Modeling Cell Communication with Time-Dependent Signaling Hypergraphs

Michael R. Schwob*, Justin Zhan**, and Aeren Dempsey*
University of Nevada, Las Vegas*
University of Arkansas**
schwom1@unlv.nevada.edu, jzhan@uark.edu, dempsa1@unlv.nevada.edu

***Abstract*—Signaling pathways describe a group of molecules in a cell that collaborate to control one or more cell functions, such as cell division or cell death. The pathways communicate by sending signals between molecules, and this process is repeated until the terminal molecule is activated and the cell function is executed. Signaling pathways are often represented as directed graphs, which does not provide enough information when modeling cell functions and reactions. Recently, directed hypergraphs have been proposed to more accurately represent reactions such as protein activation and interaction. To further improve the representation of signaling pathways, time dependency must be considered to improve the representation of cell signaling at any given time. In this paper, the importance of time dependency in modeling signaling pathways is presented. An algorithm that finds the shortest *a priori* path using time-dependent hypergraphs to more robustly model signaling pathways is adopted. The shortest time-dependent hyperpaths representing signaling pathways are an improvement to the recent adoption of hypergraphs representing these pathways. The results display the improved representation of signaling pathways and motivate the adoption of time-dependent signaling hypergraphs.**

***Index Terms*—Hypergraphs, systems biology, signaling pathways, time dependency**

I. Introduction

A group of molecules in a cell work together to control cell functions, such as cell division or cell death. This communication occurs in a signaling pathway. Once a molecule in the pathway receives a signal, it activates another molecule. This process is repeated until the cell function is executed. Within signaling pathways, many reactions occur including the activation or deactivation of proteins and complexes. The representation of these biological pathways must consider the complexity and time dependency of functions and reactions. Recent representation methods have modeled signaling pathways as hypergraphs [1]. A hypergraph is a graph that can join any number of vertices, or nodes, with a single hyperedge. Since a hypergraph can contain multiple nodes on both the tail and head of a hyperedge, hypergraphs can model complex networks where multiple nodes interact at once. For example, if multiple proteins interact simultaneously, each node representing individual proteins can be represented by a hypernode. The adoption of hypergraph theory in cell signaling provides a more accurate alternative in pathway analysis than directed and undirected graphs due to the complex reactions that occur between proteins and molecules. Hypergraphs have been used to confront questions of pathway reconstruction, enrichment, and crosstalk [2].

There is one clear advantage to representing signaling pathways with hypergraphs over graphs: convenience. A hypergraph can model multiple reactions occurring simultaneously with a single hyperedge, whereas a regular graph may have to model the complex reactions with multiple edges connecting existing and intermediary nodes. When multiple biological resources interact at once, hypergraphs provide a more convenient representation of signaling pathways than directed and undirected graphs. However, neither approach accounts for the variability of time within the biological system. Thus, time-dependent hypergraphs can more robustly model signaling pathways. The efficiency of reactions and executed functions occurring within signaling pathways changes throughout the course of hours, days, and years [3]. Time-dependent hypergraphs can improve the robustness of pathway analysis by considering the current state of signaling pathways and accounting for available resources. The human body experiences changes in reactions and efficiency that are dependent on sleep, meals, puberty, age, etc. Many biological experiences will affect cell signaling and must be considered when analyzing signaling pathways.

Most applications of time-dependent hypergraphs use a time-adaptive strategy or an *a priori* strategy [4]. Within the time-adaptive strategy, the hyperpath adjusts to new information obtained while the process is executing. An *a priori* strategy is a predetermined process, and the hyperpath does not change *en route*. Since the signaling pathway receives a signal and the molecular reactions execute until termination at transcription factors, the function and reactions are predetermined. However, since molecules degrade during the signaling process, the abundance of resources may change during longer periods of time. For larger timescales, a time-adaptive strategy would be most appropriate. However, an *a priori* approach is sufficient for typical cell communication due to the relatively small timescale of intracellular communication [5]. Generally, the most appropriate strategy to model signaling pathways will be an *a priori* strategy.

Three contributions are provided in this paper. First, time-dependent signaling hypergraphs are defined. Second, a pre-existing algorithm is modified to obtain a method for computing the shortest time-dependent hyperpath. The original algorithm computes the complete shortest acyclic hyperpath. The complete shortest acyclic hyperpath would only occur given every resource is present in the signaling pathway. However, this is unlikely to occur, so the proposed algorithm computes

the shortest time-dependent hyperpath, which considers the availability of biological resources. Although atemporal data is used in the experiment, the presence or absence of a protein is dependent upon the state of the signaling pathway. Since the pathway's state is dependent upon time, the presence of a protein is also dependent upon time. Third, it is shown that pathway analysis becomes more robust when considering available proteins and complexes within signaling pathways, which are time-dependent. If time is not considered when modeling cell communication, then the model will compute the shortest signaling hyperpath assuming all resources are available. Since many biological processes affect the availability of proteins, then a model that assumes all proteins will be available is not as robust as a model that adjusts to the availability of the pathway's resources. Because biological processes are dependent upon time, a model that considers time can adjust to the availability of resources and, therefore, provide more robust results than a model that does not consider time.

In Section II, related research to pathway analysis with signaling time-dependent hypergraphs is discussed. The discussed research will be related to time-dependent hypergraph theory and algorithms as well as current methods of modeling signaling pathways. In Section III, signaling hypergraphs, hyperpaths, shortest hyperpaths, time-dependent hypergraphs, and signaling time-dependent hypergraphs are defined. In Section IV, an approach to improve the robustness of signaling pathway analysis is discussed. Additionally, several properties and lemmas will be provided to support the proposed method. In Section V, an algorithm modified from pre-existing algorithms calculating the shortest signaling hyperpath and the shortest time-dependent hyperpath is provided. In Section VI, the experiment set-up, the datasets, and the experimental results are discussed. In Section VII, a conclusion discusses the findings and future work possibilities are presented.

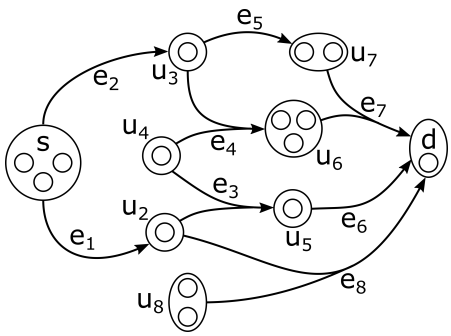

Fig. 1. Hypernodes and hyperedges connect a starting hypernode (s) and a destination hypernode (d). Within this hypergraph, there are multiple hyperpaths from s to d: (1) $(s, e_1, u_2/u_4, e_3, u_5, e_6, d)$, (2) $(s, e_2, u_3/u_4, e_4/e_5, u_6/u_7, e_7, d)$, and (3) $(s, e_1, u_2/u_8, e_8, d)$. Of these three hyperpaths, the third path is the shortest. Suppose u_8 is not available in the signaling pathway for cell communication. Then, the third hyperpath cannot complete an s-d path. Either the first or second hyperpath will become the shortest acyclic hyperpath.

II. Related Research

The researched topics of signaling hypergraphs, time-dependent factors in biological systems, time-dependent hypergraphs, a time-adaptive strategy, and an *a priori* strategy are discussed. The strategies are in context of finding the shortest hyperpath. The strategy that best models signaling pathways is decided, and the adapted algorithms are explained.

