein in the formation of the **endomesoderm** is constructed from these publications. Nodes in this pathway are considered as targets for regulating **Endo16**. Details of this pathway³ can be found in [14]. Table 4 reports the curated targets of this network (denoted by I_3).

Lastly, the **glucose metabolism** network describes the metabolism of **glucose** to **acetate** in *Escherichia coli*. In particular, it describes the Embden-Meyerhoff pathway which focuses on glycolysis or gluconeogenesis (depending on the flux direction), the **tricarboxylic acid cycle** and the carbon flux flow through the **glyoxylate** shunt. We performed manual curation by searching the *PubMed* repository using “E Coli glucose metabolism to acetate” as keywords. The search yielded 545 publications, of which 21 were relevant literature on **glucose** metabolism in *Escherichia coli* under aerobic conditions. We define targets in this network as genes or proteins in the literature that caused 50% or more change in **acetate** production when modified (knockout, knockdown or overexpression). However, not all literature included **acetate** production as a measurement. Hence, we also consider genes or proteins with significant difference in transcription or translation level between *Escherichia coli* strains with distinct difference in **acetate** production as targets. These strains are **BL21** (low **acetate** producer) and **JM109** (high **acetate** producer). Note that both **acetate** and **glucose** are considered targets by default since they are the input and output, respectively, of the metabolic reaction that we are interested in. Table 5 reports the curated targets of this network (denoted by I_4).

³Since the curated model in [14] is current till 2011, we conducted a targeted search of the relevant literature from then till 28 October 2013. The search did not yield any new insights into the curated model.

	Ras activation	MAPK-PI3K	Glucose-Stimulated Insulin Secretion	Endomesoderm Gene Regulation	Glucose Metabolism
Network notation	I ₀	I ₁	I ₂	I ₃	I ₄
BioModel ID	BIOMD0000000161	BIOMD0000000146	BIOMD0000000239	BIOMD0000000235	BIOMD0000000244
Organism or cell type	Cells of pheochromocytoma in rat adrenal medulla (pc12 cell line) and mouse embryonic fibroblast cells (NIH 3T3 cell line)	Chinese hamster ovary cells	Mouse pancreatic β cells	Sea urchin embryo	<i>Escherichia coli</i>
Related disease or biological phenomenon	Cancer (general) - Ovarian for this study	Ovarian cancer	Type 2 diabetes mellitus	Embryonic development	Glucose to acetate metabolism
Output node [†]	RasGTP on plasma membrane (RasGTP_PM)	Double phosphorylated ERK (ERKPP)	Mitochondria ATP (ATP)	Endo16 protein (Protein_E_Endo16)	Acetate (ACT)
No. of nodes	46	36	59	622	47
No. of hyperedges	43	34	45	778	109
No. of targets	5	9	6	206	16
Repository used for curation	ClinicalTrials.gov	ClinicalTrials.gov	ClinicalTrials.gov	PubMed	PubMed
Keywords used for curation	ovarian cancer drug	ovarian cancer drug	type 2 diabetes mellitus drug	sea urchin endomesoderm	E Coli glucose metabolism to acetate
Date of Curation	29 Apr 2014	29 Apr 2014	25 Jan 2013	28 Oct 2013	14 Nov 2013
Unique Drugs Curated	458	458	617	-	-
Relevant Drugs Curated	12	22	16	-	-

Table 2: Summary of the networks and curation results. [†]The output node in round bracket is the node name used in the model.

Network	Targets	Drugs or Food Constituents
I ₀	Ca ²⁺	Carboxyamidotriazole [54], Hydralazine [72]
	EGFR	EGFR antisense DNA [54]
	activated EGFR	Lapatinib [65], Gefitinib [54], Iressa [54], Vandetanib [83]
	EGF:EGFR	Matuzumab [54], Erlotinib [54], Panitumumab [89]
	dimerized EGFR	Cetuximab [5]
	cells expressing mutant Ras	Ras peptide cancer vaccine [54]
I ₁	RP	Lapatinib [65]
	Raf*	Sorafenib [85], Dabrafenib [54]
	Raf	ISIS 5132 [15], ECO-4601 [54]
	MEKPP	AZD6244 [91], gsk1120212 [30], MEK162 [54], Pimasertib [54], Trametinib [54]
	PI3K*	XL147 [54], PKI-587 [81], PKI-179 [54], BKM120 [54], BYL719 [54], SAR245409 [54]
	AktPIP	Perifosine [41]
	AktPIPP	Perifosine [41]
	AktPIP3	Perifosine [41]
	Akt	AZD5363 [54], gsk2110183 [74], gsk2141795 [3], MK-2206 [90], Triciribine [54]
	I ₂	glucose
acetyl-CoA (mitochondrial)		Benfluorex [40]
activated FBPase		CS-917 [18], MB07803 [80]
glycerol-3-phosphate		Glycerol [46]
acetyl-CoA (cytoplasm)		Methylcobalamin [23]
ferrocytochrome c		Gynostemma Pentaphyllum tea [7], Xanthohumol [7]

Table 3: Curated targets of networks I₀ to I₂.

4. DISCRIMINATIVE TOPOLOGICAL FEATURES DISCOVERY

In this section, we first formally define the notion of *discriminative topological feature* (DTF). Then, we describe the

I ₃ Targets	Reference	I ₃ Targets	Reference
Pmar1	[10]	TCF	[16]
HesC	[10]	Gro	[16], [50]
Ets1	[69]	N β :TCF	[16], [6]
Delta	[50], [69], [10]	Blimp1	[6], [16], [47]
Notch	[50]	Wnt8	[16]
SuH	[50]	Otx	[47]
GataE	[50]	Eve	[73]
Endo16 [†]	[47]	Bra	[60]
Brn1/2/4	[93]	Dri	[2]
SoxB1	[2]	Hox11/13b	[60]
cB	[16]		

Table 4: Curated targets of I₃. Targets marked by [†] are included by default due to their obvious role in regulating the output node.

