TENET: Topological Feature-based Target Characterization in Signaling Networks

Huey Eng Chua§  Sourav S Bhowmick§  Lisa Tucker-Kellogg†
C. Forbes Dewey, Jr.‡

§School of Computer Engineering, Nanyang Technological University, Singapore
†Centre for Computational Biology, Duke-NUS Graduate Medical School, Singapore
‡Division of Bioengineering, Massachusetts Institute of Technology, USA

CHUA0530|assourav@ntu.edu.sg, lisa.tucker-kellogg@duke-nus.edu.sg, cfdewey@mit.edu

May 22, 2015
Abstract

Target characterization for a biochemical network is a heuristic evaluation process that produces a characterization model that may aid in predicting the suitability of each molecule for drug targeting. These approaches are typically used in drug research to identify novel potential targets using insights from known targets. Traditional approaches that characterize targets based on their molecular characteristics and biological function require extensive experimental study of each protein and are infeasible for evaluating larger networks with poorly-understood proteins. Moreover, they fail to exploit network connectivity information which is now available from systems biology methods. Adopting a network-based approach by characterizing targets using network features provides greater insights that complement these traditional techniques. To this end, we present TENET, a network-based approach that characterizes known targets in signaling networks using topological features. TENET first computes a set of topological features and then leverages a SVM-based approach to identify predictive topological features that characterizes known targets. A characterization model is generated and it specifies which topological features are important for discriminating the targets and how these features should be combined to quantify the likelihood of a node being a target. We empirically study the performance of TENET from a wide variety of aspects, using several signaling networks from BioModels with real-world curated outcomes. Results demonstrate its effectiveness and superiority in comparison to state-of-the-art approaches.
1 Introduction

Complex intra- and inter-cellular signaling drives various biological processes such as growth, proliferation and apoptosis within systems. In systems biology, these molecular interactions are typically modelled as signaling networks [50] that provide a holistic view of the various interactions between different molecular players in the system. As signaling networks become an increasingly acceptable way for representing biological systems, various network-based computational techniques have been developed to analyze these networks with the goal of answering biological needs such as target characterization [16] and target discovery [105]. In this paper, we focus on the target characterization problem for signaling networks.

Target characterization identifies characteristics (e.g., topological features) that distinguishes targets (i.e., nodes) from other nodes in the network. These characteristics can be summarized as models which we refer to as characterization models. Traditionally, targets are characterized based on their molecular characteristics (e.g., structure and binding sites of targets [62]) and biological function (e.g., regulation of apoptosis [104]). These traditional approaches focus primarily on the target alone and are oblivious to the presence of other interacting molecules in the system. However, understanding how a target interacts with other molecules in a biological system may provide valuable and holistic insights for superior target characterization. For example, the degree centrality of a target may be leveraged to assess potential toxicity of targets since high degree nodes tend to be involved in essential protein-protein interactions [37] and are potentially toxic as a result. In particular, network-based target characterization techniques can exploit such topological features for superior characterization of targets.

Recently, there have been increasing efforts toward devising network-based target characterization techniques [41, 65, 108]. These methods focus on using topological features to characterize targets of protein-protein interaction (PPI) networks. Specifically, McDermott et al. performed characterization of targets in protein co-abundance networks using several topological features such as degree centrality. Although this study suggests that multiple topological features can be combined for superior target characterization, it did not explore how these topological features should be combined towards this goal. In contrast, Hwang et al. concluded that bridging centrality is useful in identifying targets in PPI networks. However, the complexity and diversity of biological networks make target characterization using a single feature challenging since in some networks the chosen feature may perform poorly. Indeed, Chua et al. [16] showed that bridging centrality performs well in the MAPK–PI3K network, but not in the glucose metabolism network. Zhang et al. proposed the use of machine learning techniques such as support vector machines (SVM) and logistic regression for characterizing known targets in a manually curated human PPI network using 15 topological features. In contrast to [65], their goal was to identify topological characteristics of drug targets in gen-

---

1The protein co-abundance networks are essentially protein-protein interaction (PPI) networks constructed by identifying highly differentially regulated proteins from proteomics data using specific filters.
Symbol | Description
--- | ---
θ<sub>u</sub> | Degree centrality of node u. The in, out and total degree centralities are denoted as θ<sub>in(u)</sub>, θ<sub>out(u)</sub> and θ<sub>total(u)</sub>, respectively.
α<sub>u</sub> | Eigenvector centrality of node u.
β<sub>u</sub> | Closeness centrality of node u.
γ<sub>u</sub> | Eccentricity centrality of node u.
δ<sub>u</sub> | Betweenness centrality of node u.
π<sub>u</sub> | Bridging coefficient of node u.
ζ<sub>u</sub> | Bridging centrality of node u.
κ<sub>u</sub> | Clustering coefficient of node u. The undirected, in, out, cycle and middleman clustering coefficients are denoted as κ<sub>undir(u)</sub>, κ<sub>in(u)</sub>, κ<sub>out(u)</sub>, κ<sub>cyc(u)</sub> and κ<sub>mid(u)</sub>, respectively.
µ<sub>u</sub> | Proximity prestige of node u.
ω<sub>u</sub> | Target downstream effect of node u.

### Table 1: Topological features.

eral, instead of for specific diseases. However, characterizing targets in general assumes that targets of different diseases share similar target characteristics, which may not always be true. Indeed, as we shall see in Section 4, known targets in signaling networks tend to be characterized by different sets of topological features. Consequently, target characterization based on individual disease-specific networks may yield better characterization that is specific to the disease.

A common thread running through the aforementioned target characterization techniques is their focus on PPI networks. Surprisingly, similar systematic study in curated signaling networks has been lacking in the literature. Compared to signaling networks, PPI networks may contain many false-positive PPI in the sense that although these proteins can truly physically bind they may never do so inside cells due to different localization or they are not simultaneously expressed. Furthermore, PPI networks are static. That is, the edges in PPI networks are undirected; there is neither flow of information nor mass between nodes. Hence, they lack of knowledge of the underlying mechanism (i.e., actual signal flow) causing the disease. Since network quality directly affects the results of network-based target characterization, the aforementioned limitations of PPI networks may adversely impact the search for superior characteristics of targets. Signaling networks, however, model the dynamic interaction of the biological systems and present an attractive alternative to PPI networks.

In our recent work [16], we took the first step to demonstrate how signaling networks can be effectively leveraged to identify topological features that are discriminative of targets using the Wilcoxon test. However, similar to [65], this work does not shed any insight on a predictive model to combine these features for identifying potential targets. In this paper, we address this limitation by presenting TENET (Target characterization using Network Topology), a network-based approach that characterizes known targets in signaling networks using topological features. Specifically, we use a SVM-based approach to identify the set of topological features (referred to as predictive topological features) that characterizes known targets and to generate a characterization model using these features. The model specifies which topological features are important for discriminating the tar-
gets and how these features should be combined to produce a quantitative score that identifies the likelihood of a node being a target. In particular, TENET uses feature selection to select predictive topological features and weighted misclassification cost to handle SVM training issues such as noisy labels and imbalanced data. Our empirical study on four real-world curated signaling networks demonstrates the effectiveness and superiority of TENET.

The rest of the paper is organized as follows. In Section 2, we define some terminology and introduce the topological features being considered and the target curation process that is used for identifying known targets used subsequently for validation. Then, we formally define the target characterization problem and describe the TENET algorithm in Section 3. Finally, we present the experimental results in Section 4 and conclude the paper in Section 5.

2 Preliminaries

In this section, define some terminology and topological features, and introduce the target curation process that we shall be using in the sequel.

2.1 Terminology

A biological signaling network can be modelled as a directed hypergraph $G = (V, E)$ [50] where the nodes $V$ represent molecules (e.g., proteins) and the hyper-edges $E$ represent biochemical reactions and processes. A hyperedge connects one node set $U$ to another $W$, where $U, W \subseteq V$. For instance, in the activation of $\text{ERK}$, the set $U$ in the hyperedge consists of $\text{ERK}$ and its kinase, phosphorylated $\text{MEK}$ whereas $W$ contains the phosphorylated $\text{ERK}$ ($\text{ERK}^{\text{P}}$). Analysis of directed hypergraphs is generally more complex than graphs and many graph algorithms cannot be used directly on hypergraphs. Hence, they are often transformed into graphs containing simple edges for analysis. Methods (e.g., bipartite and substrate graph representation) exist for such transformation [50]. In this paper, we use the bipartite graph representation as it retains the original information of the hypergraph [50].

Signaling networks generally contain characteristics such as feedback and feedforward loops which are common in complex regulatory control [56]. These loops in turn give rise to graph characteristics such as strongly connected components (SCC).

The activity of nodes in the signaling network are generally governed by complex interconnectivity of various nodes in the same network. We refer to a node as a candidate target if when perturbed, it modulates the activity of a specific node (referred to as disease node). A disease 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. For instance, in the MAPK-Pi3K network [36] that is often implicated in cancer, $\text{ERK}^{\text{P}}$ can be considered as a disease node due to its role in proliferation. Given a signaling
network $G = (V, E)$ and a disease node $x \in V$, let the set of nodes having a path leading to $x$ be denoted as $V_x \subseteq V$. Then, the set of candidate target nodes in $G$ relevant to $x$ is denoted as $T_x \subseteq V_x$.

Network-based analysis can be applied to signaling networks to study the characteristics and properties of these networks. In this paper, we examine a total of sixteen topological features that are summarized in Table 1. These features are selected based on their role in measuring relative importance of a node in a signaling network. The formal definitions as well as motivation for selecting these features are given in [16] (also detailed in Section 2.2).

### 2.2 Topological Features

In this subsection, we introduce the 16 topological features that we study. These features are selected based on their role in measuring relative importance of a node in a signaling network. The MAPK–PI3K network (Figure 1) is used as a running example to illustrate the features.

**Degree Centrality.** It is a local centrality measure based on the number of edges a node has [32]. For directed networks such as the signaling network, there are three variants of degree centrality, namely, in degree, out degree and total degree centrality which consider only in-going edges, only out-going edges, and all edges of a node, respectively.

**Definition 1** Given a signaling network $G = (V, E)$, in degree centrality of a node $u \in V$ is defined as $\theta_{\text{in}}(u) = \sum_{v \in V} |e_{vu}|$ where $e_{vu} \in E$ is the edge connecting node $v \in V$ to $u$. Out degree centrality and total degree centrality are denoted...
### Table 2: Summary of the curation results.

<table>
<thead>
<tr>
<th>BioModel ID</th>
<th>MAPK-PI3K (I₁)</th>
<th>Glucose-Stimulated Insulin Secretion (I₂)</th>
<th>Endomesoderm Gene Regulation (I₃)</th>
<th>Glucose Metabolism (I₄)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Related disease or biological process</td>
<td>Ovarian cancer</td>
<td>Type 2 diabetes mellitus</td>
<td>Embryonic development</td>
<td>Glucose to acetate metabolism</td>
</tr>
<tr>
<td>No. of targets</td>
<td>9</td>
<td>6</td>
<td>206</td>
<td>16</td>
</tr>
<tr>
<td>Repository used for curation</td>
<td>ClinicalTrials.gov</td>
<td>ClinicalTrials.gov</td>
<td>PubMed</td>
<td>PubMed</td>
</tr>
<tr>
<td>Keywords used for curation</td>
<td>ovarian cancer drug</td>
<td>type 2 diabetes mellitus drug</td>
<td>sea urchin endomesoderm</td>
<td>E Coli glucose metabolism to acetate</td>
</tr>
<tr>
<td>Date of Curation</td>
<td>29 Apr 2014</td>
<td>25 Jan 2013</td>
<td>28 Oct 2013</td>
<td>14 Nov 2013</td>
</tr>
<tr>
<td>Unique Drugs Curated</td>
<td>458</td>
<td>617</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Relevant Drugs Curated</td>
<td>22</td>
<td>16</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

\[
\theta_{out}(u) = \sum_{v \in V} |e_{uv}| \quad \text{and} \quad \theta_{total}(u) = \theta_{in}(u) + \theta_{out}(u), \text{ respectively.}
\]

Generally, a node with high degree centrality (hub) is considered an important node. In particular, studies have found that biological networks resemble scale-free networks [75] in that they are robust against random perturbation of non-hub nodes [1]. Specifically, a high in degree node acts as a signal integrator by integrating multiple signals while a high out degree node acts as a signal differentiator. For instance, double phosphorylated MEK (MEKP) is an out degree hub and functions as a signal differentiator.