Signaling hypergraphs have been proposed to model signaling pathways [2] [6]. Within signaling pathways, several reactions cannot be accurately modeled with directed graphs due to the complex nature of these biological functions. Directed hypergraphs can better characterize reactions that involve multiple complexes and proteins. Some of these reactions include complex assembly and dissociation, protein activation and inactivation, and combinatorial regulation. Signaling hypergraphs include more information in hyperpaths than directed graph representation, since they can represent more than one reaction at any given time instance. There are efforts to provide the most accurate weighted hypergraph to model signaling pathways [7]. Hypergraph-based learning algorithms have been proposed to classify gene expressions [8]. There has been a recent adoption of computational methods to systematically model the behavior of signaling pathways due to the growing availability of experimental cell biology data [9]. Informative proposed computational methods include hypergraphs and shortest hyperpaths. Algorithms have been proposed to calculate the shortest hyperpath using the available data [6]. Properties of hypernetworks of biological signaling pathways have been researched to improve pre-existing algorithms [10]. The modified algorithm is based on the latest proposed shortest signaling hyperpath algorithm [1].

Biological systems change over time. TGF-β signaling pathways develop in embryonic development. During this stage of development, disease processes that require TGF-β ligands vary in efficiency [11]. Fibroblast growth factors (FGF) are secreted molecules in embryonic development. FGF secretion is also dependent on time [12] [13]. Intracellular signaling often has time and concentration-dependent activities. Epidermal growth factors (EGF) and Histidine-rich glycoproteins (HRG) induce time-dependent gene expressions, resulting in distinct cellular phenotypes in MCF-7 cells, an isolated breast cancer cell. This EGF and HRG-induced communication caused the ligand-oriented biphasic induction of protein after twenty minutes [3]. Wnt signaling is involved in embryonic development and controls homeostatic self-renewal in several adult tissues. Mutations in this signaling pathway cause several hereditary diseases and are associated with intestinal cancer as well as cancer in other tissues. The signal transduction of Wnt pathways varies in efficiency depending on many factors, which includes age and available energy [14]. Mammary gland development occurs primarily after birth, controlled by steroid and peptide hormones. The amount of hormones created depends on the efficiency of the signaling pathways utilized by the mammary glands [15] [16]. The ATF6 and IRE1-XBP1 signaling pathways are important for the refolding process of endoplasmic reticulum (ER). The degradation of misfolded glycoprotein substrates require transcriptional induction that is mediated by IRE1-XBP1. The analysis of refolding ER has revealed a time-dependent transition in signaling pathways [17]. Other research papers suggest that Toll-like receptors [18], protein-signaling networks [19],

neurotrophic factors [20], antigen-specific immune responses [21], regulatory networks [22], and mood modulation in the Kinase signaling pathway are time-dependent and dynamic systems [23]. Signaling pathways have many time-dependent functions and reactions.

Time-dependent hypergraphs have been utilized in several disciplines. Most recent progress made in calculating the shortest time-dependent hyperpath has been in transportation networks [24] [25] [26] [27]. Several algorithms have been proposed to calculate the most reliable and robust time-dependent hyperpath in the context of transportation [28] [29] [30] [31]. These algorithms are based on more generalized algorithms that find the shortest stochastic hyperpath [32] [33] [34] or stochastic hyperpath [4] [35] [36] [37]. Although graph parameters have been proposed, such as the hydra number, more recent shortest time-dependent hyperpath algorithms will be adapted for this paper.

Several strategies exist for computing the shortest time-dependent hyperpath. Frequently adopted strategies include the time-adaptive strategy and *a priori* strategy.

A time-adaptive strategy in calculating the shortest time-dependent hyperpath is commonly used in transportation networks and navigation applications [28] [37] [38]. As a vehicle is *en route*, the navigation application collects data and distributes this data across its users. This dispersion of data provides a continuous update of information to users. While *en route*, the navigation application may discover a new shortest path, changing the course of navigation. In the context of transportation, an *a priori* strategy determines the shortest path before the driver leaves his/her initial position. The route has been determined, and the driver will not change his/her course.

Regarding signaling pathways, a time-adaptive strategy implies that functions may change as reactions are executed. An *a priori* strategy would mean that the biological function is chosen before the reactions begin, and the function continues until either executed or failed. Given that cell communication executes specific functions, where the functions are themselves time-dependent, the signaling pathways would be communicating a predetermined function. Although the communication may fail, the function does not change throughout the course of cell communication. However, molecules can degrade over longer time scales, indicating a time-adaptive strategy would be most appropriate for large time scales [5]. Given the relatively small timescale in cell communication, an *a priori* strategy is adopted as the comprehensive strategy to calculate the shortest time-dependent signaling hyperpath to model cell communication.

With the wide use of pathway databases to represent cellular processes, algorithms have been suggested to query compound signaling graphs [39]. Alongside these algorithms, the visual language Systems Biology Graphical Notation (SBGN) graphically represents biochemical interactions, which can accurately depict signaling hypergraphs through its bipartite graph-based notation [40]. Any computation performed with hypergraphs can be performed with regular graphs, given additional node and edge types. Hypernodes can be represented by compound nodes for use in regular graphs [39]. By introducing dummy reaction nodes, a hyperedge e can be represented as several connecting edges, where reaction nodes connect the tail node to the head node. Since hyperedges can be modeled as connecting regular edges, the theoretical hypergraph length and graph length will differ. Although a regular graph approach can similarly model cell communication, a hypergraph approach becomes more convenient to use in the proposed algorithm. Thus, an algorithm to determine these hyperpaths given the biological constraints will provide more convenient modeling than a bipartite graphing model.

Algorithms exist for finding the shortest signaling hyperpath and for finding the shortest time-dependent hyperpath. An algorithm is proposed to calculate the shortest signaling time-dependent hyperpath by adapting several algorithms to achieve this goal [1] [28] [37]. By requiring realistic time-dependent parameters, the proposed algorithm considers the biological constraints during the unlikely scenario that all resources within a signaling pathway are present.

Other research motivated our ideas includes [41]–[55].

III. Definitions

A. Signaling Hypergraphs

A hypergraph is a generalization of a graph, where an edge can join any amount of vertices. Given the restriction to represent signaling pathways as directed hypergraphs, future references to directed hypergraphs, directed hyperpaths, and directed hyperedges will simply be hypergraphs, hyperpaths, and hyperedges, respectively. Let V be a finite set of nodes and E be a finite set of edges.

Since many biological reactions involve protein complexes, a set of proteins function as a single unit in a reaction. To accurately model complexes, a definition of a *hypernode* is provided as a set of nodes $u \subseteq V$ that act together as a single unit. A *hypernode* u may contain both a complex acting as a single unit or a single node, which can represent a single protein. The set of hypernodes will be denoted $\mathcal{V}$ with the assumption that each node in V is contained in some hypernode in $\mathcal{V}$, so all nodes are accounted for in the translation to hypernodes. A *signaling hyperedge* e will be a pair ($T(e)$, $H(e)$) where each member of the tail or head is a set of hypernodes, so $T(e) \subset \mathcal{V}$ and $H(e) \subset \mathcal{V}$. A finite set of signaling hyperedges is denoted $\mathcal{E}$. Note that a *hypergraph* $\mathcal{H}$ is a directed graph when $|T(e)| = |H(e)| = 1$.