I ₄ Targets	Reference	I ₄ Targets	Reference
ACT [†]	-	Fdp	[62]
GLC [†]	-	Icd	[34]
g6P	[94]	Icd_P	[34]
ICT	[79]	Pdh	[51]
PEP	[22]	Ppc	[22]
AceB	[61]	PpsA	[57]
Acoa2act	[11]	EIIA	[31]
Cya	[59]	EIICB	[63]

Table 5: Curated targets of I₄. Targets marked by [†] has same semantics as in Table 4.

algorithm DIFFER for finding and ranking the set of DTFs for a given signaling network with a set of known targets. Note that these targets may be discovered by undertaking the curation process discussed in the preceding section.

4.1 Discriminative Topological Feature (DTF)

Given a signaling network $G = (V, E)$ and a set of known targets $T \subseteq V$, a topological feature is *discriminative* if its distributions for T and for $V \setminus T$ are “significantly” different. The similarity between the two distributions can be measured and used to determine the extent to which the distributions are different.

DEFINITION 9. Given a signaling network $G = (V, E)$ with a known set of targets $T \subseteq V$, a set of features \mathcal{X} , and a significance threshold t , let the similarity for the distribution of the i^{th} feature in \mathcal{X} between T and $V \setminus T$ be denoted as $D(\mathcal{X}_{i[T]}, \mathcal{X}_{i[V \setminus T]})$. Then, a feature $\mathcal{X}_i \in \mathcal{X}$ is **discriminative** if $D(\mathcal{X}_{i[T]}, \mathcal{X}_{i[V \setminus T]}) < t$.

In this paper, we use the p -value of the 2-tailed Wilcoxon test to compute $D(\mathcal{X}_{i[T]}, \mathcal{X}_{i[V \setminus T]})$ ⁴ and set $t = 0.05$. Specifically, in the 2-tailed Wilcoxon test for two given distributions A and B , the null hypothesis $H_0 : A = B$ is tested against the alternative hypothesis $H_1 : A \neq B$. Let n_A be the number of observations in A , w_A be the sum of the ranks for observations from A and W_A be the corresponding random variable. The p -value is given as $p = Pr(W_A \neq w_A)$ and indicates the probability of seeing a value not equal to the observed w_A .

4.2 The Algorithm DIFFER

Algorithm 1 outlines DIFFER which consists of three phases: *feature value computation*, *Wilcoxon test analysis* and *discriminative feature ranking*. DIFFER first initializes the significance threshold value t and the counters x and y (Line 1). Note that it sets t to a default value of 0.05 if it is not provided in the input. Then, in the *feature value computation* phase, the values of all features in \mathcal{X} is computed for each node in the given signaling network. The computed feature values for the target nodes and non-target nodes are stored in the $|\mathcal{X}| \times |T|$ matrix X (Line 5) and $|\mathcal{X}| \times |V \setminus T|$ matrix Y (Line 8), respectively. Next, in the *Wilcoxon test analysis* phase, the 2-tailed Wilcoxon test is performed for each feature in \mathcal{X} to determine the difference in the feature distributions between target and non-target nodes (Line 14). Finally, in the *discriminative feature ranking* phase, the p -values p_i obtained from the Wilcoxon tests are compared against the significance threshold t (Line 16). Those features whose p -values fall above t are removed and the remaining features \mathcal{X}_{dis} are ranked in increasing p -values. We deliberately omitted any correction for multiple comparisons because our purpose is to err on the side of including DTFs that may correlate with the targets. Note that our later analysis does not assume any specific level of statistical significance for the ranked DTFs.

THEOREM 1. *The worst case time complexity of DIFFER is $O(|\mathcal{X}||V|O(\mathcal{G}(\mathcal{X})) + |\mathcal{X}| \times (|T||V \setminus T|)^2 + |V|\log(|V|))$ where $O(\mathcal{G}(\mathcal{X}))$ is the worst case time complexity for computing the given feature set \mathcal{X} .*

PROOF. For the feature value computation phase, a worst case time complexity of $O(|\mathcal{X}||V|O(\mathcal{G}(\mathcal{X})))$ is required where $O(\mathcal{G}(\mathcal{X}))$ is the worst case time complexity for computing the features. Then, in the Wilcoxon test analysis phase, $O(|\mathcal{X}| \times (|T||V \setminus T|)^2)$ time is required since each Wilcoxon test requires $O((|T||V \setminus T|)^2)$ time [53]. Finally, in the discriminative feature ranking phase requires $O(|V|\log(|V|))$ in the worst case when all features are discriminative. Hence, DIFFER requires $O(|\mathcal{X}||V|O(\mathcal{G}(\mathcal{X})) + |\mathcal{X}| \times (|T||V \setminus T|)^2 + |V|\log(|V|))$ in the worst case. This can be reduced further if we assume bounds on $O(\mathcal{G}(\mathcal{X}))$. \square

Table 6 reports the time complexities for computing various structural features considered in this paper. Observe

⁴Note that other similarity measures such as *Hellinger distance* can also be used. However, it is orthogonal to the problem addressed in this paper.