**Eigenvector Centrality.** Nodes with high eigenvector centrality are well-connected to other central nodes [5]. In a signaling network, these nodes tend to be located in the network where signals either converge or diverge depending on whether these central nodes have high in-degree or out-degree. For instance, activated ErbB4 receptor (RP) which has high eigenvector centrality is connected to many other central nodes such as PI3K*, and provides a means for converging and diverging the various signals passing through the network.

**Definition 2** Given a signaling network \( G = (V, E) \), let \( N_u \) be the set of neighbors of node \( u \in V \). Then, the **eigenvector centrality** of \( u \) is defined as \( \alpha_u = \frac{1}{\lambda} \sum_{v \in N_u} \alpha_v \) where \( \lambda \) is a constant.

According to the Perron–Frobenius theorem, in the above definition \( \lambda \) has to be the largest eigenvalue of the adjacency matrix\(^2\) \( A \) if the centralities are to be non-negative [71].

**Closeness Centrality, Eccentricity Centrality and Proximity Prestige.** These features are based on the proximity of a node to other nodes in the network. Closeness centrality assigns node centrality value using the sum of the shortest path distance [32] while eccentricity centrality uses the largest shortest path distance [103]. 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).

---

\(^2\)The adjacency matrix \( A = \{a_{ij}\} \) specifies the connectivity of the network such that \( a_{ij} = 1 \) implies an edge connecting node \( i \) to \( j \).
**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* $\beta_u$, *eccentricity centrality* $\gamma_u$ and *proximity prestige* $\mu_u$ of node $u$ are defined as

$$
\beta_u = \frac{|V|}{\sum_{v \in V} l_{uv}}, \quad \gamma_u = \frac{1}{\max\{l_{uv}\}}, \quad \mu_u = \frac{|I_u|}{|V| - 1 \sum_{v \in V} l_{uv}} - 1.
$$

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 [86]. 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 [86]. 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.

**Betweenness Centrality.** This feature assigns node centrality value based on the ease in which a node can reach other nodes in the network [6].

**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, AktPIP3, 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 [41]. 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 $\theta_{\text{total}}(v)$ be the total degree of node $v \in V$, $N_v$ be the set of neighbors of $v$, and $\eta_i$ 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

$$
\pi_v = \frac{1}{\theta_{\text{total}}(v)} \sum_{i \in N_v, \theta_{\text{total}}(i) > 1} \frac{1}{\eta_i} - 1.
$$

**Definition 6** Given the inverses of betweenness centrality rank and bridging coefficient rank of node $v$ denoted as $\psi_{\delta}^v$ and $\psi_{\pi}^v$, respectively, the *bridging centrality* is defined as

$$
\zeta_v = \psi_{\delta}^v \times \psi_{\pi}^v.
$$
Ovarian Cancer Drugs in [73] and their targets in the MAPK-PI3K network [36].

### Table 3: Ovarian cancer drugs in [73] and their targets in the MAPK-PI3K network [36].

<table>
<thead>
<tr>
<th>Ovarian Cancer Drugs [73]</th>
<th>Mechanism of Action</th>
<th>Target in 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lapatinib (Phase I) [82]</td>
<td>Bind to ATP-binding site of receptor (R), preventing its autophosphorylation</td>
<td>R</td>
</tr>
<tr>
<td>Sorafenib (Phase II) [100]</td>
<td>Bind to ATP-binding site of Raf, preventing activation of Raf</td>
<td>Raf</td>
</tr>
<tr>
<td>XSI 5132 (Phase I) [18]</td>
<td>Bind to Raf mRNA to downregulate Raf expression</td>
<td>Raf</td>
</tr>
<tr>
<td>AZD6244 (Phase I) [107]</td>
<td>Bind and lock HER into inactive conformation</td>
<td>HEREP</td>
</tr>
<tr>
<td>NL471 (Phase I) [72]</td>
<td>Bind to ATP-binding site of PI3K, preventing activation of PI3K</td>
<td>PI3K</td>
</tr>
<tr>
<td>Perifosine (Phase I) [54]</td>
<td>Bind to lipid-binding PH domain of Akt</td>
<td>Akt</td>
</tr>
<tr>
<td>ECO-4401 (Phase I) [7, 72]</td>
<td>Degrade Raf through proteasomal-dependent mechanism</td>
<td>Raf</td>
</tr>
<tr>
<td>PSI-179 (Phase I) [72]</td>
<td>Inhibits PI3K and mTOR kinases</td>
<td>PI3K</td>
</tr>
<tr>
<td>BKM120 (Phase I) [72]</td>
<td>ATP competitive inhibitor of PI3K kinase</td>
<td>PI3K</td>
</tr>
<tr>
<td>AZD5363 (Phase I) [2, 72]</td>
<td>ATP-competitive anti-Akt inhibitor</td>
<td>Akt</td>
</tr>
<tr>
<td>BLY719 (Phase I, II) [72]</td>
<td>Specifically inhibits PI3K in the PI3K/Akt kinase signaling pathway, thereby inhibiting the activation of the PI3K signaling pathway</td>
<td>PI3K</td>
</tr>
<tr>
<td>Dabrafenib (Phase I, II) [72]</td>
<td>Selectively binds to and inhibits the activity of B-raf, which may inhibit the proliferation of tumor cells which contain a mutated BRAF gene</td>
<td>Raf</td>
</tr>
<tr>
<td>GSK1120212 (Phase I, II) [42]</td>
<td>Potent and selective allosteric inhibitor of MEK1/2</td>
<td>MEK</td>
</tr>
<tr>
<td>GSK2110183 (Phase I, II) [88]</td>
<td>ATP-competitive anti-Akt inhibitor</td>
<td>Akt</td>
</tr>
<tr>
<td>GSK2141995 (Phase I) [72]</td>
<td>ATP-competitive anti-Akt inhibitor</td>
<td>Akt</td>
</tr>
<tr>
<td>MEK162 (Phase I) [72]</td>
<td>Non-competitive with ATP. Binds to and inhibits the activity of MEK1/2</td>
<td>MEK</td>
</tr>
<tr>
<td>MEK-2206 (Phase I) [106]</td>
<td>Oral pan-Akt inhibitor</td>
<td>Akt</td>
</tr>
<tr>
<td>Fimaterib (Phase I) [72]</td>
<td>Selectively binds to and inhibits the activity of MEK1/2, preventing the activation of MEK1/2-dependent effector proteins and transcription factors</td>
<td>MEK</td>
</tr>
<tr>
<td>SAR245409 (Phase I) [72]</td>
<td>Inhibits both PI3K kinase and mTOR kinase</td>
<td>PI3K</td>
</tr>
<tr>
<td>Traferminib (Phase I) [72]</td>
<td>Specifically binds to and inhibits MEK-1 and -2</td>
<td>MEK</td>
</tr>
<tr>
<td>Triciribine (Phase I, II) [72]</td>
<td>Inhibits the phosphorylation, activation, and signaling of Akt-1,-2, and -3</td>
<td>Akt</td>
</tr>
</tbody>
</table>

---

Figure 2: The sea urchin Endo16 regulatory pathway. The nodes in this pathway are targets for regulating Endo16 in I3.

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 [99] by considering how close the neighbourhood is to being a
<table>
<thead>
<tr>
<th>T2DM Drugs/Food Constituents in [73]</th>
<th>Mechanism of Action</th>
<th>Target in I₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol (Phase II) [22]</td>
<td>Causes hyperglycemia</td>
<td>glucose</td>
</tr>
<tr>
<td>Avandamet (Rosiglitazone + Metformin) (Phase IV) [83]</td>
<td>Rosiglitazone is highly selective and potent agonist for the PPARγ, Metformin decreases hepatic glucose production, decreases intestinal absorption of glucose and increases peripheral glucose uptake and utilization</td>
<td>glucose</td>
</tr>
<tr>
<td>Metformin (Phase IV) [51, 98]</td>
<td>Inhibit respiratory chain complex 1 of mitochondria, thereby reducing cellular energy transiently and activating AMPK. Increase peripheral glucose uptake and utilization</td>
<td>glucose</td>
</tr>
<tr>
<td>Benfluorex (Phase II) [53]</td>
<td>Reduce β-oxidation (process by which fatty acid molecules break down in mitochondria to generate acetyl-CoA) rates and ketogenesis. Reduce gluconeogenesis from lactate/pyruvate</td>
<td>acetyl-CoA, glucose</td>
</tr>
<tr>
<td>Berberine (Phase III) [90]</td>
<td>Mimick insulin action, improve insulin action by activating AMPK, reduce insulin resistance through PPK-dependent up-regulation of insulin receptor expression, inducing glycolysis, promoting Glu-1 secretion and modulating its release, inhibition of sPP-4</td>
<td>insulin, glucose</td>
</tr>
<tr>
<td>Cod (Phase II) [57]</td>
<td>Increase insulin-stimulated glucose uptake</td>
<td>intracellular glucose</td>
</tr>
<tr>
<td>Deproteinised hemoderivative of call blood (Actovegin) (Phase III) [61]</td>
<td>Actovegin is composed of small molecules present under normal physiological conditions, therefore pharmacokinetic and pharmacodynamic studies to determine its active substance are not feasible. Increase glucose uptake and improve oxygen uptake under conditions of ischemia</td>
<td>glucose</td>
</tr>
<tr>
<td>Tagatose (Phase III) [25]</td>
<td>Interfere with carbohydrate absorption by inhibiting intestinal disaccharidases and glucose transport</td>
<td>plasma glucose</td>
</tr>
<tr>
<td>Vinegar [45]</td>
<td>Reduce plasma renin activity and aldosterone concentration, Reduce serum glucose</td>
<td>serum glucose</td>
</tr>
<tr>
<td>Rice [40]</td>
<td>Increase dietary glycemic load</td>
<td>blood glucose</td>
</tr>
<tr>
<td>cs-917 (Phase II) [24]</td>
<td>Competitively inhibits fructose-1,6-bisphosphatase (FBPase) at the AMP binding site</td>
<td>activated FBPase</td>
</tr>
<tr>
<td>ms07803 (Phase II) [95]</td>
<td>Fructose-1,6-bisphosphatase (FBPase) inhibitor</td>
<td>activated FBPase</td>
</tr>
<tr>
<td>Glycerol [59]</td>
<td>Synthesized from glycerol using glycerol kinase</td>
<td>glycerol-3-phosphate</td>
</tr>
<tr>
<td>Methylcobalamin [30]</td>
<td>Provides methyl group that couples to c to synthesize acetyl-CoA</td>
<td>acetyl-CoA</td>
</tr>
<tr>
<td>Gynostemma Pentaphyllum tea (Phase II) [4, 67]</td>
<td>Increase superoxide dismutase (SOD). Strong inhibition on IL-6 and Ftg25 mRNA expression and weak inhibition on TNF-α mRNA expression. Gypenosides, the major components of Gynostemma pentaphyllum, increase Bax, reduce Bcl-2 and stimulate release of cytochrome c, AIF (apoptosis-inducing factor), and Endo G (endonuclease G) from mitochondria. Ferricytochrome c may be required for depolarization of the mitochondrial</td>
<td>ferrocytochrome c</td>
</tr>
<tr>
<td>Xanthohumol [4, 19]</td>
<td>Binds annexin V-FITC, cleaves FAP-1 and activates procaspases-3, -8 and -9. Depolarizes mitochondrial leading to release of cytochrome c. Inhibits Akt, Wnk1, mTOR, Bcl-2 and survivin. Ferricytochrome c reduction to ferrocytochrome c may be required for depolarization of the mitochondrial</td>
<td>ferrocytochrome c</td>
</tr>
</tbody>
</table>

Table 4: T2DM drugs or food constituents in [73] and their targets in the glucose-stimulated insulin secretion network [44].