Hypergraphs can join any number of hypernodes with a single hyperedge. For example, the hyperedge e_3 in Fig. 1 connects the three hypernodes u_2, u_4, and u_5. In this example, $T(e) = \{u_2, u_4\}$ and $H(e) = \{u_5\}$. A hypernode may contain multiple nodes. The hypernode u_6 contains three nodes. This property of a hypernode enables a group of proteins to be represented as a single compound, which is important if the compound is a reactant in biochemical reactions.

Each positive regulator is represented as a hypernode to correctly model positive regulation. The set of hypernodes u is added to the tail of the signaling hyperedge if u is a positive regulator for a reaction. If all positive regulators must be present for the reaction to execute, all regulators are added to the tail of the signaling hyperedge. Thus, signaling

hyperedges can represent multiple positive regulators. If any positive regulators can initiate the reaction, a copy of the signaling hyperedge is made for each regulator.

A *signaling hypergraph* is notated $\mathcal{H} = (\mathcal{V}, \mathcal{E})$, where V is a finite set of nodes, $\mathcal{V} \subseteq 2^V$ is a set of hypernodes, and $\mathcal{E}$ is a finite set of signaling hyperedges [2].

Signaling hypergraphs represent reactions among more than two molecules, complexes, and combinatorial positive regulations. Additionally, signaling hypergraphs can model post-translational modifications (PTMs) and complex arrangements. However, negative regulation and more complex regulatory logic are not yet represented as signaling hypergraphs.

B. Hyperpaths

Many definitions of directed hypergraph paths exist [56] [57]. One definition extends to signaling hypergraphs [1]. An *s-d path* $P(s, d)$ is an alternating sequence of hypernodes and hyperedges starting at a hypernode s and ending at a hypernode d where $s, d \in \mathcal{V}$. This definition is notated

$$P(s, d) = (u_1, e_1, u_2, e_2, ..., u_{k-1}, e_{k-1}, u_k), \tag{1}$$

where $s = u_1$, $d = u_k$ and for every $1 \leq i < k$, $u_i \in T(e_i)$ and $u_{i+1} \in H(e_i)$ [56]. An intermediary hypernode belongs to both $H(e)$ and $T(e)$. A path $P(s, d)$ is *simple* if it does not contain repeated hypernodes or hyperedges. $P(s, d)$ is a *simple cycle* if u_1 and u_k are both in the tail of e_1. The signaling hypergraph $\mathcal{H}$ is *acyclic* if it does not contain any simple cycles for any hypernode pair $s, d \in \mathcal{V}$.

Simple paths do not capture all associated hypernodes and hyperedges in the path, since simple paths report an alternating sequence of hypernodes and hyperedges. Thus, signaling reactions involving multiple reactants and products are inaccurately modeled with simple paths. All reactants must be present for all products of the signaling reaction to be present, a notion developed in hyperpath literature [56] [57].

The set of hyperedges e for which a hypernode $u \in \mathcal{V}$ and $u \in H(e)$ is the *backward star* $BS(u)$. Thus, the set $BS(u)$ contains all hyperedges going into u. Given a hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$ and a hypernode $s \in \mathcal{V}$, the hypernode $u \in \mathcal{V}$ is *B*-connected to s in $\mathcal{H}$ if either of the following conditions are met: (i) $u = s$ or (ii) a hyperedge $e \in BS(u)$ exists such that for all $w \in T(e)$, w is *B*-connected to s. The following notation is adopted: $\mathcal{B}_{\mathcal{H}}(s)$ denotes the set of hypernodes that are *B*-connected to s in $\mathcal{H}$. If positive regulators are included in the tail $T(e)$, the definition of B-connectedness will be modified: all reactants and positive regulators in the tail of the reaction must be present for all products of a signaling reaction to be present.

A *sub-hypergraph* $\mathcal{H}' = (\mathcal{V}_{\mathcal{H}'}, \mathcal{E}_{\mathcal{H}'})$ of $\mathcal{H}$ consists of subsets of hypernodes and hyperedges that satisfy $\mathcal{V}_{\mathcal{H}'} \subseteq \mathcal{V}$ and $\mathcal{E}_{\mathcal{H}'} \subseteq \mathcal{E}$ of $\mathcal{H}$ respectively, with the property that for every hyperedge $e \in \mathcal{E}_{\mathcal{H}'}$, both $T(e)$ and $H(e) \in \mathcal{V}_{\mathcal{H}'}$. Given two hypernodes $s, d \in \mathcal{V}$ and a hypergraph $\mathcal{H}$, an s-d B-hyperpath $\Pi(s, d)$ is a sub-hypergraph of $\mathcal{H}$ where $d \in \mathcal{B}_{\Pi(s,d)}(s)$ and $\Pi(s, d)$ has the deletion of all hyperedges and hypernodes that are not necessary to keep hypernode d B-connected to hypernode s. Thus, the hypergraph $\mathcal{H}$ is simplified, and d is B-connected to s. Only the hypernodes and hyperedges in $\Pi(s, d)$ must be used to B-connect the hypernodes s and d. The set of hypernodes B-connected to s in $\Pi(s, d)$ is the same set of hypernodes $\mathcal{V}_{\Pi(s,d)}$ in $\Pi(s, d)$.

Several lemmas and corollaries have been provided in hypergraph literature to reveal fundamental properties, such as the following: if $\Pi(s, t)$ is a B-hyperpath in $\mathcal{H}$, then $\mathcal{V}_{\Pi(s,t)} \subseteq \mathcal{B}_{\mathcal{H}}(s)$. One proposed traversal algorithm identifies the set of hypernodes that are B-connected to hypernode s [56].

C. Shortest Hyperpath

Since there may exist many s-d routes within a hypergraph $\mathcal{H}$, there may be multiple hyperpaths in $\mathcal{H}$. The objective is to find the most efficient hyperpath, representing a minimal amount of reactions from s-d. Adopting notation from [1], a hyperpath $\Pi^*(s, d)$ of $\mathcal{H}$ is computed with a minimal amount of hyperedges:

$$\Pi^*(s, d) = \underset{\Pi: T \in \mathcal{B}_{\Pi}(S)}{\operatorname{argmin}} |\mathcal{E}_{\Pi}| \tag{2}$$

All sub-hypergraphs where d is B-connected to s in Π comprises Π; thus, this set of sub-hypergraphs contain all s-d hyperpaths. Since reactions in signaling hypergraphs likely involve a small number of proteins, signaling hypergraphs are considered a special case of directed hypergraphs.

Several properties exist regarding a hyperpath $\Pi(s, d)$. An ordering exists $o : \mathcal{V} \mapsto \mathbb{R}$ of the hypernodes in a hypergraph $\mathcal{H}$. This function maps each hypernode in $\mathcal{V}$ to a real number $\mathbb{R}$. A valid ordering exists if every $e \in \mathcal{E}$ and every pair of hypernodes $u \in T(e)$ and $w \in H(e)$, $o(u) < o(w)$ [58]. Given a valid ordering in hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$, hypergraph $\mathcal{H}$ is acyclic. Given this acyclic hypergraph exists, there exists some hypernode $u \in \mathcal{V}$ where $BS(u) = \emptyset$. Additionally, given the hyperpath $\Pi(s, d)$, s is the only hypernode where $BS(s) = \emptyset$ [1].