Algorithm 1 Algorithm DIFFER

Input: Signaling network $G = (V, E)$, set of known targets T , set of topological features \mathcal{X} , significance threshold t (optional)
Output: set of ranked discriminative topological features \mathcal{X}_{dis}

```

1:  $t, x, y \leftarrow \text{INIT}(t)$ 
2: for iteration  $i=1$  to  $|\mathcal{X}|$  do
3:   for iteration  $j=1$  to  $|V|$  do
4:     if  $\text{ISTARGET}(V_j)=\text{true}$  then
5:        $X_{(i,x)} \leftarrow \text{COMPUTEFEATUREVALUES}(\mathcal{X}_i, G, V_j)$ 
6:        $x \leftarrow \text{INCREMENTCOUNTER}(x)$ 
7:     else
8:        $Y_{(i,y)} \leftarrow \text{COMPUTEFEATUREVALUES}(\mathcal{X}_i, G, V_j)$ 
9:        $y \leftarrow \text{INCREMENTCOUNTER}(y)$ 
10:    end if
11:  end for
12: end for
13: for iteration  $i=1$  to  $|\mathcal{X}|$  do
14:    $p_i \leftarrow \text{WILCOXON}(X_i, Y_i)$ 
15: end for
16:  $\mathcal{X}_{dis} \leftarrow \text{RANKDISCRIMINATIVEFEATURES}(\mathcal{X}, p, t)$ 

```

Structural Features	Time Complexity
Degree centrality	$O(V)$ [12]
Eigenvector centrality	$O(V ^2)$ [37]
Closeness centrality	$O(V ^3)$ [36]
Eccentricity centrality	$O(V E)$ [77]
Betweenness centrality	$O(V E)$ [9]
Bridging centrality	$O(V ^2 + E)$ [12]
Bridging coefficient	$O(V ^2)$ [12]
Clustering coefficient	$O(V ^{2.373})$ [86]
Proximity prestige	$O(V ^2 + E)$ [12]
Target downstream effect	$O(V ^2 + E)$ [12]

Table 6: Time complexities of structural features.

that closeness centrality computation has the maximum worst case complexity ($O(|V|^3)$ [36]) amongst all features.

THEOREM 2. *The worst case space complexity of DIFFER is $O(|E| + |\mathcal{X}||V|)$.*

PROOF. First, the algorithm requires $O(|V| + |E| + |\mathcal{X}|)$ space for storing the inputs to the algorithm. Then, in the feature value computation phase, a worst case space complexity of $O(|\mathcal{X}||V|)$ is required for storing all the feature values. Next, in the Wilcoxon test analysis phase, $O(|\mathcal{X}|)$ space is needed to store the p -values obtained. Finally, in the discriminative feature ranking phase, $O(|\mathcal{X}|)$ space is again needed to store the ranks of all features in the worst case when all features are considered discriminative. Hence, DIFFER requires $O(|E| + |\mathcal{X}||V|)$ space in the worst case. \square

Remark. Observe that the performance of DIFFER is affected by the target curation quality. The manual curation is limited by coverage of the reference repository (*e.g.*, *PubMed*). Literature evidence which are not found in the repository will not be curated and potential targets may be missed. These missing targets may impact DIFFER since the DTFs are discovered by comparing the feature distribution of the targets against the non-targets. The missing targets would be likely to affect the results if they have different feature distributions from the known targets. This impact could be significant when the set of missing targets is large relative to the size of the set of known targets. For instance, a feature that is deemed discriminative based on the known target size may be considered non-discriminative when the missing targets are considered.

5. EMPIRICAL STUDY

DIFFER is implemented in Java and the Wilcoxon test is performed using *R*. The experiments are performed on

a computer system using a 64-bit operating system with 8GB RAM and a dual core processor running at 3.60GHz. In this section, we investigated the performance of DIFFER using the signaling networks and curated targets described in Section 3. Unless specified otherwise, we set $t = 0.05$.

5.1 DTFs of the Networks

In this set of experiments, we applied DIFFER to the five signaling networks to obtain sets of ranked DTFs. Table 7 reports the ranks of the DTFs for these networks. Note that for I_4 , the default significance threshold $t = 0.05$ did not return any DTF. Hence for this network, we relaxed t to 0.2 for the purpose of generating DTFs for our empirical study. We can make the following observations. First, the significance threshold directly affects the selection of DTFs and the size of \mathcal{X}_{dis} but not their ranking. For instance, for I_4 , $|\mathcal{X}_{dis}|$ is 0 and 3 when t is 0.05 and 0.2, respectively. The choice of the value of t is application-specific. Second, the number of DTFs differs across networks. In some networks such as I_3 , the targets are characterized by as many as nine features whereas in another network I_0 , only four features are discriminative enough. Note that it is also possible for a network to have no DTF (*e.g.*, I_4 for $t = 0.05$). This implies that the feature set \mathcal{X} may not be appropriate for characterizing the targets in this network and additional features need to be considered. Third, different DTFs perform differently on different networks. For instance, bridging centrality, a feature found to yield good targets in a variety of PPI networks [28] performed well on I_1 and I_3 , but poorly on I_2 and I_4 . Hence, a single DTF cannot be used for distinguishing targets from non-targets in all signaling networks.

5.2 DTFs Finds Biologically Relevant Targets

In this set of experiments, we examine the discriminative power of DTFs in terms of identifying known targets in the upper 50-percentile of nodes ranked according to the DTFs. Note that a low p -value in the 2-tailed Wilcoxon test indicates a difference in the feature distribution between the targets T and non-targets $V \setminus T$. This difference could be due to the targets either generally having higher or lower values than the non-targets. Hence, ranking based on a DTF may result in majority of the known targets either in the upper 50-percentile rank (high-to-low DTF) or the lower 50-percentile rank (low-to-high DTF). For instance, in I_0 , ranking based on κ_{undir} , a low-to-high DTF, resulted in 4 out of the 5 known targets in the lower 50-percentile. For simplicity, we redefine the low-to-high DTFs (marked with #) using their inverse values (*i.e.*, $\frac{1}{z}$ rather than z) so that majority of known targets are found in the upper 50-percentile.