![Directed triangle graphs](image-url)

Figure 3: Directed triangle graphs adapted from [26].

clique where every node within the clique is connected to every other node in it [1]. The original definition was meant for undirected graph. A variety of definition exists [26] when edge directions are considered (Figure 3).

**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 $\kappa_{\text{undir}}(u)$, $\kappa_{\text{in}}(u)$, $\kappa_{\text{out}}(u)$, $\kappa_{\text{cyc}}(u)$ and $\kappa_{\text{mid}}(u)$, respectively, are defined as $\kappa_{\text{undir}}(u) = \frac{(A^T A)}{\mu_{\text{total}}(u)(\mu_{\text{total}}(u) - 1)}$, $\kappa_{\text{in}}(u) = \frac{(A^T A)_{ii}}{\mu_{\text{in}}(u)(\mu_{\text{in}}(u) - 1)}$, $\kappa_{\text{out}}(u) = \frac{(A^T A)_{ii}}{\mu_{\text{out}}(u)(\mu_{\text{out}}(u) - 1)}$, $\kappa_{\text{cyc}}(u) = \frac{(A^T A)_{ii}}{\mu_{\text{cycle}}(u)(\mu_{\text{cycle}}(u) - 1)}$, and $\kappa_{\text{mid}}(u) = \frac{(A^T A)_{ii}}{\mu_{\text{middle}}(u)(\mu_{\text{middle}}(u) - 1)}$. Where $\mu_{\text{total}}(u)$, $\mu_{\text{in}}(u)$, $\mu_{\text{out}}(u)$, $\mu_{\text{cycle}}(u)$, and $\mu_{\text{middle}}(u)$ represent the degree of node $u$ in the graph.
Table 5: Targets that are crucial for acetate production in the glucose metabolism network in [55]. * indicates targets that are included by default due to their direct involvement (either as input or output) in the metabolic reaction being studied.

<table>
<thead>
<tr>
<th>Target ID in BIOMD0000000244 (I4)</th>
<th>Name</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT</td>
<td>acetate</td>
<td>+</td>
</tr>
<tr>
<td>Glc</td>
<td>glucose</td>
<td>[109]</td>
</tr>
<tr>
<td>G6P</td>
<td>glucose-6-phosphate</td>
<td>[109]</td>
</tr>
<tr>
<td>Ict</td>
<td>isocitrate</td>
<td>[94]</td>
</tr>
<tr>
<td>Pfp</td>
<td>phosphoenolpyruvate</td>
<td>[10, 27]</td>
</tr>
<tr>
<td>AceB</td>
<td>enzyme for reaction from ACoA to ACT</td>
<td>[9]</td>
</tr>
<tr>
<td>Cya</td>
<td>adenylate cyclase</td>
<td>[76]</td>
</tr>
<tr>
<td>Fdp</td>
<td>fructose-1,6-bisphosphatase</td>
<td>[78]</td>
</tr>
<tr>
<td>Icd</td>
<td>unphosphorylated isocitrate dehydrogenase</td>
<td>[46]</td>
</tr>
<tr>
<td>IcdP</td>
<td>phosphorylated isocitrate dehydrogenase</td>
<td>[46]</td>
</tr>
<tr>
<td>Pdh</td>
<td>pyruvate dehydrogenase</td>
<td>[69]</td>
</tr>
<tr>
<td>Ppc</td>
<td>phosphoenolpyruvate synthase</td>
<td>[27]</td>
</tr>
<tr>
<td>EIIA</td>
<td>unphosphorylated pts protein EIIA</td>
<td>[43]</td>
</tr>
<tr>
<td>EIIc</td>
<td>pts protein EIIc (ptsB)</td>
<td>[79]</td>
</tr>
</tbody>
</table>

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 node $w$ and the likelihood of $w$ causing off-target effect [15].

**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 $\theta_{\text{total}(w)}$ be the total degree of $w$. The **target downstream effect** of $v$ is defined as $\omega_v = \sum_{w \in W} (\rho_{v,w} \times \theta_{\text{total}(w)})$.

### 2.3 Target Curation Process

In this section, we describe the target curation process used to identify the set of benchmark targets required by TENET. Manual curation of literature generates substantially lower error rates than text mining-based approaches [89]. In this article, we study two categories of networks: networks associated with human diseases.
and networks describing biological processes. Table 2 summarizes the curation results of the four signaling networks we studied.

**Human Disease-Related Networks.** Amongst the four networks we study, two are associated with human diseases. These networks are the MAPK–PI3K [36] and the glucose-stimulated insulin secretion networks [44]. 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 [73].
2. Obtain the targets of these drugs and compounds via drug related databases [102] and literature survey.
3. Identify the targets that are in the scope of the signaling network.

**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 curated targets are listed in Tables 3 to 5 and Figure 2.

### 3 Target Characterisation

In this section, we formally define the target characterization problem and describe the TENET algorithm.

#### 3.1 Topological Feature-based Target Characterization

Intuitively, the goal of topological feature-based target characterization is to use a set of predictive topological features to characterize known targets in a network. Hence, the topological feature-based target characterization problem can be formulated as a supervised learning problem. In a supervised learning problem, a training set \( \{(x_i, f(x_i))\} \) is given where \( f(x_i) \) is the predictor of \( x_i \) and the goal is to learn some target function \( f : X \rightarrow Y \) which can be applied to predict unseen data \( w \). The problem can be subdivided into two categories: regression when the predictor yields a continuous outcome and classification when the outcome is discrete. A regression problem can be converted into a binary classification problem by specifying a threshold \( h \) and assigning \( x_i \) with \( f \) greater than \( h \) to one class and the remaining to the other class. We advocate that the topological feature-based target characterization problem is best represented as a regression problem. In this problem, we are interested in finding out how likely one node is a target relative to another node based on a set of predictive topological features. This is
different from the target classification problem where we want to find out the class membership of a node. Note that the regression problem can be converted into a classification problem by specifying a threshold $h$ and assigning nodes having target function greater than $h$ to the target class and the rest to the non-target class.

Although we examine sixteen topological features, as we shall see later, not all features are relevant to a given signaling network. In fact, incorporating irrelevant features may adversely impact the performance of the prediction model. Hence, it is important to learn a set of predictive topological features that best characterizes targets (referred to as topological feature selection) for a given network. Formally, it is defined as follows.

**Definition 9** Given a signaling network $G = (V, E)$ and a disease node $x \in V$, let $T_x \subseteq V$ and $\mathcal{X}_{all}$ denote the set of known targets in $G$ relevant to $x$, and the set of topological features of $G$, respectively. Then, the goal of topological feature selection is to find a set of predictive topological features $\mathcal{F} \subseteq \mathcal{X}_{all}$ that maximizes the prediction accuracy for $f(\xi(u, F))$ subject to the following conditions:

$$
\begin{align*}
  f(\xi(u, F)) &= 1 & \text{when } u \in T_x, \\
  f(\xi(u, F)) &= 0 & \text{otherwise}. \\
\end{align*}
$$

Then the topological feature-based target characterization problem is formally defined as follows.

**Definition 10** Given a signaling network $G = (V, E)$, a disease node $x \in V$, $T_x$, and $\mathcal{X}_{all}$, let $\mathcal{F}$ denote the set of predictive topological features. Then, for a threshold $h$, the goal of the topological feature-based target characterization problem is to identify a set of predictive topological features $\mathcal{F} \subseteq \mathcal{X}_{all}$ using topological feature selection and learn a characterization model $g(\xi(u, \mathcal{F}))$ subject to the conditions

$$
\begin{align*}
  g(\xi(u, \mathcal{F})) &\in \mathbb{R}, \\
  g(\xi(u, \mathcal{F})) &\geq h & \text{when } u \in T_x, \\
  g(\xi(u, \mathcal{F})) &< h & \text{otherwise},
\end{align*}
$$

that maximizes the target prediction for $g(\xi(u, \mathcal{F}))$.

Figure 4 depicts a pictorial overview of the topological feature-based target characterization problem. For example, given the MAPK-P13K signaling network, its associated disease node ERKPP, the set of known targets in this network, and the topological features in Table 1, the goal of this problem is to produce the followings: (a) Identify the set of predictive topological features $\mathcal{F} = \{ \delta, \pi, \theta_{in}, \theta_{out} \}$ and (b) learn a characterization model $g(\xi(ERKPP, \mathcal{F}))$. Note that in Definition 10, there is no need to explicitly specify a threshold $h$ if we are only interested in obtaining the relative rankings of the nodes. The threshold is required if we want to assign class labels (e.g., target class) to the nodes.
3.2 SVM-based Target Characterization

We employ support vector classification (SVC) to select predictive topological features and support vector regression (SVR) to generate the characterization model. The SVC and SVR are typically formulated as constrained optimization problems and solved using the Lagrangian multiplier method. In general, SVM models contain multiple parameters, such as the cost parameter $C$ and parameters related to the kernel function, that affect the learning and performance of the models [11]. We follow the method in [39] for training the SVM. The feature values are scaled linearly to the range of $[0, 1]$ for each signaling network to avoid features with larger ranges dominating those with smaller ranges. We use stratified cross-validation (described below) and grid-search [39] on the training data to identify the best values of the model parameters. Note that cross-validation helps us to avoid the issue of overfitting the data whereas stratification enables us to keep the percentage of targets in the different folds similar to the original dataset. The best parameter is the one that yields the best average prediction accuracy for the cross-validation process. Wherever possible, we use a 10-fold stratified cross-validation since larger fold numbers reduce pessimistic bias and 10 folds generally give good performances [52].

Several non-trivial issues, namely, irrelevant or redundant features, noisy labels and imbalanced data set, need to be addressed in training the SVM model for characterizing targets. In particular, we use feature selection to select for appropriate features to be used in the SVM model and cost-sensitive learning to handle the issue of noisy labels and imbalanced data set. We examine three feature selection approaches, namely, backward stepwise elimination (BSE) [63], Wilcoxon-ROC based elimination (WRE) and WRE-BSE. BSE is classifier-aware whereas WRE is classifier-independent. WRE-BSE which performs WRE followed by BSE is a hybrid approach. Note that compared to classifier-independent methods, classifier-aware methods interact with the classifiers and such interaction can lead to better classification results [84]. However, they are typically computationally expensive and run the risk of model-overfitting. Cost sensitive learning is an algorithmic approach that chooses an appropriate strategy specific to the classifier to overcome the bias introduced by imbalanced data and the noise caused by uncertainty in labelling. We use weighted misclassification cost (WMC), an approach that proportionates the misclassification cost of the training data according to class. In particular, we use a variable $C_i$ as the cost parameter $C$:

$$C_i = \begin{cases} C^+ & \text{if } y_i = +1 \\ C^- & \text{if } y_i = -1 \end{cases}$$

subject to the constraints $C^+ + C^- = 1$, $C^+ > 0$ and $C^- > 0$ where $y_i$ is the

---

4The training data was sampled from the original data such that the ratio of the targets to non-targets is similar to that of the original data.

5In our study, we set a lower bound of one target in all our test sets.
class predictor and \(C^+\) and \(C^-\) denote the misclassification cost of the target and non-target classes, respectively.