D. Time-dependent Hypergraphs

Given two sets of hypernodes u and w and a hyperedge $(u, w) \in \mathcal{E}$, the set of all possible departure times from u to w is denoted as $DT(u, w)$. Let $DT(u)$ denote the set of all possible departure times from the set of hypernodes u. Adopting the following notation from [28], this set can be written as $DT(u) = \bigcup_{\forall (u,w) \in \mathcal{E}} DT(u, w)$, where $\forall w \in \mathcal{V} \setminus \{s\}$. Additionally, $DT(w)$ denotes the set of all possible arrival times at the final set of hypernodes w.

The random variable denoting the arrival time at w when leaving from u at time t along (u, w) will be $X(u, w, t)$. A probability mass function can be employed on the discrete random variable $X(u, w, t)$ and is denoted as $Pr\{X(u, w, t) = t_i\} = p_{uwt}(t_i)$, such that $\forall t_i \in I(u, w, t)$ where $I(u, w, t) = \{t_1, t_2, ..., t_{k(u,w,t)}\}$ represents the set of possible arrival times at the set of hypernodes w when leaving u at time t along hyperedge (u, w).

The total number of possible departure times, κ, can be denoted by

$$\kappa = \sum_{(u,w) \in \mathcal{E}, t \in DT(u,w)} \kappa(u, w, t), \tag{3}$$

where $\kappa(u, w, t)$ represents the path from u to w along hyperedge (u, w) when departing at time t. Thus, κ can represent the size of the hypergraph, or the size of the input.

An *a priori* strategy is defined as follows. A strategy is a function S with domain $Dm(S) \subseteq \{(u, t) : u \in \mathcal{V} \setminus \{d\}, t \in DT(u)\}$. This assigns each pair $(u, t) \in Dm(S)$ a successor hyperedge $(u, w) \in FS(u)$, where $FS(u)$, read "Forward Star," denotes the set of all hyperedges $e \in \mathcal{E}$ such that $u \in T(e)$. This strategy must satisfy two conditions. If $(u, t) \in Dm(S)$, $S(u, t) = (u, w)$, and $w = d$, then $t \in DT(u, w)$. However if $w \neq d$, then $(w, t') \in Dm(S), \forall t' \in I(u, w, t)$. Thus, there will be multiple intervals, where t' indicates a departure time for the partial interval. A complete interval from s to d can be partitioned into multiple intervals as long as there is no wait between the intervals.

This strategy provides routing choices for departing from all hypernodes and departing times in the domain $Dm(S)$ towards the destination hypernode d. Thus, any communication in a signaling pathway leaving u at time t travels along the hyperedge $S(u, t)$.

It is important to note that a hypernode may contain a single node. In this case, the proposed method can still compute the shortest time-dependent signaling hyperpath.

E. Signaling Time-dependent Hypergraphs

By implementing the previously defined time-dependent hypergraphs with signaling hypergraphs, time-dependent signaling hypergraphs are subsequently defined. There are several things to note.

Since biological processes are predefined and a model of signaling pathways with *a priori* time-dependent hypergraphs is chosen, the time-dependent signaling hypergraphs will not need a time-adaptive strategy. However, time-adaptive strategies can be implemented at the cost of computational efficiency [4]. Within recent stochastic time-dependent hypergraph research, a large emphasis has been placed on time-adaptive strategies for time-dependent hypergraphs [28] [36] [37].

The assumption is made that departure times are positive and that the communication in the signaling time-dependent hypergraph cannot pause at an intermediate hypernode w. Thus, if it is possible to arrive at hypernode w at time t_i, then it is possible to leave hypernode w at time t_i. A process cannot wait at intermediate hypernodes. Within the hypergraph model, a hypernode can contain either a compound of many proteins or a single protein. This flexibility allows translation from directed graphs to hypergraphs without a loss of information.

IV. Approach

It is important to note several properties of time-dependent signaling hypergraphs. To run a shortest acyclic hyperpath algorithm, hypergraphs must have proper ordering.

Adopting several constraints from [1], binary variables α_u and α_e exist for any hypernode $u \in \mathcal{V}$ and hyperedge $e \in \mathcal{E}$, respectively. A hypernode u is included in the sub-hypergraph $\mathcal{H}'$ if and only if $\alpha_u = 1$. Likewise if $\alpha_e = 1$, the corresponding hyperedge e is included in $\mathcal{H}'$. The following constraints exist for the α variables:

$$\forall u \in \mathcal{V} \setminus \{s\} : \begin{cases} \alpha_u \leq \sum_{e \in BS(u)} \alpha_e & BS(u) \neq \emptyset \\ \alpha_u = 0 & otherwise \end{cases} \tag{4}$$

$$\forall e \in \mathcal{E} : \sum_{u \in T(e)} \alpha_u \geq |T(e)| \alpha_e \tag{5}$$

$$\forall e \in \mathcal{E} : \sum_{u \in H(e)} \alpha_u \geq |H(e)| \alpha_e \tag{6}$$

$$\alpha_t = 1 \tag{7}$$

If a real-valued order variable, notated o, and α can simultaneously satisfy (4)-(7), then the resulting sub-hypergraph $\mathcal{H}(\alpha) = (V(\alpha), \mathcal{V}(\alpha), \mathcal{E}(\alpha))$ has a valid ordering, where within any hyperedge e, the hypernodes in the head must have an order value greater than the corresponding hypernodes in its tail. Thus, $\forall e$ such that $\alpha_e = 1; \forall (u, w) \in T(e) \times H(e) : o_u < o_w$.

A final constraint is introduced in the context of time-dependent signaling hypergraphs. The set $\mathcal{P}_p$ corresponds to the set of proteins that p can communicate with, where $p \in T(e)$ for some hyperedge e. Any given protein cannot communicate with all proteins in a signaling pathway. Therefore, a protein set $\mathcal{P}_{p_1}$ may include different proteins and compounds than a second set $\mathcal{P}_{p_2}$, implying that protein p_1 can communicate with a different set of proteins than p_2. Given a time, a protein u can communicate with a protein w along the hyperedge e, where $w \in \mathcal{P}_u$ and $(u, w) \in T(e) \times H(e)$. These stated constraints lead to the following lemma.

Lemma 1. *The set of communicating proteins in $\mathcal{H}(\alpha)$ form an acyclic path along hyperpath P.*

Proof. Let u be a protein in a simple cycle $P = (u_1, e_1, u_2, ..., u_{k-1}, e_{k-1}, u_k)$ where the final protein u_k belongs to the set of tails to hypernode u_1, denoted $u_k \in T(e_1)$. Let protein w have the smallest order value, where $w = \text{argmin}_{u_i \in P} o(u_i)$, and $w \in H(e)$. There must exist a protein u_j, where $u_j \in T(e)$ since P is a simple cycle. Thus, $o(u_j) < o(w)$, contradicting the fact that the protein w minimizes the order function in the simple path P; so, P must be acyclic. Since cell communication is modeled with shortest paths, proteins will not engage in cycles. Rather, the proteins will communicate once along the shortest signaling path if they are in the shortest signaling path. A protein p_i must communicate with a protein p_j along hyperedge e_i, such that $p_j \in P_{p_i}$. Therefore, $(p_i, p_j) \in T(e_i) \times H(e_i)$. This protein p_j communicates to another protein in the simple hyperpath $P = (p_1, e_1, p_2, ..., p_{k-1}, e_{k-1}, p_k)$, where $p_i \notin \mathcal{P}_{p_j}$, since P is acyclic. Hence, the set of communicating proteins in $\mathcal{H}(\alpha)$ form an acyclic path. □

At different times, different proteins may be present within a biological system. Thus, the protein p_i must belong to the set $T(e)$ of the hyperedge e.