Figure 3 plots the performance of the DTFs. Observe that 71% of the features are able to identify at least 70% of targets in the upper 50-percentile rank of the nodes. In particular, all the DTFs in I_2 identified at least 83.3% of the targets. In I_0 and I_1 , we note that a DTF that is ranked higher than another is also able to identify a higher percentage of targets. For instance, β , the top-ranked feature in I_1 , identified 90% of the targets compared to μ (ranked 6th), which identified only 60% of the targets.

We further examine non-target nodes that are present in the upper 50-percentile ranks of the top-3 DTFs in various networks. The goal is to find out how relevant these nodes are as targets for the networks. Note that we exclude I_4 from this particular study as its DTFs have p -values greater than 0.05. Several of these non-target nodes are found to

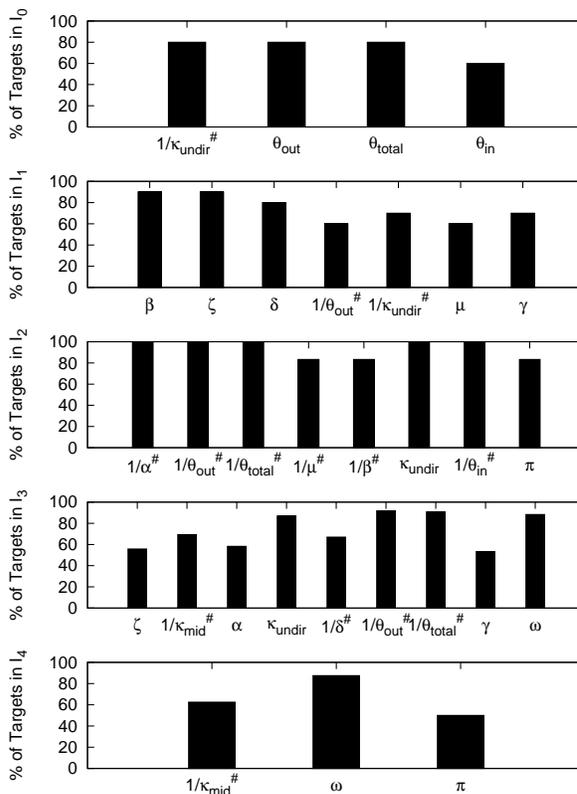


Figure 3: Percentage of targets in upper 50-percentile of nodes when ranked using DTFs.

be promising therapeutic targets. Due to space constraints, we only highlight some of them here. In I_0 , **diacylglycerol (DAG)**, a non-target node, is implicated in cancer cell migration and facilitates signaling in cancer cells [4]. Moreover, a recent study reveals that deletion of **DAG kinase**, an enzyme that metabolizes DAG, improves the clinical potential of chimeric antigen receptor (CAR)-transduced T cells in the treatment of cancer [66]. In a review of novel treatments of type II diabetes mellitus, several enzymes (*e.g.*, **phosphoenolpyruvate (PEP) carboxykinase**) and mitochondrial complex I (**rotenone-sensitive NADH:ubiquinone oxidoreductase**) have been proposed for treating diabetes [82]. In particular, **PEP carboxykinase** regulates the production of PEP, a non-target in I_2 . **NADH**, another non-target in I_2 is a component of mitochondrial complex I. In summary, DTFs demonstrate promising results toward identifying biologically-relevant targets.

5.3 Correlation of DTFs

Next, we examine the correlation of DTFs to gain more insights on their characteristics. Table 8 reports DTFs having correlation greater than 0.68⁵. We make the following observations. First, DTFs are more likely to be positively-correlated than negatively-correlated. Second, the extent of correlation amongst DTFs differ across networks. For instance, all the DTFs in I_0 have strong correlation whereas none of the DTFs in I_4 are correlated. Third, degree centralities tend to be strongly correlated to others. For instance, θ_{total} is strongly correlated with κ_{undir} in I_0 and I_2 .

⁵Correlation in the range of [0.68–1] is considered strong [78].

Features	I ₀		I ₁		I ₂		I ₃		I ₄	
	p-value	Rank	p-value	Rank	p-value	Rank	p-value	Rank	p-value	Rank
Undirected clustering coefficient κ_{undir}	0.02	1	0.014	5	0.016	5	3.69×10^{-7}	4	0.211	-
Out degree θ_{out}	0.024	2	0.013	4	0.006	2	1.11×10^{-4}	6	0.652	-
Total degree θ_{total}	0.039	3	0.066	-	0.009	3	7.28×10^{-4}	7	0.848	-
In degree θ_{in}	0.043	4	0.59	-	0.035	6	0.124	-	0.557	-
Middleman clustering coefficient κ_{mid}	0.763	-	0.89	-	0.051	-	5.60×10^{-8}	2	0.117	1
Target downstream effect ω	0.511	-	0.085	-	0.107	-	0.033	9	0.142	2
Bridging coefficient π	0.801	-	0.19	-	0.037	7	0.177	-	0.155	3
Bridging centrality ζ	0.077	-	1.77×10^{-3}	2	0.347	-	6.67×10^{-11}	1	0.23	-
Proximity prestige μ	0.48	-	0.022	6	0.015	4	0.902	-	0.349	-
Eigenvector centrality α	0.198	-	0.116	-	0.003	1	3.51×10^{-7}	3	0.433	-
Betweenness centrality δ	0.186	-	1.92×10^{-3}	3	0.139	-	3.54×10^{-5}	5	0.475	-
Closeness centrality β	0.902	-	3.52×10^{-4}	1	0.015	4	0.118	-	0.755	-
Cycle clustering coefficient κ_{cyc}	0.763	-	0.828	-	0.051	-	0.472	-	0.793	-
Eccentricity centrality γ	0.698	-	0.03	7	0.176	-	0.012	8	0.798	-
In clustering coefficient κ_{in}	0.763	-	0.594	-	0.259	-	0.115	-	0.923	-
Out clustering coefficient κ_{out}	0.763	-	0.94	-	0.491	-	0.637	-	0.978	-

Table 7: Rank of DTFs.