**Data partitioning (stratified sampling).** Note that for the networks studied, the target class for all nodes including the test set is known. This is for the purpose of validating our approach later in the experiments. We partition the data into training and test set by following two rules. First, there should be at least one target in the test set. This allows us to determine if TENET is able to rank the curated target higher than other nodes. Second, the ratio of target nodes to non-target nodes should mirror that of the original data set as close as possible. This ensures the real distribution of targets versus non-targets in the networks is retained. Using these two rules, we determine the number of targets and non-targets in the test set for each network. Then, the targets and non-targets are randomly selected from the original data set to generate the test set. Finally, the remaining nodes form the training set. The test set of \(I_1\) to \(I_4\) are provided in Table 6. Note that these same rules are followed when generating individual folds from the training set.

**Cross validation.** We use cross validation (illustrated in Figure 5) for training the SVM. Briefly, the training data (two matrices, one for candidate targets and one for candidate non-targets) is partitioned using stratified sampling into multiple folds.
folds. For the training, multiple iterations are needed to explore different sets of SVM parameters. In each iteration, $x$ models corresponding to the number of folds are generated. For each model, one of the folds is excluded from the training and used for validating the model. Then, the model accuracies are averaged. Cross validation terminates when the SVM parameters have been satisfactorily explored. In TENET, we use the grid-search approach (detailed in [39]) for exploring the SVM parameters. The best SVM model is the one with parameters yielding the highest average model accuracy.

### 3.3 The TENET algorithm

Given a signaling network $G = (V, E)$, a disease node $x \in V$, a known target set $T_x \subseteq V$, a set of topological features $\mathcal{X}_{all}$ and a step-size of the misclassification cost $s$, TENET identifies the set of predictive structural features and a characterization model that best characterizes these known targets. Note that $\mathcal{X}_{all}$ and $s$ are optional inputs and are set to default values$^8$ if they are not given. The known targets $T_x$ can be extracted by following the curation process described in [16] (Section 2.3). The TENET algorithm is given in Algorithms 1 to 5 and is comprised of three phases, namely, the pruning phase (Algorithm 2), the feature extraction phase and the model training phase. First, the pruning phase identifies relevant...
Algorithm 1: Algorithm TENET

**Input:** Signaling network $G$, disease node $x$, target set $T_x$, topological feature set $X_{all}$ (optional) and step-size of the misclassification cost $s$ (optional).

**Output:** Predictive topological feature set $F$ and characterization model $M$.

1. $F, M, X_{all} \leftarrow \text{INITIALIZE}(F, M, X_{all})$  // Phase 1
2. $V_{\text{candidate}} \leftarrow \text{FILTERCANDIDATE}(G, x)$;  // Phase 2
3. $H \leftarrow \text{EXTRACTFEATURE}(G, V_{\text{candidate}}, X_{all})$;
4. $F, M \leftarrow \text{TRAINMODEL}(H, T_x, s)$;  // Phase 3

Algorithm 2: Procedure FILTERCANDIDATE

**Input:** Signaling network $G$, disease node $x$.

**Output:** Candidate target node set $V_{\text{candidate}}$.

1. $G_{BI} \leftarrow \text{CONVERT2BIPARTITEGRAPH}(G)$
2. $G_{DAG} \leftarrow \text{CONVERT2DAG}(G_{BI})$
3. $U \leftarrow \text{GETROOTNODES}(G_{DAG})$
4. foreach iteration $i=1$ to $|U|$ do
5.  $G_{DAG} \leftarrow \text{INDEX}(G_{DAG}, U_i, \text{null})$
6. $V_{\text{candidate}} \leftarrow \text{ASSESSREACHABILITY}(G_{DAG}, x)$

nodes (denoted as $V_{\text{candidate}}$) that shall be used for training the SVM. Then, the feature extraction phase extracts all the topological features (denoted as $X_{all}$) of each candidate node and stores them in a $|V_{\text{candidate}}| \times |X_{all}|$ matrix $H$. Finally, in the model training phase, TENET learns the optimal set of predictive topological features $F$ and the best model parameters of the characterization model $M$. We shall now describe these phases in turn.

**Phase 1: Pruning.** In this phase, TENET prunes nodes that do not have paths leading to the disease node $x$. This phase yields a set of potential candidate nodes $V_{\text{candidate}} \subset V$ and is used to reduce the subsequent computation. In the pruning process, the given network $G$ is first preprocessed into a bipartite graph and then converted into a directed acyclic graph (DAG) (Algorithm 3), a graph with consistent topological ordering, to facilitate indexing of nodes (Algorithm 4). Note that the node indices shall be used subsequently to perform reachability check to identify the nodes to be pruned. We adopt the method in [23] for bipartite graph conversion. In order to convert the bipartite graph into its DAG representation, we adopt the approach in [92] to identify SCCs and replace each SCC with a representative node (referred to as meta node). Then, we adopt the indexing approach of [13] to index the DAG. This indexing approach performs depth-first traversal to assign each node $v$ a preorder index (when $v$ is first visited) and a postorder index (when all descendant nodes of $v$ are visited). Finally, an index-based reachability algorithm is used to determine if there exists a path from each node $v$ to the disease node $x$ (denoted as $v \rightarrow x$). Given a node $v$ and $x$, let $w$ be the descendant of $v$ that is not in the spanning tree (referred to as non-spanning tree node) and $v.preorder$ and $v.postorder$ denote the preorder and postorder indexes of $v$, respectively. A path $v \rightarrow x$ exists if any of the following conditions are satisfied [13]:

1. $v.preorder \leq x.preorder$ and $v.postorder \geq x.postorder$
Algorithm 3: Procedure CONVERT2DAG

Input: Bipartite graph $G_{B} = (V_{B}, E_{B})$.
Output: Directed acyclic graph (DAG) $G_{DAG}$.

1. $G_{DAG} \leftarrow G_{B}$
2. $G_{DAG}.SCC \leftarrow \text{GETSCC}(G_{DAG})$
3. foreach iteration $i = 1$ to $|G_{DAG}.SCC|$ do
   4. $V \leftarrow \text{INSERTNODE}(V, v_{meta,i})$
   5. $X \leftarrow \text{GETNODESINSCC}(G_{DAG}, SCC_{i})$
   6. foreach iteration $j = 1$ to $|X|$ do
      7. $N \leftarrow \text{GETNEIGHBOURSNOTINSCC}(X_{j}, X)$
     8. foreach iteration $k = 1$ to $|N|$ do
        9. $E \leftarrow \text{REPLACEEDGE}(E, (X_{j}, N_{k}), (v_{meta,i}, N_{k}))$
       10. $E \leftarrow \text{REPLACEEDGE}(E, (N_{k}, X_{j}), (N_{k}, v_{meta,i}))$
      11. $V \leftarrow \text{REMOVENODE}(V, X_{j})$

Algorithm 4: Procedure INDEX

Input: DAG $G_{DAG} = (V_{DAG}, E_{DAG})$, child node $u$, parent node $v$.
Output: DAG $G'_{DAG}$.

1. if $v.preorder = \text{null}$ then
   2. $v.preorder \leftarrow \text{SETTONEXTINDEX}(v.preorder)$
3. foreach iteration $i = 1$ to $|V_{DAG}|$ do
   4. $w \leftarrow \text{GETCHILDNODE}(v)$
   5. $G_{DAG} \leftarrow \text{INDEX}(G_{DAG}, w, v)$
   6. $v.descendants \leftarrow \text{INSERTNODE}(v.descendants, w)$
   7. $v.descendants \leftarrow v.descendants \cup w.descendants$
   8. $v.NSTNodes \leftarrow v.NSTNodes \cup w.NSTNodes$
   9. $v.postorder \leftarrow \text{SETTONEXTINDEX}(v.postorder)$
10. else if $v.preorder \neq \text{null}$ and $u \neq \text{null}$ then
11. $u.descendants \leftarrow \text{INSERTNODE}(u.descendants, v)$
12. $u.NSTNodes \leftarrow \text{INSERTNODE}(u.NSTNodes, v)$
13. if $u \neq \text{null}$ then
14. $u.descendants \leftarrow u.descendants \cup v.descendants$
15. $u.NSTNodes \leftarrow u.NSTNodes \cup v.NSTNodes$

2. $w.preorder \leq x.preorder$ and $w.postorder \geq x.postorder$

Note that the pruning step is beneficial in improving execution time for larger sparsely connected networks and for disease node that are positioned further upstream. For instance, in the MAPK-P38K network, no nodes are pruned when we select ERKPP (downstream) as the disease node whereas 17 nodes (47.2%) are pruned when activated Ras (RasGTP) (upstream) is selected.

Phase 2: Feature Extraction. In this phase, for all nodes in $V_{\text{candidate}}$, TENET extracts all the topological features in Table 1 for characterizing the known targets.

Phase 3: Model Training. Given a matrix of topological feature values $H$, a target set $T$, and a step-size of the misclassification cost $s$, this phase identifies a set of predictive topological features $F$ and the best parameters for configuring the characterization model $M$. First, the misclassification cost of the target class $C^{+}$ is initialized to a default value of 0.5. Then, feature selection is used to obtain the predictive topological feature set $F$. We iterate over three different feature selection approaches (BSE, WRE and WRE-BSE). Next, the step-size $s$ is used to step
Algorithm 5: Procedure TRAINMODEL

**Input:** Matrix of topological feature values $H$, known target set $T_x$ and step-size of the misclassification cost $s$.

**Output:** Predictive topological feature set $F$ and characterization model $M$.

1. $C^+ \leftarrow \text{INITIALIZE}()$
2. $\phi_{\text{best}}, F, M \leftarrow \text{SELECTFEATURES}(H,T_x,C^+)$
3. **foreach** iteration $i = 1$ to $\frac{1}{s} - 1$ **do**
4.   $C^+ \leftarrow i \times s$
5.   $\phi_{\text{curr}} \leftarrow \text{TUNE}\text{SVM}(F,T_x,C^+)$
6.   **if** $\phi_{\text{curr}} > \phi_{\text{best}}$ **then**
7.     $\phi_{\text{best}} \leftarrow \phi_{\text{curr}}$
8.     $M \leftarrow \text{SET}\text{BEST}\text{PARAMWEIGHT}(M,C^+)$

Algorithm 6: Algorithm WRE

**Input:** Matrix of topological feature values $H$, target set $T_x$ and misclassification cost for target class $C^+$.

**Output:** Prediction accuracy $\phi$, predictive topological feature set $F$ and characterization model $M$.

1. $R_{\text{Wilcoxon}} \leftarrow \text{WILCOXONFILTER}(H,T_x)$
2. $R_{\text{ROC}} \leftarrow \text{ROCFILTER}(H,T_x)$
3. $F \leftarrow R_{\text{Wilcoxon}} \cap R_{\text{ROC}}$
4. $\phi, M \leftarrow \text{TUNE}\text{SVM}(F,T_x,C^+)$

through the range of misclassification cost $(0 – 1)$. In each iteration, the misclassification cost of the target class $C^+$ is incremented according to the number of iterations completed, before the SVM training (Algorithm 5) is performed to obtain the parameter settings of the characterization model $M$ with the best accuracy.