Corollary 1. *All hypernodes in a sub-hypergraph are B-connected to hypernode s in $\mathcal{H}(\alpha)$.*

The following optimization problem is introduced, using the discussed constraints adapted to time-dependent signaling pathways from [1]:

$$\underset{\alpha,o}{\operatorname{argmin}} \sum_{e\in\mathcal{E}} \alpha_e, \tag{8}$$

which is subject to constraints (4)-(7).

A set of *a priori* path-strategies is denoted $\mathcal{S}$. Let $\mathcal{L}$ denote the set of s-d hyperpaths in a hypergraph $\mathcal{H}(\alpha)$. The set $\mathcal{L}$ can be partitioned into disjoint subsets, denoted $\mathcal{L}^i$, where $1 \leq i \leq q+1$, given a sub-hyperpath $P = (s = u_1, ..., u_{q+1})$. For $1 \leq i \leq q$, paths in $\mathcal{L}^i$ contain the path $P^i = (u_1, ..., u_i)$, where the arc (u_i, u_{i+1}) is not included. However, hyperpaths in $\mathcal{L}^{q+1}$ contain path P. This notation and the following theorem have been adapted from [37].

Theorem 1. *Given the disjoint sets $\mathcal{L}^i$, where $1 \leq i \leq q+1$, the following statements are equivalent.*

1. $P \in \mathcal{L}^i$.

2. P is a s-d hyperpath in $\mathcal{H}(\alpha)$.

3. There is a unique path-strategy $\mathcal{S}^i \in \mathcal{S}^i_P$ corresponding to the hyperpath P in the hypergraph $\mathcal{H}(\alpha)$.

Proof. The equivalence of the first two statements follows from the fact that P must be a hyperpath of $\mathcal{H}(\alpha)$, and $\mathcal{L}^i$ represents disjoint subsets of $\mathcal{H}(\alpha)$. If $P \in \mathcal{L}^i$, then P is a s-d hyperpath in $\mathcal{H}(\alpha)$. The equivalence of the first two statements with the third statement follows from the one-to-one correspondence between paths and path-strategies and between strategies and hyperpaths. □

Using this theorem, a sub-hypergraph $\mathcal{H}^i$ can represent its corresponding subset $\mathcal{L}^i$. Thus, each sub-hypergraph $\mathcal{H}^i$ defines a time-dependent hypergraph, where the set of *a priori* strategies is $\mathcal{S}^i$ and the set of path-strategies is notated as the subset $\mathcal{S}^i_P \in \mathcal{S}^i$. Thus, the following corollary is provided for time-dependent hypergraphs.

Corollary 2. *For the set of path-strategies $\mathcal{S}^i_P$, the weight of the best a priori strategy in $\mathcal{S}^i$ is a lower bound.*

V. Algorithm

The original RunMILP algorithm does not consider time dependency for biological resources within signaling pathways. This algorithm's "shortest signaling hyperpath" is only accurate when all proteins, enzymes, and compounds within the signaling pathway are available. However, this is unlikely to occur given the dynamic nature of biological systems. The proposed algorithm considers time dependency when determining the shortest signaling hyperpath, which models cell communication that is dependent on the realistic availability of biological resources.

The *Modified RunMILP*$(\mathcal{H}, \mathcal{B}_\mathcal{H}(s), s, d)$ requires a hypergraph, denoted $\mathcal{H}$, with a starting hypernode and destination hypernode belonging to $\mathcal{V}$ and the set of hypernodes that are B-connected to the starting hypernode, including the destination hypernode. A sub-hypergraph $\mathcal{H}'$ is generated by removing one or more proteins from the signaling pathway, where $\mathcal{H}' = \mathcal{H}(\mathcal{B}_\mathcal{H}(s))$ is the induced sub-hypergraph on the B-connected set of hypernodes (line 1). The removal of proteins can reflect the state of a signaling pathway throughout various biological processes. Then, α and O is solved for, where α and o are optimal objectives, using an adopted mixed integer linear program (MILP) from [1]. This program optimizes Equation (8) and stores the optimal objective (lines 2-3). A set of shortest acyclic hyperpaths, R, is created and initializes the set to be null (line 4). While the amount of hyperedges in the hyperpath is minimal, the hyperpath is added to R. Then, a constraint is added such that the amount of hyperedges in the sub-hypergraph is less than or equal to the amount of hyperedges in the optimal sub-hypergraph (lines 5-7). Given the new constraint, the MILP is re-solved on the sub-hypergraph and stores the new optimal objectives (line 8). Finally, the set of shortest signaling hyperpaths is returned (line 10). If multiple hyperpaths have the same optimal objective score, the entire set of hyperpaths are returned. One example of a tied result can be found in Table IX, where there are two shortest signaling hyperpaths following the removal of the Tyrosine-protein kinase LCK in the IL2 signaling pathway. All shortest acyclic hyperpaths are plausible signaling hyperpaths for cell communication given different time instance.

Algorithm 1 Modified RunMILP$(\mathcal{H}, \mathcal{B}_\mathcal{H}(s), s, d)$

Require: $\mathcal{H}, \mathcal{B}_\mathcal{H}(s); s \in \mathcal{V}, d \in \mathcal{V}, d \in \mathcal{B}_\mathcal{H}(s)$

1: $\mathcal{H}' = \mathcal{H}(\mathcal{B}_\mathcal{H}(s))$;
2: $\alpha, o :=$ Solve Equation (8) on $\mathcal{H}', s, d$ and t;
3: opt $:= |\mathcal{E}(\alpha)|$;
4: $R := \emptyset$;
5: **while** $|\mathcal{E}(\alpha)| =$ opt **do**
6: $\quad R := R \cup \mathcal{H}(\alpha)$;
7: $\quad$ Add constraint such that $\Sigma_{e\in\mathcal{E}_{\mathcal{H}'}} \alpha_e \leq |\mathcal{E}(\alpha)|$;
8: $\quad \alpha, o :=$ Re-solve the MILP on $\mathcal{H}', s,$ and t;
9: **end while**
10: **return** R;

VI. Experiment

A. NCI-Pathway Interaction Database

The National Cancer Institute - Pathway Interaction Database (NCI-PID) is a curated collection of biological data regarding biomolecular interactions and important cellular processes occurring in signaling pathways [59]. NCI-PID was a collaborative project from 2006 to 2012. While listing over 200 data sets, this experiment will focus on the following four signaling pathway data sets provided by NCI-PID: DNA-PK, PDGFR-α, p53, and PDGFR-β_1.

B. Reactome Pathway Database

Reactome is an open-source curated and peer-reviewed pathway database founded in 2003 for the visualization, interpretation, and analysis of pathway data [60]. The three selected signaling pathways from this database are Ceramide, Ca2+, and Purine catabolism.

C. Pathway Commons

Pathway Commons is a web resource for biological pathway data that integrates signaling pathways from public pathway and interactions databases. Pathway Commons contains over 37,600 pathways and three million interactions. The following five signaling pathways were selected from Pathway Commons for this experiment: PDGFR-β_2, IL2-mediated, Regulation of Telome, IL2, and Gastrin CCK2R 240212.