5.4 Effect of Varying Target Set Size

Recall that the set of curated targets may miss some known targets which inevitably affects the size of the known target set. In this set of experiments, we study the influence of the size of targets and non-target sets on DTF identification on I₃, the largest network in this study. First, we examine the distribution of the top-3 DTFs. As observed in Figures 4(a)-(c), the QQ-plots⁶ indicate that the distributions of targets and non-targets are different. Next, we simulate the scenario of missing targets in the curated set by randomly adding nodes from the non-target set to the target set. Figures 4(d)-(l) depict that varying the number of non-target nodes do affect the distribution. When we perform the Wilcoxon test for these new target sets, we observe that modifying the target set did not significantly affect DTF identification since seven⁷ out of the nine cases considered in Figure 4 have *p*-values lesser than 0.05. Hence, DIFFER is relatively robust against changes in target set size.

5.5 Role of Bridging Nodes

The experiments in Section 5.1 reveal that bridging centrality does not necessarily appear among the top-ranked DTFs. Note that this finding is in contrast to recent studies where bridging centrality was significant in characterizing nodes as drug targets [28]. In our final set of experiments, we ask whether bridging nodes (nodes with high bridging centrality) lie in the vicinity of the targets or have a role in regulating these targets. Hence, we studied two networks, one in which bridging centrality performed well (I₁), and the other in which it performed poorly (I₄). We identified the top-5 bridging nodes, and for these bridging nodes that are not targets, we examined their neighbourhood to see how far they are from known targets. In both networks, we found that bridging nodes are located near target nodes (within 2 hops). Note that a larger sample size would be required for concluding whether bridging nodes are significantly closer to target nodes than would be expected by chance, but our preliminary work raises the possibility that bridging centrality may have indirect predictive value in networks where it does not predict targets directly. Biologically, it is possible

⁶ In a QQ-plot (e.g., Figure 4), two distributions are identical if the plotted values (blue '+') all fall on the line $y=x$ (red line).

⁷ For bridging centrality and eigenvector centrality, the *p*-values were 0.695 and 0.698, respectively when 50 non-target nodes were added.

Network	Positively-Correlated DTFs	Negatively-Correlated DTFs
I ₀	$(\theta_{out}, \theta_{total}), (\theta_{out}, \theta_{in}), (\theta_{total}, \theta_{in})$	$(\kappa_{undir}, \theta_{out}), (\kappa_{undir}, \theta_{total}), (\kappa_{undir}, \theta_{in})$
I ₁	$(\beta, \delta), (\beta, \gamma)$	$(\zeta, \delta), (\delta, \kappa_{undir})$
I ₂	$(\alpha, \theta_{out}), (\alpha, \mu), (\theta_{out}, \theta_{total}), (\theta_{total}, \mu), (\theta_{total}, \theta_{in}), (\mu, \theta_{in}), (\beta, \theta_{in})$	$(\alpha, \kappa_{undir}), (\theta_{out}, \kappa_{undir}), (\theta_{total}, \kappa_{undir}), (\mu, \kappa_{undir}), (\beta, \kappa_{undir}), (\kappa_{undir}, \theta_{in})$
I ₃	$(\zeta, \delta), (\alpha, \theta_{total}), (\kappa_{undir}, \omega), (\theta_{out}, \theta_{total}), (\theta_{total}, \gamma)$	$(\kappa_{mid}, \gamma), (\kappa_{undir}, \delta), (\delta, \omega), (\theta_{out}, \gamma)$
I ₄	-	-

Table 8: DTFs with strong correlation.

that target proteins may be disproportionately regulated by indirect effects of proteins that function as bridging nodes.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we present an algorithm called DIFFER that analyzes the topology of a signaling network to identify a set of discriminative topological features (DTFs) for distinguishing target from non-target nodes. We investigated 16 topological features over five signaling networks whose targets were determined by extensive manual curation. DIFFER computed the DTF results, and we observed that there was no single DTF that could characterize targets for all signaling networks. This finding contrasts with the results in [28] where bridging centrality was leveraged to identify drug targets in multiple PPI networks. It also contrasts with the findings in [71] where low clustering coefficient feature is used as a predictor of disease-related genes in diabetes-related gene-gene interaction network. Based on the results of this paper, we believe that characterization of targets in signaling networks is highly dependent on network topology and demands analysis of multiple structural features. The results of this paper are an important first step in this regard. As part of future work, we aim to investigate the effect of using other metrics (e.g., Kolmogorov-Smirnov statistics) to identify DTFs and to examine how DTFs can be leveraged to provide superior solution to the *target prioritization problem* in signaling networks.

Acknowledgments. This work was supported in part by a Singapore MOE AcRF Tier 1 Grant RGC 1/13.