The BSE approach is a well-known greedy approach that progressively removes features from the naïve SVM model (built using all topological features) and trains a new best model after each feature removal. The elimination process stops when removal of additional features result in a worse average accuracy of the validation set prediction. In contrast, the WRE approach (Algorithm 6) performs two statistical tests, namely, one-tailed Wilcoxon Rank-Sum (referred to as Wilcoxon) and receiver operating characteristics (referred to as ROC). The results are used to eliminate features that do not discriminate between targets and non-targets in a significant manner (based on Wilcoxon) and that do not classify targets well (based on ROC). Note that we perform two 1-tailed Wilcoxon tests and for each test; $p$-values smaller than 0.05 are considered significant. Hence, we take the difference of the $p$-values for both test hypotheses (referred to as $p$-value difference) and remove features with $p$-value difference less than 0.9. For the ROC analysis, features with AUC less than 0.7 [38] are considered poor performers and are removed. The best characterization model is found by training the SVM using the remaining features. The WRE-BSE approach (Algorithm 7) first performs WRE followed by BSE.
Algorithm 7: Algorithm WRE-BSE

Input: Matrix of topological feature values $H$, target set $T_x$, misclassification cost for target class $C^+$
Output: Prediction accuracy $\phi$, predictive topological feature set $\mathcal{F}$, characterization model $\mathcal{M}$

1. $\mathcal{R}_{\text{Wilcoxon}} \leftarrow \text{WILCOXONFILTER}(H, T_x)$
2. $\mathcal{R}_{\text{ROC}} \leftarrow \text{ROCFILTER}(H, T_x)$
3. $\mathcal{F} \leftarrow \mathcal{R}_{\text{Wilcoxon}} \cap \mathcal{R}_{\text{ROC}}$
4. $\phi_{\text{prevBest}}, \text{param}_{\text{prevBest}} \leftarrow \text{TUNE}\text{SVM}(\mathcal{F}, T_x, C^+)$

repeat

6. foreach iteration $i = 1$ to $|\mathcal{F}|$ do
7. \[ \phi_i, \text{param}_i \leftarrow \text{TUNE}\text{SVM}(\mathcal{F} - \mathcal{F}_i, T_x, C^+) \]
8. $\mathcal{F}_i, \phi_{\text{currBest}}, \text{param}_{\text{currBest}} \leftarrow \text{GETFEATURETOREMOVE}(\phi_i, \text{param}_i)$
9. if $\phi_{\text{currBest}} > \phi_{\text{prevBest}}$ then
10. \[ \mathcal{F} \leftarrow \mathcal{F} - \mathcal{F}_i \]
11. $\mathcal{M} \leftarrow \text{param}_i$

until $|\mathcal{F}| = 1$ or $\phi_{\text{currBest}} < \phi_{\text{prevBest}}$;

<table>
<thead>
<tr>
<th>Structural Features</th>
<th>Time Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree centrality</td>
<td>$O(</td>
</tr>
<tr>
<td>Eigenvector centrality</td>
<td>$O(</td>
</tr>
<tr>
<td>Closeness centrality</td>
<td>$O(</td>
</tr>
<tr>
<td>Eccentricity centrality</td>
<td>$O(</td>
</tr>
<tr>
<td>Betweenness centrality</td>
<td>$O(</td>
</tr>
<tr>
<td>Bridging centrality</td>
<td>$O(</td>
</tr>
<tr>
<td>Bridging coefficient</td>
<td>$O(</td>
</tr>
<tr>
<td>Clustering coefficient</td>
<td>$O(</td>
</tr>
<tr>
<td>Proximity prestige</td>
<td>$O(</td>
</tr>
<tr>
<td>Target downstream effect</td>
<td>$O(</td>
</tr>
</tbody>
</table>

Table 7: Time complexity for computing the different features. The proofs of those algorithm complexities that are provided without citation are given in Section 3.4.

3.4 Complexity Analysis

In this subsection, we present the complexity analysis of TENET. We start by providing the complexity analysis for the computation of the topological features considered (summarized in Table 7) in TENET.

Degree Centrality.

Theorem 1 Computation of degree centrality requires $O(|V|)$ time in the worst case.

Proof 1 It takes $O(|V|)$ time to iterate through all the nodes in the graph since it requires constant time to retrieve the number of edges associated to each node.

Bridging Coefficient.

Theorem 2 Computation of bridging coefficient requires $O(|V|^2)$ time in the worst case.

Proof 2 For each node in the network, the computation of the bridging coefficient iterates through all the neighbours of the node (Definition 5). In the worst case, the
network is a single strongly connected component where every node is a neighbour of all other nodes. Hence, calculating the bridging coefficient of all nodes in the network requires $O(|V|^2)$ in the worst case.

**Bridging Centrality**

**Theorem 3** Computation of bridging centrality requires $O(|V|^2)$ time in the worst case.

**Proof 3** For each node in the network, the computation of the bridging centrality is a product of the inverse of the betweenness rank and the bridging coefficient rank (Definition 6). Hence, computation requires $O(|V|^2 + |V| + |E|)$ time in the worst case. This can be further simplified into $O(|V|^2 + |E|)$.

**Proximity Prestige**

**Theorem 4** Computation of proximity prestige requires $O(|V|^2 + |E|)$ time in the worst case.

**Proof 4** In Definition 3, the set of nodes having at least one path leading to node $u$ ($I_u$) and the shortest path distance are needed for calculating the prestige value. Using the ASSESSREACHABILITY procedure in Algorithm 2 to obtain $I_u$ requires $O(|V|^2 + |E|)$ time. Space is required to store the node and edge information of the input graph. The shortest path distance can be found using Dijkstra’s algorithm [20] and the computation requires $O(|E| + |V|\log_2 |V|)$ time using Fibonacci heaps [31]. In the worst case, $O(|V|^2 + |E|)$ time is needed since $O(|V|^2) > O(|V|\log_2 |V|)$.

**Target Downstream Effect**

**Theorem 5** Computation of target downstream effect requires $O(|V|^2 + |E|)$ time in the worst case.

**Proof 5** According to Definition 8, the computation of the target downstream effect for each node requires iterating through each of its downstream nodes. The downstream nodes can be found by using the ASSESSREACHABILITY procedure in Algorithm 2. The time required to compute the reachability of the nodes is $O(|V|^2 + |E|)$. In the worst case, the network is a single strongly connected component where every node has a path leading to all other nodes. Hence, calculating the target downstream effect of all nodes in the network requires $O(2|V|^2 + |E|)$ in the worst case and can be simplified to $O(|V|^2 + |E|)$.

**TENET Algorithm**

**Theorem 6** Given a signaling network $G$, a disease node $x$, a target set $T_x$, a feature set $X_{all}$ and the step-size of the misclassification cost $s$, the Algorithm TENET has worst-case time complexity $O(|V|^2 + |E|)^2 + O(G(X_{all})) + O(T(\cdot)))$ where $G(X_{all})$ is the worst-case time complexity for extracting the features and $O(T(\cdot))$ is the worst-case time complexity of the feature selection method used.
**Proof 6** In the FILTERCANDIDATE algorithm, the conversion of the input signaling network $G = (V, E)$ to a bipartite graph $G_{BI} = (V_{BI}, E_{BI})$ takes $O(|V_{BI}| + |E_{BI}|)$ time. In DAG conversion, $O(|V_{BI}| + |E_{BI}|)$ time is required for finding SCC using [92]. In the worst case, the signaling network is a complete directed graph and CONVERT2DAG takes $O(|V_{BI}|^2 + |E_{BI}|)$ time since $|V_{BI}| < |V_{BI}|^2$. In the indexing of the DAG graph $G_{DAG} = (V_{DAG}, E_{DAG})$, the depth-first traversal requires $O(|V_{DAG}| + |E_{DAG}|)$ time while computing the set of nodes that can reach $x$ takes $O(|V_{DAG}|)$ time. Hence, FILTERCANDIDATE algorithm takes $O(|V_{BI}|^2 + |E_{BI}|)$ time since $|V_{BI}| = |V| + |E|$, $|E_{BI}| = \sum_{(U,W) \in E} (|U| + |W|)$ and $(|V_{BI}| + |E_{BI}|) \geq (|V_{DAG}| + |E_{DAG}|).

The time complexity of the EXTRACTFEATURE procedure (denoted as $O(G(\cdot))$) depends on the features to be extracted. Amongst the features we consider, closeness centrality has the highest time complexity ($O(|V|^3)$) (Table 7).

The time complexity of the TRAINMODEL procedure is dependent on the time complexities of the feature selection approach (denoted as $O(T(\cdot))$) and the training of the misclassification cost. Three feature selection approaches are explored. In BSE, a greedy approach is used for selecting a feature for removal at each iteration and a new SVM model is trained and tuned accordingly using the TUNESVM procedure. TUNESVM uses the grid search approach described in [8] to tune the SVM parameters. The tuning process takes $O(i^p \times k)$ where $i$, $p$ and $k$ are the number of iterations$^9$ required for the grid search, the number of parameters to be tuned, and the time complexity of training a SVM, respectively. According to [93], standard SVM training takes $O(m^3)$ time where $m$ is the training set size. Hence, TUNESVM has $O(i^p \times |V|^3)$ time complexity since the training set size is approximately equal to the data set size ($|V|$). In the worst case, algorithm BSE removes all but one feature. This takes $O(|X_{alt}|^2 \times i^p \times |V|^3)$ time where $X_{alt}$ is the set of topological features. In WRE, the statistical-based filter requires two steps, namely, WILCOXONFILTER and ROCFILTER to find the predictive feature set. Performing the Wilcoxon test for a particular topological feature requires $O((gh)^2)$ time [70] where $g$ and $h$ are the target and non-target class sizes, respectively, and $|V| = g + h$. Generating ROC for a particular topological feature requires $O(|V|^2)$ time [28]. Hence, WILCOXONFILTER and ROCFILTER require $O((gh)^2|X_{alt}|)$ and $O(|V|^2|X_{alt}|)$ time complexities, respectively. The intersection of the two sets of features generated by WILCOXONFILTER and ROCFILTER takes $O(|V|)$ time in the worst case [21]. Hence, WRE requires $O(|V|^2|X_{alt}| + i^p \times |V|^3)$ time. This can be further simplified to $O(i^p \times |V|^3)$ since $|V| > |X_{alt}|$ in most signaling networks. The time complexity of WRE-BSE is $O(|X_{alt}|^2 + i^p \times |V|^3)$, the maximum of the time complexities of BSE and WRE. The training of the misclassification cost takes $O((\frac{1}{\alpha} - 1)(i^p \times |V|^3))$. Hence, taken together, TRAINMODEL procedure has time complexity of $O(O(T(\cdot)) + (\frac{1}{\alpha} - 1)(i^p \times |V|^3))$.

---

$^9$The number of iterations $i$ needed for the grid search is dependent on the number of tuning level $l$, the range of the parameter to be searched $r$ and the step size during the search $s$. Formally, it is defined as $i = l \times \frac{r}{s}$. 

22
4 Results and discussion

**TENET** is implemented using Java. We shall now present the experiments conducted to study the performance of **TENET** and report some of the results here (additional results are given in Supplementary Material). 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. We characterize four signaling networks (referred to as individual networks) in **BioModels** (I₁ to I₄ in Table 8) and a combined network that is generated by iteratively performing a union of the nodes and edges in individual networks. The resulting combined network is a graph consisting of four disconnected¹⁰ subgraphs, each representing one individual network. For the combined network, we use each of the signaling network as the test set in turn (C₁ to C₄ in Table 8) and examine the effects of generating characterization models from individual networks and from the combined network. 

¹⁰The node and edge sets of the individual networks are disjoint.
in T{ENET} is performed on each individual network within the combined network. Section 3.2 describes the generation of the training and test data. We study different variants of T{ENET} (Table 9) by varying the SVM training approach.

### 4.1 Performance Metrics

We evaluate the performance of T{ENET} based on prediction accuracy\textsuperscript{11} ($\phi$), sensitivity (TPR), specificity (TNR) and precision (PPV) of the generated characterization models using the same training and test set. The definitions are as follows: $\phi = \frac{TP}{TP+FN}$, $TPR = \frac{TP}{TP+FN}$, $TNR = \frac{TN}{FP+TN}$ and $PPV = \frac{TP}{TP+FP}$ where TP, TN, FP and FN denote true positive, true negative, false positive and false negative prediction, respectively. Note that PPV is set to 0 when the classifier did not make any positive prediction. We include an additional metric feature reduction factor (FRF) to compare the performance of the feature selection methods. Formally, $FRF = 1 - \frac{|F|}{X_{all}}$ where $X_{all}$ is the entire set of features considered in the study. The performance of different characterization models is compared using an integrated performance score\textsuperscript{12} $\mathcal{P} = \sum_{m \in M} val_m$ where $M = \{\phi(val), \phi(test), TPR, TNR, PPV\}$ and $val_m$ is the value of metric $m$. Note that a larger score indicates better performance.