D. Using Atemporal Data

The three databases utilized in this experiment provide atemporal signaling pathway datasets. It would be both costly and time-consuming for any temporal signaling pathway dataset to be produced. To justify time dependency in the experiment, several proteins were eliminated from the signaling hypergraph. The elimination of proteins from a signaling pathway modifies the connectivity of the network. The justification for the removal of proteins to reflect time dependency within the signaling hypergraph is provided below.

Many biological processes affect the availability of resources. For every removed protein in the experiment, several academic papers discuss its potential absence from the network. Therefore, the availability of proteins are directly affected by the biological processes occurring in the body. For example, the absence of Telomerase reverse transcriptase occurs following chromosomal mutation [61] and platelet-derived growth factor proteins are not frequently activated for adult women [62]. Since biological processes are not constant nor are some frequent, the occurrence of these processes change over time. Therefore, the biological processes affecting the availability of proteins within signaling pathways are time-dependent, and the removal of several proteins can reflect time dependency within the pathway. Despite the atemporal nature of the datasets available, the removal of proteins can reflect the time-dependent nature of signaling pathways.

E. Setup

These datasets were chosen for their complete variety of size for signaling pathways. Data used in this experiment were in the BioPAX format [63]. This experiment utilized a modified parser provided in PaxTools to interpret BioPAX data [64].

Extracted data from all three databases are not time-dependent. Rather, the downloaded datasets are a complete hypergraph, containing all proteins and connections between proteins. Since the data is not inherently time-dependent, the sub-hypergraphs resulting from the removal of proteins will reflect the justified time-dependent nature of biological resources. To more robustly model signaling pathways with respect to the availability of biological resources, several hypernodes are removed from the complete hypergraph, and the subsequent sub-hypergraph $\mathcal{B}_{\mathcal{H}}(o)$ is utilized throughout the algorithm. The removal of biological resources reflects time dependency since every biological process degrades or activates proteins, enzymes, and complexes. Thus, all resources will not always be present in each signaling pathway.

Twelve signaling pathways were chosen from NCI-PID, Reactome, and Pathway Commons to display the effect of time dependency on signaling pathways in a variety of conditions. The selected signaling pathways have a variety of node count, hypernode count, edge count, and hyperedge count (see Table I). Some pathways' nodes can be greatly condensed to hypernodes, such as the DNA-PK pathway. Other pathways' nodes, such as the PDGFR-β_1 pathway, cannot be greatly condensed to hypernodes. The condensing of nodes to hypernodes occurs when a set of elements can be represented as a single compound, where each compound is represented as one hypernode. Likewise, some selected pathways' edges can be highly condensed to hyperedges, such as the PDGFR-α pathway. Evidently, a pathway that has a highly condensed hypernode count will likely have a highly condensed hyperedge count.

It is important to note that the Purine catabolism signaling pathway does not contain any complexes composed of proteins within its pathway. Therefore, all nodes will remain as separate nodes, which implies that all edges will remain as separate edges. There is no translation from nodes to hypernodes or edges to hyperedges, yet the proposed algorithm can still determine the shortest signaling hyperpath, revealing another robust quality of Modified RunMILP.

In preparation for the experiment, several sub-hypergraphs were created for each signaling pathway by removing various hypernodes that reflect reasonable variations dependent upon time. Each of the removed hypernodes were contained in the shortest signaling hyperpath given all resources were available. By removing these proteins, a different hyperpath will become the shortest signaling hyperpath, displaying the effect of time dependency on biological resources and, consequently, cell communication.

Before analyzing the PDGFR-β, Purine catabolism, and Gastrin CCK2R 240212 signaling pathways, several neighboring proteins were removed to more accurately model biological events, such as mutation, phosphorylation, and condensation of multiple molecules. The resulting network contains a lower level of connectivity due to the loss of neighboring hypernodes. Despite the removal of several neighboring hypernodes, the proposed algorithm still computes the shortest signaling path.

Throughout this experiment, the proposed algorithm was implemented on a MacBook Pro operating on macOS High Sierra with a 2.9 GHz Intel Core i7 processor. The algorithm was run in a Jupyter Notebook using Python and the *halp* and *ndex* packages.

F. Results and Analysis

Within each signaling hypergraph, there exists one or more shortest paths for the entire system. If a protein or complex within the shortest path is not present in the signaling pathway at various time instances, then the shortest path for the entire hypergraph will not be utilized, and a longer hyperpath will execute the cell communication. To reveal the effect of time-dependent resources on cell communication, only proteins that are in the complete shortest acyclic hyperpath were removed.

To display that signaling hypergraphs are time-dependent, several proteins were removed from the pathways, reflecting

TABLE I
PROPERTIES OF THE SELECTED SIGNALING PATHWAYS

Signaling Pathway	Node #	Hypernode #	Edge #	Hyperedge #
DNA-PK	14	5	66	4
PDGFR-α	26	18	43	19
p53	59	52	181	66
PDGFR-β_1	139	132	1277	701
PDGFR-β_2	128	77	1019	170
IL2 Mediated	55	33	619	186
Regulation of Telome	71	57	229	145
Ceramide	58	19	248	38
Ca2+	70	55	563	346
IL2	66	15	210	27
Purine catabolism	43	43	149	149
Gastrin CCK2R 240212	136	33	251	33

that the available resources in signaling pathways are time-dependent. A unique sub-hypergraph reflects each time instance. With different sub-hypergraphs, each reduced pathway may result in a different set of shortest acyclic hyperpaths. The shortest acyclic hyperpath thus differs between time instances. Additionally, these differences affect the proteins used in cell communication and the length of the shortest signaling hyperpath that executes the cell communication.

Although a signaling hypergraph may have many nodes, a complex was treated as a single hypernode. Therefore, if the signaling pathways contain many proteins, the collective nodes were treated as a hypernode. A large amount of complexes within a signaling pathway resulted in very limited variability in the shortest acyclic hyperpaths for some signaling pathways because the cardinality of the set of sub-hyperpaths is relatively small. Since nodes in a complex were not independently B-connected to the starting node in different hypergraphs, this may result in a very limited number of possible hyperpaths.

The signaling pathways analyzed in this experiment were selected to represent the diversity in signaling hypergraph size. The removal of a single hypernode in a smaller hypergraph would affect the set of hyperpaths more drastically than it would affect a larger hypergraph's set of hyperpaths. Since the cardinality of the set of hyperpaths for the DNA-PK pathway is less than that of the PDGFR-β_1, it is expected that a single missing hypernode would affect the set of B-connected hypernodes within the DNA-PK pathway greatly, and the length of its shortest acyclic hyperpath would not likely be volatile. Since there is no way to alter any of the DNA-PK shortest paths and maintain a B-connected set from the starting hypernode to the destination hypernode, there is only a single signaling hyperpath, and consequently one shortest acyclic hyperpath. This is the expected result for such a pathway as it only has pathways that was one hyperedge long.

However, larger signaling pathways, such as the PDGFR-α signaling pathway and p53 pathway, contain a larger set of hyperpaths, so the removal of several hypernodes and their respective hyperedges would not eliminate every hyperpath. Thus, these large signaling pathways are more forgiving in the lack of proteins present in the pathway and are more likely to result in different sets of hyperpaths from the starting hypernode to the destination hypernode. Although the proteins and complexes utilized in cell communication may change, the length of the shortest acyclic hyperpath remained relatively constant. This can be attributed to the abundance of hyperpaths available in the signaling hypergraph.