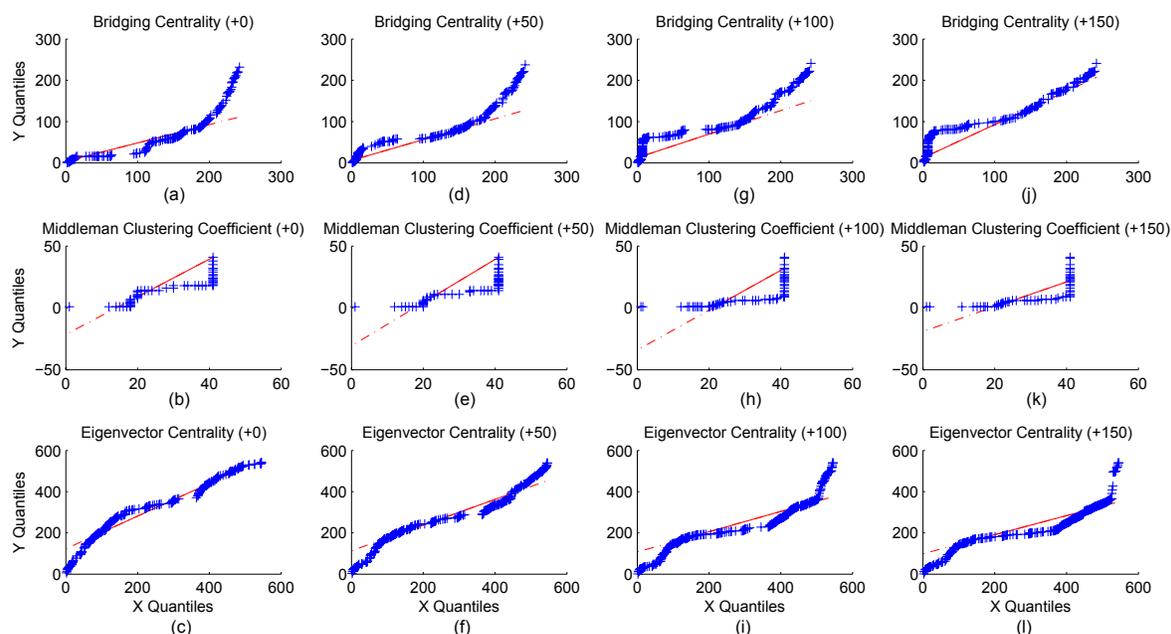


Figure 4: [Best viewed in color] Quantile-quantile (QQ) plot of the target set (x-axis) versus the non-target set (y-axis) for the top-3 DTFs in I_3 . Note that (+x) labels indicates that x non-target nodes have been randomly transferred to the target set.

7. REFERENCES

- [1] R. Albert. Scale-free networks in cell biology. *Journal of Cell Science*, 2005.
- [2] G. Amore et al. Spdeadringer, a sea urchin embryo gene required separately in skeletogenic and oral ectoderm gene regulatory networks. *Developmental Biology*, 2003.
- [3] C. Arteaga. The phosphatidylinositol-3 kinase/mTOR pathway: new agents. *Breast Cancer Research*, 2011.
- [4] F. Baenke et al. Hooked on fat: the role of lipid synthesis in cancer metabolism and tumour development. *Disease models & mechanisms*, 2013.
- [5] J. Bangham. Therapeutics: Cetuximab constricts conformational contortionist. *Nature Reviews Cancer*, 2005.
- [6] S. Ben-Tabou de-Leon et al. Deciphering the Underlying Mechanism of Specification and Differentiation: The Sea Urchin Gene Regulatory Network *Sci. STKE*, 2006.
- [7] A. Bhuyan et al. Resting membrane potential as a marker of apoptosis: studies on *Xenopus oocytes* microinjected with cytochrome *c*. *Cell death and differentiation*, 2001.
- [8] P. Bonacich et al. Eigenvector-like measures of centrality for asymmetric relations. *Social Networks*, 2001.
- [9] U. Brandes. A Faster Algorithm for Betweenness Centrality. *Journal of Mathematical Sociology*, 2001.
- [10] C. Byrum et al. Blocking Dishevelled signaling in the noncanonical Wnt pathway in sea urchins disrupts endoderm formation and spiculogenesis, but not secondary mesoderm formation. *Developmental Dynamics*, 2009.
- [11] D.-E. Chang et al. Acetate Metabolism in a PTA Mutant of *Escherichia coli* W3110: Importance of Maintaining Acetyl Coenzyme A Flux for Growth and Survival. *Journal of Bacteriology*, 1999.
- [12] H. Chua et al. One Feature Doesn't Fit All: Characterizing Topological Features of Targets in Signaling Networks. www.cais.ntu.edu.sg/~assourav/TechReports/DIFFER-TR.pdf, 2014.
- [13] H. Chua et al. PANI: A Novel Algorithm for Fast Discovery of Putative Target Nodes in Signaling Networks. *BCB*, 2011.
- [14] H. Chua et al. In Silico Identification of Endo16 Regulators in the Sea Urchin Endomesoderm Gene Regulatory Network. *ACM SIGHIT*, 2012.
- [15] C. Cunningham et al. A phase I trial of c-Raf kinase antisense oligonucleotide ISIS 5132 administered as a continuous intravenous infusion in patients with advanced cancer. *Clin Cancer Res*, 2000.
- [16] E. Davidson et al. A Genomic Regulatory Network for Development. *Science*, 2002.
- [17] N. Emanuele et al. Consequences of alcohol use in diabetics. *Alcohol Health and Research World*, 1998.
- [18] M. Erion et al. MB06322 (CS-917): a potent and selective inhibitor of fructose 1, 6-bisphosphatase for controlling gluconeogenesis in type 2 diabetes. *PNAS*, 2005.
- [19] I. Espinosa et al. Tagatose: from a sweetener to a new diabetic medication? *Expert opinion on investigational drugs*, 2010.
- [20] N. Eungdamrong et al. Compartment-Specific Feedback Loop and Regulated Trafficking Can Result in Sustained Activation of Ras at the Golgi. *Biophysical Journal*, 2007.
- [21] G. Fagiolo. Clustering in complex directed networks. *Phys. Rev. E*, 2007.
- [22] W. Farmer et al. Reduction of aerobic acetate production by *Escherichia coli*. *Appl. and Environmental microbiology*, 1997.
- [23] J. Fontecilla-Camps et al. Structure-function relationships of anaerobic gas-processing metalloenzymes. *Nature*, 2009.
- [24] L. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1979.
- [25] M. Hatakeyama et al. A computational model on the modulation of mitogen-activated protein kinase (MAPK) and Akt pathways in heregulin-induced ErbB signalling. *Biochem J*, 2003..
- [26] A. Hopkins. Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol*, 2008.
- [27] E. Hu et al. White rice consumption and risk of type 2 diabetes: meta-analysis and systematic review. *BMJ: British Medical Journal*, 2012.
- [28] W.-C. Hwang et al. Identification of information flow-modulating drug targets: a novel bridging paradigm for drug discovery. *Clin Pharmacol Ther*, 2008.
- [29] RxList Inc. RxList – The Internet Drug Index. <http://www.rxlist.com>. Accessed 3 Sept 2013.
- [30] J. Infante et al. Safety and efficacy results from the first-in-human study of the oral MEK 1/2 inhibitor GSK1120212. *J Clin Oncol*, 2010.
- [31] S. Jahn et al. A role for EIIANtr in controlling fluxes in the central metabolism of *E. coli* K12. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 2013.
- [32] N. Jiang et al. A kinetic core model of the glucose-stimulated insulin secretion network of pancreatic β cells. *Mammalian Genome*, 2007.
- [33] C. Johnston et al. Vinegar: medicinal uses and antiglycemic effect. *Medscape General Medicine*, 2006.