### 4.2 Kernel Selection

We experimented with several kernels: linear, radial basis function (RBF), sigmoid and polynomial. The parameters relevant to each kernel type and the ranges of these parameters that we tested are found in Tables 10 and 11, respectively. Figure 6 plots the results of T{ENET}-naïve (which considers all structural features)

\[ \text{Kernel} \quad | \quad \text{Formula} \quad | \quad \text{Parameters} \]

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Type</th>
<th>Range Tested</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C$</td>
<td>SVM parameter</td>
<td>$[2^{-12}, 2^{12}]$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>kernel parameter</td>
<td>$[2^{-12}, 2^{12}]$</td>
</tr>
<tr>
<td>$d$</td>
<td>kernel parameter</td>
<td>$[2-6]$</td>
</tr>
<tr>
<td>$c_0$</td>
<td>kernel parameter</td>
<td>$[2^{-12}, 2^{12}]$</td>
</tr>
</tbody>
</table>

Table 10: SVM kernel types and their associated parameters \textsuperscript{[66]}. Nodes marked with ♯ are known targets.
Figure 6: Performance of different kernels using the TENET-naïve approach.
Table 12: Best model parameters for the various signaling networks using the TENET-naive approach with different kernels. The parameters for the TENET-naive approach with linear kernel is found in Table 13.

<table>
<thead>
<tr>
<th>Kernels</th>
<th>Best model Parameters</th>
<th>I₁</th>
<th>I₂</th>
<th>I₃</th>
<th>I₄</th>
</tr>
</thead>
<tbody>
<tr>
<td>RBF Kernel</td>
<td>Best model C</td>
<td>2⁻¹⁰.⁸</td>
<td>2⁻¹⁰.⁸</td>
<td>2⁻¹⁰.⁸</td>
<td>2⁻¹⁰.⁸</td>
</tr>
<tr>
<td></td>
<td>γ</td>
<td>2⁻¹⁰.⁸</td>
<td>2⁻¹⁰.⁸</td>
<td>2⁻¹⁰.⁸</td>
<td>2⁻¹⁰.⁸</td>
</tr>
<tr>
<td>Sigmoid Kernel</td>
<td>Best model C</td>
<td>2⁻¹⁰.⁸</td>
<td>2⁻¹⁰.⁸</td>
<td>2⁻¹⁰.⁸</td>
<td>2⁻¹⁰.⁸</td>
</tr>
<tr>
<td></td>
<td>γ</td>
<td>2⁻¹⁰.⁸</td>
<td>2⁻¹⁰.⁸</td>
<td>2⁻¹⁰.⁸</td>
<td>2⁻¹⁰.⁸</td>
</tr>
<tr>
<td></td>
<td>C₀e₀f₀</td>
<td>2⁻¹⁰.⁸</td>
<td>2⁻¹⁰.⁸</td>
<td>2⁻¹⁰.⁸</td>
<td>2⁻¹⁰.⁸</td>
</tr>
<tr>
<td>Polynomial Kernel</td>
<td>Best model C</td>
<td>2⁻¹⁰.⁸</td>
<td>2⁻¹⁰.⁸</td>
<td>2⁻¹⁰.⁸</td>
<td>2⁻¹⁰.⁸</td>
</tr>
<tr>
<td></td>
<td>γ</td>
<td>2⁻¹⁰.⁸</td>
<td>2⁻¹⁰.⁸</td>
<td>2⁻¹⁰.⁸</td>
<td>2⁻¹⁰.⁸</td>
</tr>
<tr>
<td></td>
<td>C₀e₀f₀</td>
<td>2⁻¹⁰.⁸</td>
<td>2⁻¹⁰.⁸</td>
<td>2⁻¹⁰.⁸</td>
<td>2⁻¹⁰.⁸</td>
</tr>
<tr>
<td></td>
<td>Degree</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

using the various kernels. The choice of kernel did not affect the accuracy of the validation and the test sets. This implies that the training data is likely to be linearly separable. The execution time, however, is affected by the number of parameters involved in the kernels and the training set size [80] (size of network). Henceforth, we shall use the linear kernel for the rest of the experiments since it yielded the same accuracy as other kernels but is faster in terms of training speed. The parameters for the best models in this experiment is reported in Table 12. Note that in I₁, the sensitivity (TPR) and precision (PPV) are zero irrespective of the kernels. This highlights a need to use additional techniques (e.g., feature selection) to improve the characterization models. For subsequent experiments, we use the linear kernel as it yielded the same accuracy as other kernels but is faster to train.

4.3 Feature selection

First, we examine the performance of different feature selection approaches (TENET-B, TENET-R and TENET-H) and compare it with TENET-naive for different signaling networks. Note that in this set of experiments, we study the effect of the feature selection approaches in isolation. The effect of incorporating WMC into the SVM shall be investigated later. Table 14 reports the predictive feature sets for each network using different approaches. In total, 24 experiments were conducted since there are three feature selection methods and eight networks (I₁ to I₄ and C₁ to C₄). Amongst these 24 experiments, 25% of the predictive feature sets consist of only one feature while the remaining had multiple features (ranging from 4 to 15 features). This supports our previous observation [16] that multiple features result in better prediction of known targets. Observe that in Table 14, bridging centrality is not always in the predictive feature set (e.g., I₂). Figure 7 plots the performances of different feature selection approaches. We can make several observations. First, no single approach performs consistently well on all performance metrics. In fact, network topology plays an important role in feature selection. For instance, I₄ has
### Best model \( C \) parameter for the various signaling networks using various approaches with linear kernel.

<table>
<thead>
<tr>
<th>Data</th>
<th>TENET-B</th>
<th>TENET-R</th>
<th>TENET-H</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>( T ), ( \pi ), ( \theta ), ( \omega )</td>
<td>( \delta, \zeta, \beta, \gamma, \alpha )</td>
<td>( \delta, \zeta, \beta, \gamma, \alpha )</td>
</tr>
<tr>
<td>1</td>
<td>( R_{in} )</td>
<td>( \delta, \zeta, \beta, \gamma, \alpha )</td>
<td>( \delta, \zeta, \beta, \gamma, \alpha )</td>
</tr>
<tr>
<td>2</td>
<td>( \theta, \omega )</td>
<td>( \delta, \zeta, \beta, \gamma, \alpha )</td>
<td>( \delta, \zeta, \beta, \gamma, \alpha )</td>
</tr>
<tr>
<td>3</td>
<td>( \beta, \gamma, \alpha )</td>
<td>( \delta, \zeta, \beta, \gamma, \alpha )</td>
<td>( \delta, \zeta, \beta, \gamma, \alpha )</td>
</tr>
<tr>
<td>4</td>
<td>( \gamma, \alpha )</td>
<td>( \delta, \zeta, \beta, \gamma, \alpha )</td>
<td>( \delta, \zeta, \beta, \gamma, \alpha )</td>
</tr>
<tr>
<td>5</td>
<td>( \alpha )</td>
<td>( \delta, \zeta, \beta, \gamma, \alpha )</td>
<td>( \delta, \zeta, \beta, \gamma, \alpha )</td>
</tr>
<tr>
<td>6</td>
<td>( \beta, \gamma, \alpha )</td>
<td>( \delta, \zeta, \beta, \gamma, \alpha )</td>
<td>( \delta, \zeta, \beta, \gamma, \alpha )</td>
</tr>
<tr>
<td>7</td>
<td>( \gamma, \alpha )</td>
<td>( \delta, \zeta, \beta, \gamma, \alpha )</td>
<td>( \delta, \zeta, \beta, \gamma, \alpha )</td>
</tr>
<tr>
<td>8</td>
<td>( \alpha )</td>
<td>( \delta, \zeta, \beta, \gamma, \alpha )</td>
<td>( \delta, \zeta, \beta, \gamma, \alpha )</td>
</tr>
</tbody>
</table>

### Features selected by various feature selection approaches.

extremely high density of edges (ratio of edges to nodes) compared to other net-
Figure 7: Performance of different feature selection approaches.
works. The connectivity features of such networks become less informative and other features such as target downstream effect becomes more important. Hence, the most appropriate feature selection approach is dependent on the signaling network. However, we note that for larger sized networks, a larger number of features are informative (regardless of feature selection approach). This is perhaps because larger networks provide greater richness of context and diversity of structure in the sub-networks. Since network sizes are growing and network analysis demands applicability to larger networks, future methods might benefit particularly from the use of multiple features. Second, feature selection generally led to an improvement in prediction accuracy (87.5% for validation data set and 50% in test data set) over the naive approach. An exception is C4 in which feature selection resulted in poorer performance. In C4, the characterization model is generated using I1, I2 and I3 as training data whereas I4 is used as the test data. The characteristics of the known targets in the training data may be quite different from that of the test data. Indeed, from Table 14, we observe that bridging coefficient $\pi$ is included in the predictive topological feature set of C4, but not in I4. Including redundant features may lead to poorer performance. Third, the models generally have high specificity due to imbalanced data set. Fourth, TENET-R has the best runtime performance, followed by TENET-H and TENET-B. The poorer performance of TENET-B is due to the interaction of the feature selection approach with the classifier (classifier-aware approach) which is different from TENET-R where the feature selection approach is a wrapper layer that sits on top of the classifier. Finally, the size of the networks used for training affects the runtime performance. In general, larger size networks require longer runtime. In Section 4.7, we report TENET’s performance on the human cancer signaling network containing more than 2500 nodes.

4.4 Effect of varying WMC

Intuitively, when we vary the WMC, we expect that as the target misclassification cost $C^+$ increases, the prediction accuracy, sensitivity, specificity and precision would display a negative skewed, increasing, decreasing and positive skewed distribution, respectively. This is because a large $C^+$ eventually results in a model that is likely biased towards classifying data as targets. The effect of varying the WMC are reflected in Figures 8 to 15. From the figures, we noted the following trends. First, amongst the individual networks, only I3 (Figure 10) displays the expected trends. This could be due to the extreme small target size (1 or 2) in the test set that resulted in extreme fluctuations in the performance metrics and deviation from the expected trends. Hence, the target size of the test set can have significant impact on the observed results. Second, the performance of the combined networks C1 (Figure 12), C2 (Figure 13) and C4 (Figure 15) resembles that of I3, possibly due to the large size of I3 dominating over other networks used for training. This implies large training networks can have undue influence on the characterization model. Third, sensitivity generally improves whereas specificity generally deteriorates when the target misclassification cost is set higher than the non-target
Figure 8: Performance of TENET variants incorporating feature selection approach and WMC for the MAPK-P13K network.

Figure 9: Performance of TENET variants incorporating feature selection approach and WMC for the glucose-stimulated insulin secretion network.
misclassification cost ($C^+ > C^-$). The choice of an appropriate model depends on the application. Fourth, the prediction accuracy tends to display a skewed distribution where accuracy initially increases (or remains constant) with increasing $C^+$,
and then decreases with increasing $C^+$. Fifth, individual networks and combined networks behave differently. In individual networks, prediction accuracy, sensitivity and precision generally improve when $C^+$ is set larger than $C^-$. However, in combined networks, sensitivity improves whereas other performance criteria deteriorates when $C^+$ is set larger than $C^-$. Hence, there is no single universal best value of $C^+$ and the choice of $C^+$ depends on the network.