A table is provided for each of the twelve signaling pathways analyzed. The effect of time dependency on these signaling pathways can be determined by the change from shortest path to the timed shortest path. The "shortest path" displays the shortest path for the complete signaling hypergraph, without the removal of any hypernodes. These paths can be determined from the original RunMILP algorithm. "Shortest path length" represents the amount of processes that execute in the complete shortest acyclic hyperpath. The "shortest path (timed)" displays the shortest path following the removal of the specified hypernode in the table ("removed hypernode"). These paths are the results from the proposed Modified RunMILP algorithm. Likewise, the "shortest path length (timed)" represents the length of the newly considered shortest path. Hypergraphs were used in the algorithm to identify the shortest signaling hyperpath, since the length of the entire cell communication process can be determined from the cardinality of the set of hyperedges. However, the results can be described in either bipartite graphs or, simply, graphs once the shortest signaling path is known. Thus, each test shares a single shortest signaling path in simple path form.

Since the DNA-PK signaling hypergraph has five hypernodes with four hyperedges, there is no variation in hyperpath length. Thus if the starting hypernode is set as the metabolite Inositol hexakisphosphate (IP6) and the destination hypernode as the DNA repair gene XRCC5, there is only one possible path for any communication to occur, which is directly between the two hypernodes (Table II).

The PDGFR-α signaling pathway has a larger set of hypernodes and hyperedges, so a variety of paths can execute functions and communications between a starting and destination hypernode (Table III). The selected starting hypernode, Phosphatidylinositol 4,5-bisphosphate (PIP_2), can utilize a chain of PLCG1, PRRT2, and ELK1 or a chain of DAG, PRRT2, and ELK1 to terminate at the destination hypernode, the oncogene FOS. The removal of PLCG1 results in the timed shortest path. Since PLCG1 is activated by the Proto-oncogene tyrosine-protein kinase Src, it is not always present in a signaling pathway [65]. Therefore, its removal is justified.

Although the third and fourth hypernodes are identical in these two paths, the second hypernode differs, displaying robustness in the signaling pathway's communication. At various times, either of the interchangeable hypernodes may not be present due to time dependency. Although both of the shortest paths are the same length, the proteins and complexes involved differ.

The p53 pathway, similar to the PDGFR-α signaling pathway, has the same hyperpath length with different proteins and complexes functioning in the hyperpath at different time instances (Table IV). The Cyclin-A2 protein (CCNA2) is removed to reflect different resources available at different time instances. Levels of CCNA2 are well synchronized with the progression of the cell cycle, so it is not atypical to see a lack of CCNA2 in a signaling pathway during the cellular phase G2 [66].

The PDGFR-β_1 pathway is the largest signaling pathway tested. Therefore, this pathway has the largest variety of sub-hypergraphs and, consequently, hyperpaths. After removing several hypernodes, the length of the shortest path between PIP_2 and phosphate increases (see Table V). Therefore, this dataset illustrates that, due to time dependency, the proteins and complexes used in the shortest acyclic hyperpath may change and the length of the hyperpath may change depending on which resources are available in a signaling pathway. The removed resources are PDGFβ, PDGFRβ, CRK, and RAPGEF1. The platelet-derived growth factor proteins, PDGFβ and PDGFRβ, are activated when their receptor's kinase activity is de-repressed, which is not a frequent biological process for adult women [62]. If a woman's cell communication were to be modeled, the presence of PDGFβ and PDGFRβ should likely be removed from the shortest path. The CRK molecule is a member of an adapter protein family that participates in the Reelin signaling cascade of DAB1. If DAB1 is not present, which is likely to occur in the PDGFR-β signaling pathway, then CRK will not execute in cell communication [67]. Additionally, the nucleotide exchange factor RAPGEF1 has a limited amount of resources that it reacts with, where the presence of any of these resources is time-dependent [68]. Therefore, to robustly model cell communication in this pathway given its time-dependent resources, time dependency must be considered.

The PDGFR-β_2 pathway is a derivative of the same signaling pathway from which the PDGFR-β_1 pathway is derived. By selecting different starting and destination hypernodes for the PDGFR-β_2 pathway, the effect of time dependency on the shortest path length is shown to be different. In this dataset, the stress-activated protein kinase MKK7 is removed. Since MKK7 is activated during stressful events, it is absent in signaling pathways quite frequently [69]. Therefore, the removal of this protein kinase reflects real time-dependent parameters. The removal of MKK7 causes the PDGFR-β_2 pathway to change its second hypernode in the shortest path, bypassing the now removed MKK7 hypernode (Table VI).

Although the removal of a hypernode may not affect the shortest path's length, it can still affect the timed shortest path. The dual-specificity protein kinase MEK1 is removed from the IL-2 Mediated signaling pathway and the Ceramide pathway (Tables VII and VIII). This protein kinase requires phosphorylation of two conserved Ser/Thr residues to become active [70]. Since this requirement is not always satisfied, the absence of MEK1 is expected to occur frequently in many signaling pathways, such as the IL-2 Mediated and Ceramide pathways. For the non-mediated IL2 signaling pathway, the cytosol LCK is removed. LCK can be inhibited by kinase inhibitors in the form of novel drugs [71]. These inhibitors can treat inflammation and autoimmune disorders, and they have become more prevalent in the biotech and pharmaceutical industries. This hypernode is removed to reflect how some elements within signaling pathways can become absent due to drugs. Following the removal of this hypernode, the IL-2 pathway contains two shortest paths, labeled (1) and (2) in Table IX. Since a tie for shortest path was obtained from the proposed algorithm, both results are reported. If an individual were to lack MEK1 and LCK, their cell communication's path would significantly differ from the shortest path provided by currently accepted modeling techniques. Since the shortest path for the IL-2 pathway maintains shortest path length with different elements, the representation of its communication is more robust, as it considers the proteins and compounds present within an individual's body given a specific time.

The Ca2+ pathway's shortest path length changes when removing the 1,2-diacyl-sn-glycerolipid C00641 (Table X). This glycerolipid is formed by the condensation of one, two, or three fatty acid molecules on glycerol. When fatty acids are not present in a signaling pathway, C00641 is not activated. Therefore, its removal is not only justified but required to improve the robustness of modeling cell communication when it is not active [72].

The Regulation of Telome pathway is selected for the experiment to reveal how vital the existence of some elements are to their signaling pathway. By removing Telomerase reverse transcriptase (TERT) from this pathway, there is no destination hypernode for the signaling hypergraph (Table XI). Therefore, there is not a shortest path when TERT is absent in the pathway. This absence is usually a result of chromosomal mutation [61]. Although this mutation is not prevalent in all individuals, it is an example of how biological systems can mutate over time, making the resources in a signaling pathway time-dependent.

After removing the two inosines C05512 and D00054 from the Purine catabolism pathway, the shortest path length increases from six processes to seven processes (Table XII). C05512 is a purine 2'-deoxyribonucleoside that is a reactant of several enzymes. If these enzymes are not present in the signaling pathway, which is a plausible event for many adults, C05512 cannot execute cell communication [73]. Therefore, this dataset exhibits how an element may be present in a signaling hypergraph. However, it cannot be in the shortest path since it is unable to react with available resources.