- [34] M. Kabir et al. Metabolic regulation analysis of icd-gene knockout Escherichia coli based on 2D electrophoresis with MALDI-TOF mass spectrometry and enzyme activity measurements. *Applied microbiology and biotechnology*, 2004.
- [35] R. Kaminski et al. Minimal intervention strategies in logical signaling networks with ASP. *Theory and Practice of Logic Programming*, 2013.
- [36] U. Kang et al. Centralities in Large Networks: Algorithms and Observations. *SIAM*, 2011.
- [37] C. Kiss et al. Identification of influencers measuring influence in customer networks. *Decision Support Sys.*, 2008.
- [38] S. Klamt et al. Hypergraphs and cellular networks. *PLoS Comput Biol*, 2009.
- [39] A. Klip et al. Cellular mechanism of action of metformin. *Diabetes Care*, 1990.
- [40] C. Kohl et al. Effects of Benfluorex on Fatty Acid and Glucose Metabolism in Isolated Rat Hepatocytes: From Metabolic Fluxes to Gene Expression. *Diabetes*, 2002.
- [41] S. Kondapaka et al. Perifosine, a novel alkylphospholipid, inhibits protein kinase B activation. *Molecular Cancer Therapeutics*, 2003.
- [42] O. Kotte et al. Bacterial adaptation through distributed sensing of metabolic fluxes. *Molecular systems biology*, 2010.
- [43] C. Kuhn et al. Monte Carlo analysis of an ODE Model of the Sea Urchin Endomesoderm Network. *BMC Sys. Biology*, 2009.
- [44] C. Lavigne et al. Prevention of skeletal muscle insulin resistance by dietary cod protein in high fat-fed rats. *American Journal of Physiology-Endocrinology And Metabolism*, 2001.
- [45] N. Le Novère et al. BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucl. Acids Res.*, 2006.
- [46] T. Li et al. Performance of batch membrane reactor: Glycerol-3-phosphate synthesis coupled with adenosine triphosphate regeneration. *Biotech. and Bioengineering*, 2001.
- [47] C. Livi et al. Expression and function of blimp1/krox, an alternatively transcribed regulatory gene of the sea urchin endomesoderm network. *Developmental Biology*, 2006.
- [48] J. Luo et al. Targeting the PI3K-Akt pathway in human cancer: Rationale and promise. *Cancer Cell*, 2003.
- [49] F. Machicao et al. Pleiotropic neuroprotective and metabolic effects of Actovegin's mode of action. *Journal of the neurological sciences*, 2012.
- [50] S. Materna et al. Logic of gene regulatory networks. *Curr Opin Biotechnol*, 2007.
- [51] A. Murarka et al. Metabolic Analysis of Wild-type Escherichia coli and a Pyruvate Dehydrogenase Complex (PDHC)-deficient Derivative Reveals the Role of PDHC in the Fermentative Metabolism of Glucose. *Journal of Biological Chemistry*, 2010.
- [52] J. Nacher et al. A global view of drug-therapy interactions. *BMC pharmacology*, 2008.
- [53] N. Nagarajan et al. Reliability and efficiency of algorithms for computing the significance of the Mann-Whitney test. *Computational Statistics*, 2009.
- [54] National Cancer Institute. NCI Drug Dictionary. <http://www.cancer.gov/drugdictionary/>. Accessed Sept 2014.
- [55] M. Newman. The mathematics of networks, 2006.
- [56] NIH. ClinicalTrials.gov. <http://www.clinicaltrials.gov>. Accessed 5 May 2014.
- [57] R. Patnaik et al. Stimulation of glucose catabolism in Escherichia coli by a potential futile cycle. *Journal of bacteriology*, 1992.
- [58] G. Pavlopoulos et al. Using graph theory to analyze biological networks. *BioData Mining*, 2011.
- [59] A. Perrenoud et al. Impact of global transcriptional regulation by ArcA, ArcB, Cra, Crp, Cya, Fnr, and Mlc on glucose catabolism in Escherichia coli. *Journal of bacteriology*, 2005.
- [60] I. Peter et al. The endoderm gene regulatory network in sea urchin embryos up to mid-blastula stage. *Dev. Biology*, 2010.
- [61] J.-N. Phue et al. Transcription levels of key metabolic genes are the cause for different glucose utilization pathways in E. coli B (BL21) and E. coli K (JM109). *Journal of Biotechnology*, 2004.
- [62] J.-N. Phue et al. Glucose metabolism at high density growth of E. coli B and E. coli K: Differences in metabolic pathways are responsible for efficient glucose utilization in E. coli B as determined by microarrays and Northern blot analyses. *Biotechnology and Bioengineering*, 2005.