### 4.5 Best TENET variant

We identify the best TENET variant (Table 15) using the integrated performance score $\mathcal{P}$. We note the following. First, the best TENET variant is network dependent. Second, variants incorporating both WMC and feature selection generally perform well. Specifically, setting $C^+$ greater than $C^-$ led to better results. Third, TENET variants based on individual networks ($I_1$ to $I_4$) outperforms that based on combined networks ($C_1$ to $C_4$). The poorer performance of the combined networks may be due to insufficient number of training networks, inappropriate or insufficient features used for training or that signaling networks by nature have distinct characteristics and it is just not possible to have a generalized model. Finally, the predictive topological features differ across networks (Tables 14 and 15). Hence, as we mentioned in Section 1, a single set of predictive topological features may not effectively characterize known targets in all signaling networks. When we compare the results with that in our previous work, we note that the set of predictive topological features are different from the discriminative topological features.
Figure 13: Performance of TENET variants incorporating feature selection approach and WMC for $C_2$.

Figure 14: Performance of TENET variants incorporating feature selection approach and WMC for $C_3$. 
Figure 15: Performance of TENET variants incorporating feature selection approach and WMC for $C_4$.

Table 15: Summary of best TENET variant for different networks. $C^+$ values are provided in bracket besides approaches using WMC. $\Delta x = \frac{x_{\text{best}} - x_{\text{naive}}}{x_{\text{naive}}}$, where $x_{\text{best}}$ and $x_{\text{naive}}$ are the values of performance metric $x$ of the best TENET variant and TENET-naive, respectively. * marks instances where $x_{\text{naive}} = 0$ and ^ marks the best models selected for generating the characterization model.

(DTF) identified in [16] although there was an overlap of at least 50% of the features\textsuperscript{13}. The difference is due to the different approach used to identify the features. The characterization models\textsuperscript{14} generated by these DTFs also yielded poorer average ROC (0.873) than that generated using TENET (0.913) (Approach DIFFER in Figure 16).

\textsuperscript{13}We consider only $I_1$ to $I_3$ and exclude $I_4$ from this comparison as no DTF was found at $p$-value less than 0.05.

\textsuperscript{14}We use SVM with WMC and WRE to generate the characterization models.
In this subsection, we describe the experiments that compare TENET against other state-of-the-art approaches. We perform three sets of experiments for comparison with network-unaware techniques, PPI network-based techniques and network-aware target prioritization approaches. Recall that state-of-the-art techniques such as [41, 65, 108] focus on PPI networks instead of signaling networks. To the best of our knowledge, there does not exist any target characterization technique for signaling networks. However, one way to investigate the performance of TENET is to examine how well the characterization model generated by it prioritizes known targets. Intuitively, target prioritization aims to rank the nodes according to their

---

**Figure 16:** Performance of different prioritization approaches.
Table 16: Summary of targets obtained from review articles. Targets that are present in the network and in TENET top-10 ranked nodes are marked as * and underlined, respectively.

<table>
<thead>
<tr>
<th>Network</th>
<th>Targets in review article</th>
<th>Total targets</th>
<th>% of TENET top-10 ranked nodes overlapping with review targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>I₁</td>
<td>BAB25, PKRC1, EVI1, PIK3CA, POF1, MYC, PIK3R1, AKT12, AURKA, Kras², Ras², CTNNB1, CDKN2A, APC, KIT, SMAD4, IGBP2, SAT2, ARH1, PEG3, PLAG1, RPS6KA2, TP53, BRCA1, BRCA2, PTEN, OPCLM1, WWOX, DAPK1, CDH13, MLH1, ICA1, DNASC15, MUC2, CDKN1A, RASSF1, SOCS1, SOCS2, Pycard, Snf1</td>
<td>41</td>
<td>50%</td>
</tr>
<tr>
<td>I₂</td>
<td>GLP-1, GLP-1 receptor, DPP-4, NAP-24.11, SGLT, amylin, PPAR, ATP-sensitive potassium channel, α-glucosidase, glucokinase, AMP kinase, carnitine palmitoyltransferase-1, glycogen synthase kinase-3, PTP-18, pyruvate dehydrogenase, fructose-1,6-bisphosphatase, 11β-hydroxysteroid dehydrogenase 1, siruin 1, acyl-CoA-diacylglycerol acyltransferase 1, phosphoenolpyruvate carboxykinase, glucose-6-phosphatase, PPARγ coactivator 1α, acetyl-CoA carboxylase, mitochondrial rotenone-sensitive NADH:ubiquitine oxidoreductase (complex I), leptin, ghrelin, resistin, C-peptide, protein kinase C, AGE, RAGE, glutamine:fructose-6-phosphate, PARP, VEGF, aldose reductase, vitamin C, vitamin E, GPR40, GPR119, GPR41, GPR43, GPR120, GPR109A, dopamine-2 receptor, m3 subtype muscarinic receptor, 5-hydroxytryptamine 2c subtype serotonin receptor, imidazoline, glucagon receptor, retinoid X receptor, colesalvalam, IL-1β, chemokine receptor 2, angiotensin receptor, thioredoxin-interacting protein, Kv1.1 channel, PBR21, w-3 PUFAs, ZnT8, diacylglycerol acyl transferase 1</td>
<td>60</td>
<td>50%</td>
</tr>
<tr>
<td>I₃</td>
<td>csk-3, frizzled, n-H-tcr², RhesC, Wnt8, Hox11/13b, Su(N)², Blimp1², Otx², Bra², Foxa², Gata6², Notch²</td>
<td>14</td>
<td>90%</td>
</tr>
<tr>
<td>I₄</td>
<td>sck, pta, acs², perxa², pykA, pykF², fadR, ppc², pyc, zwf, ztl², galP, glucokinase², glucose²</td>
<td>14</td>
<td>50%</td>
</tr>
</tbody>
</table>

Potential of being a target based on some importance measures (e.g., gene expression level [14]). In the following, we first describe the comparisons with network-unaware techniques, then that with PPI networks and finally that with the network-aware target prioritization approaches.

### 4.6.1 Comparison with network-unaware approaches

We compared TENET’s prediction against those derived from non-network-based approaches, specifically, targets that were predicted by various experimental techniques and consolidated within review articles. The targets for I₁, I₂, I₃ and I₄ were derived from [3], [97], [60] and [87], respectively. Table 16 summarizes the targets found in the review articles (referred to as review targets). In general, there is an overlap between the targets in the network and those in the reviews. Note that in signaling networks, the same gene and protein often exist in multiple forms and such representations may be manifested in the top-10 ranked nodes. For instance,
in TENET top-10 ranks of \( I_3 \), there are multiple versions of \( n_\beta^{-TCF} \)\(^{15} \). When we take this multiple forms into consideration, the percentage of TENET top-10 ranks overlapping with review targets are 50\%, 50\%, 90\% and 50\% for \( I_1 \), \( I_2 \), \( I_3 \) and \( I_4 \), respectively. Next, we examine if the remaining targets are biologically relevant by checking for correspondence with our curated targets in Section 2.3. Only three targets\(^{16} \) do not correlate with curated targets. This implies that top ranking nodes in TENET correlate well with existing biological knowledge. That is, there is good correspondence between TENET top-10 ranked nodes and existing biological knowledge. Note that TENET also out-performs other state-of-the-art approaches in terms of the overlap achieved between the top-10 ranked nodes and the review targets and curated targets (Figure 17).

\( ^{15} \)\( n_\beta^{-TCF} \) is present as protein P \( n_\beta^{-TCF} \), protein M \( n_\beta^{-TCF} \) and protein \( n_\beta^{-TCF} \) which represent \( n_\beta^{-TCF} \) protein form in \( PMC \), mesoderm and endoderm cells, respectively.

\( ^{16} \)PP2A in \( I_1 \) and ferricytochrome c, dihydroxyacetone-phosphate and succinyl-CoA from \( I_2 \)
RefSeq or UniProt identifiers. We annotate the nodes in the four networks using UniProt and Entrez identifiers when majority of the nodes are proteins and genes, respectively. For clarity, networks I₁, I₂ and I₄ are annotated using UniProt whereas I₃ is annotated using Entrez. We follow the following rules during annotations:

1. When multiple annotation are available, select the one with organism matching that of the given network.

2. When a node does not have a valid annotation, it inherits an annotation (where available) related to an edge that the node is associated with. For example, glucose do not have a corresponding UniProt identifier. It is involved in a reaction where glucokinase (UniProt ID=P52792) catalyzes glucose to glucose-6-phosphate. Hence, it shall inherit the UniProt identifier of glucokinase.

3. When a node is a result of post-translational modification (e.g., phosphorylation), it inherits the annotations of the original node. For example, phosphorylated ERK shall have the same annotations as unphosphorylated ERK.

ToppNet, which prioritizes nodes based on functional annotations, returns no results for all four networks and is excluded from Figure 16. This could imply that the database in ToppNet are lacking in functional annotations related to these networks. Unlike ToppNet whose analysis is reliant on the quality of its functional annotations database, TENET analysis depends only on the structure of the network which is inherent in the signaling network given by the user.

ProphNet provides several prioritization features. In particular, we are interested in prioritization of a given set of nodes and prioritization of a ProphNet-generated node set for a given disease and we refer to them as ProphNet A and ProphNet B, respectively. ProphNet A returns no results for I₁ to I₄ whereas ProphNet B returns a prioritized ProphNet-generated node set for I₁ and I₂ when “ovarian fibromata” and “diabetes mellitus, insulin-dependent, 2” were given as the input disease, respectively. Note that for I₃ and I₄, we were not able to find a related disease tag in ProphNet B for a meaningful query. The ProphNet-generated and prioritized node sets (referred to as ProphNet nodes) where mapped to the nodes in I₁ and I₂. The mapping is performed according to the following rules:

1. When a node does not have a clear, unambiguous 1-1 ProphNet node mapping, it is mapped to a related ProphNet node. For example, gs(Grb2\-SOS) is mapped onto ProphNet nodes GRB2, SOS1 and SOS2 and inherit the ProphNet score of these nodes.

2. When multiple ProphNet node versions are available, all ProphNet nodes are mapped to the network node. For example, different ProphNet nodes

---

17 ToppNet database contain human and mouse genes and uses GO as functional annotations whereas I₁ to I₄ are networks related to mouse (I₁ and I₂), sea urchin (I₃) and E. Coli (I₄).
such as PPP2R5E, PPP2R3A and PPP2R2B are used to represent different components of PPP2A. All of these ProphNet nodes and their values are mapped to PPP2A.

This mapping resulted in some nodes being mapped to one or more ProphNet nodes. Ambiguity in the node prioritization occurs when a node is mapped to multiple ProphNet nodes. We resolve this ambiguity by generating two new set of prioritization ranks called ProphNet(Max) and ProphNet(Sum). In the former rank, a node will be assigned the highest ProphNet ranks among the mapped ProphNet nodes. In the latter rank, a node will be given a score that is the sum of the ProphNet values for all the mapped ProphNet nodes. Then, the nodes are ranked in decreasing order of this score. Note that node mappings have to be performed either during annotations for ToppNet or when comparing the prioritized ProphNet nodes. Ambiguity arise when the signaling networks contain the same protein or gene in different forms or cells as different nodes whereas the tools expect each node to represent a different protein or gene. Hence, a fair comparison between TENET and these two approaches becomes difficult. Note that in both I₁ and I₂ (Figure 16), TENET outperforms ProphNet in terms of ROC AUC and AUPR. For execution time, TENET also outperforms ProphNet when training was performed offline. Hence, in subsequent experiments, we shall focus on comparing TENET and other signaling network-based approaches (random prioritization, DIFFER, LSA and NetworkPrioritizer).