- [63] A. Picon et al. Reducing the glucose uptake rate in Escherichia coli affects growth rate but not protein production. *Biotechnology and Bioengineering*, 2005.
- [64] E. Pieroni et al. Protein networking: insights into global functional organization of proteomes. *Proteomics*, 8, 2008.
- [65] C. Qiu et al. Mechanism of activation and inhibition of the HER4/ErbB4 kinase. *Structure*, 2008.
- [66] M. Riese et al. Enhanced effector responses in activated CD8+ T cells deficient in diacylglycerol kinases. *Cancer Res.*, 2013.
- [67] A. Ritz et al. Signaling hypergraphs. *Trends in biotech.*, 2014.
- [68] P. Roberts et al. Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. *Oncogene*, 2007.
- [69] E. Röttinger et al. A Raf/MEK/ERK signaling pathway is required for development of the sea urchin embryo micromere lineage through phosphorylation of the transcription factor Ets. *Development*, 2004.
- [70] G. Scardoni et al. New Frontiers in Graph Theory, 2012.
- [71] A. Sharma et al. Gene prioritization in type 2 diabetes using domain interactions and network analysis. *BMC Genomics*, 11, 84, 2010.
- [72] N. Singh et al. Molecular modeling and molecular dynamics studies of hydralazine with human DNA methyltransferase 1. *ChemMedChem*, 2009.
- [73] J. Smith et al. A spatially dynamic cohort of regulatory genes in the endomesodermal gene network of the sea urchin embryo. *Developmental Biology*, 2008.
- [74] A. Spencer et al. Novel AKT Inhibitor Afuresertib In Combination With Bortezomib and Dexamethasone Demonstrates Favorable Safety Profile and Significant Clinical Activity In Patients With Relapsed/Refractory Multiple Myeloma. *Blood*, 2013.
- [75] C. Stark et al. The BioGRID Interaction Database: 2011 update. *Nucleic Acids Research*, 2011.
- [76] R. Steriti. Berberine for Diabetes Mellitus Type 2. *Natural Medicine Journal*, 2010.
- [77] F. Takes et al. Computing the eccentricity distribution of large graphs. *Algorithms*, 2013.
- [78] R. Taylor. Interpretation of the correlation coefficient: a basic review. *Journal of diagnostic medical sonography*, 1990.
- [79] M. van de Walle et al. Proposed mechanism of acetate accumulation in two recombinant Escherichia coli strains during high density fermentation. *Biotechnology and bioengineering*, 1998.
- [80] P. van Poelje et al. Fructose-1, 6-bisphosphatase inhibitors for reducing excessive endogenous glucose production in type 2 diabetes. *Diabetes-Perspectives in Drug Therapy*, 2011.
- [81] A. Venkatesan et al. Bis(morpholino-1,3,5-triazine) Derivatives: Potent Adenosine 5'-Triphosphate Competitive Phosphatidylinositol-3-kinase/Mammalian Target of Rapamycin Inhibitors: Discovery of Compound 26 (PKI-587), a Highly Efficacious Dual Inhibitor. *J. of Med. Chem.*, 2010.
- [82] E. Verspohl. Novel pharmacological approaches to the treatment of type 2 diabetes. *Pharmacological Reviews*, 2012.
- [83] D. Vitagliano et al. The tyrosine kinase inhibitor ZD6474 blocks proliferation of RET mutant medullary thyroid carcinoma cells. *Endocr Relat Cancer*, 2011.
- [84] D. Watts et al. Collective dynamics of 'small-world' networks. *Nature*, 1998.
- [85] S. Wilhelm et al. Discovery and development of sorafenib: a multikinase inhibitor for treating cancer. *Nat Rev Drug Discov*, 2006.
- [86] V. Williams. Multiplying matrices faster than Coppersmith-Winograd. *ACM STOC*, 2012.
- [87] D. Wishart et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res*, 2008.
- [88] S. Wuchty et al. Centers of complex networks. *Journal of Theoretical Biology*, 2003.
- [89] X. Yang et al. From Xenomouse® Technology to Panitumumab (ABX-EGF). *The Oncogenomics Handbook*, 2005.
- [90] T. Yap et al. First-in-man clinical trial of the oral pan-AKT inhibitor MK-2206 in patients with advanced solid tumors. *Journal of Clinical Oncology*, 2011.
- [91] T. Yeh et al. Biological characterization of ARRY-142886 (AZD6244), a potent, highly selective mitogen-activated protein kinase kinase 1/2 inhibitor. *Clin Cancer Res*, 2007.
- [92] M. Yıldırım et al. Drug-target network. *Nature Biotech.*, 2007.
- [93] C.-H. Yuh et al. Brn1/2/4, the predicted midgut regulator of the endo16 gene of the sea urchin embryo. *Dev. Biology*, 2005.
- [94] L. Zhao et al. Evaluation of Combination Chemotherapy. *Clin Cancer Res*, 2004.