### 4.6.3 Comparison with network-aware target prioritization approaches

Target prioritization is the process of ranking nodes in a network according to their likelihood of being a target based on some criteria (e.g., sensitivity, gene expression level, score generated by a characterization model). That is, given a signaling network \( G = (V, E) \), the target prioritization problem assigns a target rank \( r_u \) for each node \( u \in V \). Given \( u, v \in V \), \( u \) is more likely than \( v \) to be a target if \( r_u < r_v \). It is potentially useful in helping to plan experiments since resources are limited and experiments can be costly and time-intensive. This is especially true in drug development [68]. Note that target characterization do not generate “new” targets, but instead produces a model that characterizes known targets. In contrast, target prioritization may generate “new” targets by virtue of the fact that high ranking nodes that are not in the set of known targets have greater potential to be “new” targets. Characterization models (described in Section 3.1) generated by TENET can be used for generating prioritization scores to rank nodes in a signaling network. Recall that TENET generates characterization models using an approach based on support vector machines (SVM). SVM (e.g., \( \epsilon \)-support vector regression) can yield models that produce a continuous outcome (i.e., regression scores) instead of discrete outcome (i.e., classes). Hence, the regression scores can then be used for prioritizing nodes. Figure 18 depicts the overview of the target prioritization process using the output of TENET. Specifically, TENET produces a characterization model when given a signaling network (e.g., MAPK/PI3K network),
a relevant disease node (e.g., ERKPP) and a list of known targets (e.g., Akt, Raf, etc.) as inputs. This characterization model can then be applied on any signaling network $G$ to obtain a regression score for each node in $G$. Finally, the nodes in $G$ are ranked in decreasing order of the regression score. Nodes that are ranked top but not in the set of known curated targets may be potential “new” targets.

For our study, we compare TENET with several network-aware target prioritization approaches, namely, random prioritization, LSA [33] and NetworkPrioritizer [47]. In random prioritization, the nodes were randomly assigned a rank in the range $[1, V]$ where $V$ is the number of nodes in the network and we assume that no ranking ties are present. LSA was performed using Copasi [85] with the following configuration: \{task=sensitivities; subtask=time series; function=all variables of the model; and variable=all parameter values\}. We consider both Weighted Borda Fuse (WBF) and Weighted AddScore Fuse (WASF) in NetworkPrioritizer and consider all features provided. Note that uniform weights were used for rank aggregation since we do not have prior knowledge of the best weights or features to consider. For TENET, we use the characterization model to generate prioritization ranks of known targets. Specifically, we apply the SVM models to obtain these ranks. The SVM type is set to $\epsilon$-support vector regression ($\epsilon$-SVR) with default $\epsilon$ value ($1 \times 10^{-3}$) and the SVM parameters are set according to the best models for each network (Tables 13, 15 and 17). Note that the nodes are ranked in decreasing order of the regression score and higher ranked nodes are more likely to be targets.

First, the experimental results reveal that the normalized ranks of a given node vary widely using different approaches. Figures 19, 20 and 21 plot the normalized

\[18\] In $\epsilon$-SVR, the error function is an $\epsilon$-insensitive loss function and error smaller than $\epsilon$ is ignored [8].
Figure 20: Normalized rank of nodes in test set of I₂.

Figure 21: Normalized rank of nodes in test set of I₄.

Figure 22: Normalized rank of top 10 nodes ranked using TENET in test set of I₅.


<table>
<thead>
<tr>
<th>Best model C</th>
<th>C₁</th>
<th>C₂</th>
<th>C₃</th>
<th>C₄</th>
</tr>
</thead>
<tbody>
<tr>
<td>TENET-W</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC Ratio ID=1</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=2</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=3</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=4</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=5 (TENET-naive)</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=6</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=7</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=8</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=9</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>TENET-WB</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC Ratio ID=1</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=2</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=3</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=4</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=5 (TENET-B)</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=6</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=7</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=8</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=9</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>TENET-WR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC Ratio ID=1</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=2</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=3</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=4</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=5 (TENET-W)</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=6</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=7</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=8</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=9</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>TENET-WR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMC Ratio ID=1</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=2</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=3</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=4</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=5 (TENET-W)</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=6</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=7</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=8</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
<tr>
<td>WMC Ratio ID=9</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
<td>2⁻¹⁰</td>
</tr>
</tbody>
</table>

Table 17: Best model C parameter for the various combined signaling network variants (C₁ to C₄) using different approaches with linear kernel.

The normalized rank of a node u for a particular approach x is denoted as Ψₙ₉₉(x):u and defined as follows.

\[
Ψₙ₉₉(x):u = \frac{Ψₓ:u}{MAX_{i∈V}(Ψₓ:i)}
\]

where Ψₓ:u is the rank of u based on x, V is the set of nodes in the given signaling network and MAX(·) is the maximum operator. We use the normalized rank for comparison since the range of ranks for each approach is different. In these figures,
we mark the known targets in the figures using (*) for ease of reading. The DIFFER approach refers to the method used to find discriminative topological features (DTF) in our previous work [16]. Specifically, we perform SVM using WMC and WRE. We use the WRE feature selection approach to ensure that the DTFs are used specifically for training the SVM and tested the entire range of WMC (i.e., [0–1]). The characterization model with the best performance score is then used to generate the prioritization ranks of DIFFER. Table 18 provides a summary of the best characterization models of DIFFER. Compared to TENET, the performance scores of DIFFER’s characterization models are consistently lower. Note that an ideal ranking approach should assign higher normalized ranks, represented by shorter histogram, to known targets. From these figures, we note that normalized ranks of a given node can vary widely using different prioritization approaches. For instance, in Figure 24, the normalized rank of G3P, a known target, is relatively high when prioritized using TENET, random prioritization and LSA. However, it is given relatively low ranking by WBF and WASF. In another network C1 (Figure 23), WBF and WASF assigned higher ranks to Raf(*), another known target, compared to TENET and random prioritization. Hence, an approach that performs better for one particular network can perform poorly in another.

Table 18: Summary of best DIFFER characterization model for different networks.

<table>
<thead>
<tr>
<th>SVM cost parameter C</th>
<th>I1</th>
<th>I2</th>
<th>I3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>2^6</td>
<td>2^-10</td>
<td>2^12</td>
</tr>
</tbody>
</table>

Figure 23: Normalized rank of nodes in test set of C1.
Figure 24: Normalized rank of nodes in test set of C₂.

Figure 25: Normalized rank of top 15 nodes ranked using TENET in test set of C₃. *none* is the node used to represent degraded proteins in this network.

Figure 26: Normalized rank of nodes in test set of C₄.
Next, we discuss the ability of each method to predict known targets in the test set of two networks ($I_1$ and $I_4$) in particular. For a given network $G = (V, E)$, containing a set of known targets $T \subset V$, an ideal target prioritization approach should rank all targets higher\(^\text{19}\) than non-targets. That is, the set of targets should be contained within the top-$\left(\frac{|T|}{|V|} \times 100\right)$% ranks of the network. For clarity, a target $t \in V$ is correctly predicted if it is given a rank within the top-$\left(\frac{|T|}{|V|} \times 100\right)$%. In the MAPK/PI3K network ($I_1$), all approaches except LSA\(^\text{20}\) correctly predicted the known target (MEKPP) (Figure 19). In another network (glucose metabolism network, $I_4$) (Figure 21), TENET performed the best as it correctly predicted the two known targets (EIIA and isocitrate). In contrast, random prioritization and WASF correctly predicted EIIA but missed isocitrate (ranked third and fourth in random prioritization and WASF, respectively). The remaining approaches performed worse and missed both targets. LSA and WBF ranked EIIA and isocitrate as second and fifth, respectively whereas DIFFER did not yield any results as no discriminative topological features (DTF) were found for this network. Note that the ROC AUC is generally used to compare the performance of different characterization models and the results of the ROC AUC can be found in Figure 16. From this figure, we observe that TENET outperforms other approaches in terms of the quality of the prioritization results, particularly for individual networks, and is comparable in terms of runtime performance when SVM training is performed offline (TENET (Regression only)).

4.7 Scalability

Although existing signaling networks tend to be small (tens to hundreds of nodes) in size, they are expected to grow. We tested TENET to assess its scalability to larger networks. The largest curated signaling network we obtain from Biomodels.net was $I_3$, endomesoderm gene regulatory network with over 600 nodes (previous experiments). We further tested TENET on a larger network with 2635 nodes and 43735 edges. This network is obtained from a human signaling network [58] curated by Edwin and colleagues and is implicated in cancer. The targets (324 nodes, 12.3% of dataset) in this network that are deemed relevant to cancer are reported in [17]. We performed 10-fold cross validations and the test set (239 nodes) contains 27 targets (11.3% of test set). TENET took 2 days for the analysis using the following configuration: C parameter range=$[2^{-8} - 2^8]$, libsvm shrinking parameter=1, libsvm SVM type=NU_SVC, $C^+$=0.7. Note that we use this SVM configuration to avoid numeric instability when performing SVM training for the larger network. The weighted misclassification cost was set to 0.7 as larger values in the range of [0.6 - 0.8] were found to perform better (Table 5 in main text). The best SVM model has C parameter $2^8$ and has associated predictive features $X_{all} \setminus \theta_{in}$. The need for multiple predictive features may be due to the presence

\(^{19}\) Note that in real situations, it is unlikely for us to know the exact ranks within the set of targets.

\(^{20}\) In LSA, MEKPP is ranked third out of four nodes in the test set.
Approach | ROC AUC | AUPR  
--- | --- | ---  
**TENET** | 0.718 | 0.251  
Random | 0.54 | 0.127  
DIFFER | 0.627 | 0.215  

Table 19: Comparison of different approaches on a large network.

of different types of cancer targets (e.g., different hallmarks) which may be characterized differently. Hence, we performed further experiments by classifying the targets based on 8 different hallmarks suggested in [35] and characterizing each hallmark category. The hallmark targets were curated using information from the OMIM database [34]. The results from hallmark-based characterization are similar to our previous results. That is, predictive features remain as $x_{all} \setminus \theta_{in}$ regardless of hallmarks. This further highlights the need of using multiple predictive features to characterize targets and a single feature (e.g., bridging centrality) may not be effective. In addition, we note that certain topological features (e.g., closeness centrality) we study requires the computation of the shortest path which has $O(|V|^3)$ time complexity using Floyd-Warshall algorithm [29] where $|V|$ is the size of the network. Hence, this may impose an upper limit on the size of the network that TENET can handle. We can address this limitation by extending TENET with techniques that estimate shortest path for large networks such as [81] and [110].

Finally, we compared TENET to other state-of-the-art signaling network-based approaches. Note that comparison could not be performed on LSA and NetworkPrioritizer. For the former, no dynamic information of the large cancer network was available whereas for the latter, the program did not complete the analysis due to memory issues. We observe that TENET (ROC AUC=0.718 and AUPR=0.251) outperforms other approaches in terms of ROC AUC and AUPR (Table 19).

5 Conclusions & Future Work

We propose TENET, a SVM-based approach that characterizes known targets in signaling networks using topological features by identifying a set of predictive topological features and using them to generate a characterization model. TENET uses feature selection to remove redundant features, thereby improving prediction accuracy of the characterization models and WMC to improve other performance criteria (e.g., sensitivity). Our empirical study reveals that the characterization models generated by TENET outperforms state-of-the-art approaches in prioritizing signaling and PPI networks. In summary, the contribution of this work is a machine learning-based framework that affords flexibility in characterizing signaling networks of different sizes and with different number of known targets. Although TENET is evaluated on a small\textsuperscript{21} number of signaling networks, it can easily incorporate additional signaling networks without any modification to the framework. As part of future work, we intend to explore how the characterization models learnt by

\textsuperscript{21}Manual target curation, a time-intensive process, is needed to identify known targets of signaling networks for validating our experimental results.
TENET can be leveraged for target prioritization of signaling networks with *unknown* targets.
References


[77] Phue, J.-N. et al. (2004). Transcription levels of key metabolic genes are the cause for different glucose utilization pathways in *E. coli* B (BL21) and *E. coli* K (JM109). *J. Biotechnol.,* 109(1-2), 21-30.