The Gastrin CCK2R 240212 pathway exhibits volatile shortest path length through the absence of several hypernodes. When removing the protein PLCG1 and the enzyme PLA2G4A, the shortest path increases from five processes to eight (Table XIII). As explained for the PDGFR-α pathway, the absence of PLCG1 occurs frequently. Likewise, PLA2G4A

is an enzyme that gets metabolized into an eicosanoid through the release of arachidonic acid. If phospholipids are not in a state of communicable hydrolysis, which would occur when a salt of a weak acid or weak base is dissolved in water, then PLA2G4A cannot get metabolized [74]. Thus, we remove it from the signaling pathway. This is another example of a potentially present enzyme that we cannot include in a shortest path, simply because it cannot react or execute cell communication. The time dependency of the proteins and enzymes within this signaling pathway reveal that the shortest path and cell communication are also time-dependent.

VII. Conclusion and Future Work

With the recent proposal to model signaling pathways with hypergraphs, there is a potential to improve accuracy in modeling cell communication. Proposed hypergraph-based algorithms more accurately reflect the complexity of signaling pathways. However, recent literature does not account for time dependency when analyzing signaling pathways. As with many biological systems and processes, signaling pathways use time-dependent biological resources to execute cell communication. At different time instances, a signaling pathway may have a different set of proteins and complexes present, altering the pathway of cell communication. Thus, the proposal to include time-dependent constraints when modeling cell communication proves more robust than current modeling techniques, since it considers the availability of biological resources.

Through several proved lemmas and corollaries, it is evident that a time-dependent signaling hypergraph is acyclic given a valid ordering. Therefore, all hypernodes in a sub-hypergraph are B-connected to a starting hypernode. These properties were fundamental in modifying the proposed RunMILP algorithm to function with time-dependent hypergraphs.

Signaling pathways were chosen from several databases: NCI-PID, Reactome, and Pathway Commons. The selected signaling pathways were chosen for their variety in size, regarding both nodes and edges. After converting the BioPAX data into hypergraphs, these selected signaling pathways further represented a variety of sizes. While modeling smaller signaling pathways, it was evident that the shortest path length rarely changed. While modeling larger pathways, the length of the shortest path and its contents differed. In both cases, the shortest signaling hyperpath is dependent upon the available resources. After removing several neighboring nodes in the PDGFR-β, Purine catabolism, and Gastrin CCK2R 240212 signaling pathways, the length of the shortest path increased.

The removed hypernodes in the twelve signaling pathways represent realistic expectations of available resources within each pathway at various time instances. Therefore, the resulting sub-hypergraph can be expected to occur due to a variety of conditions. The availability of these resources affects the shortest path when considering time dependency. Thus, it was shown through reasonable expectations that time-dependent signaling hypergraphs more robustly model cell communication.

By analyzing signaling hypergraphs *a priori*, the modification of a proposed MILP algorithm was utilized to compute the shortest B-hyperpaths within twelve signaling pathways. By modifying each signaling pathway before computing the shortest path, it was shown that the resulting shortest acyclic hyperpath was dependent upon the biological resources available in the system. Thus, the robustness of modeling signaling pathways can be improved by utilizing time-dependent signaling hypergraphs.

Since each signaling pathway is in continuous flux, the proteins and complexes present will be difficult to record. Both tedious and costly, the measurement of available resources within signaling pathways may not always be accessible. Therefore, it would be difficult to curate a time-dependent dataset of signaling pathways; however, this development would be beneficial to the future understanding of signaling pathways. Additionally, there can be efforts made to forecast available proteins and complexes within specific signaling pathways.

Since real-time transportation data is generated simultaneously for many users on a transportation network from navigation applications, significant experimental data sets are available. By integrating this experimental data with rather complete hypergraphs, one can test many topological properties of time-dependent hypergraphs. Social media platforms may provide additional large sets of experimental data to integrate with hypergraphs, where hyperedges are interactions between multiple hypernodes, or groups of users.

Although obtaining large temporal datasets may be difficult in some fields, time-dependent hypergraphs can be applied to a variety of domains. In neural imaging, they can model neural oscillation and activity. In logistics, time-dependent hypergraphs can model the coordination of people, facilities, and supplies. In computer science, they can be applied to several fields, such as machine learning, program optimization, and data mining. Due to the convenience of hypergraph properties and the influence of time on many networks, time-dependent hypergraphs may be applied to various disciplines.

VIII. Acknowledgments

This work was supported in part by the United States Department of Defense under Grant W911NF1710088, W911NF1810437, W911NF1810246, and 75585-NS-RI, in part by the National Science Foundation under Grant 1625677 and Grant 1710716. We gratefully acknowledge all the organizations for their support.

References

[1] A. Ritz *et al.*, "Pathway analysis with signaling hypergraphs," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 14, no. 5, pp. 1042–1055, Sep. 2017.

[2] A. Ritz *et al.*, "Signaling hypergraphs," *Trends in Biotechnology*, vol. 32, no. 7, pp. 356–362, Jul. 2014.

[3] Nagashima *et al.*, "Quantitative transcriptional control of erbb receptor signaling undergoes graded to biphasic response for cell differentiation," *Journal of Biological Chemistry*, vol. 282, pp. 4045–4056, Dec. 2006.

[4] L. R. Nielsen *et al.*, "Bicriterion shortest hyperpaths in random time-dependent networks," *IMA Journal of Management Mathematics*, vol. 14, no. 3, pp. 271–303, Jul. 2003.

[5] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, "General principles of cell communication," in *Molecular Biology of the Cell. 4th edition*. Garland Science, 2002.

TABLE II
EFFECT OF TIME DEPENDENCY ON THE DNA-PK SIGNALING PATHWAY

	DNA-PK
Shortest path	IP6 → XRCC5
Shortest path length	2

TABLE III
EFFECT OF TIME DEPENDENCY ON THE PDGFR-α SIGNALING PATHWAY

	PDGFR-α
Shortest path	PIP_2 → PLCG1 → PRRT2 → ELK1 → FOS
Shortest path length	5
Removed hypernode	PLCG1
Shortest path (timed)	PIP_2 → DAG → PRRT2 → ELK1 → FOS
Shortest path length (timed)	5

TABLE IV
EFFECT OF TIME DEPENDENCY ON THE P53 SIGNALING PATHWAY

	p53
Shortest path	ATM → DYRK2 → MAPK14 → CCNA2 → PPP2CA
Shortest path length	5
Removed hypernode	CCNA2
Shortest path (timed)	ATM → DYRK2 → MAPK14 → CDK2 → PPP2CA
Shortest path length (timed)	5

TABLE V
EFFECT OF TIME DEPENDENCY ON THE PDGFR-β_1 SIGNALING PATHWAY

	PDGFR-β_1
Shortest path	PIP_2 → PDGFB → CRK → RAPGEF1 → RAP1 → ARAP1 → phosphate
Shortest path length	7
Removed hypernodes	PDGFB, PDGFRB, CRK, RAPGEF1
Shortest path (timed)	PIP_2 → PIK3CG → EPS8 → ABL1 → VAV2 → RAP1B → ARAP1 → phosphate
Shortest path length (timed)	8

TABLE VI
EFFECT OF TIME DEPENDENCY ON THE PDGFR-β_2 SIGNALING PATHWAY

	PDGFR-β_2
Shortest path	CHEBI:15996 → MAP2K7 → MAPK8 → JUND → MYC
Shortest path length	5
Removed hypernode	MAP2K7
Shortest path (timed)	CHEBI:15996 → MAP2K4 → MAPK8 → JUND → MYC
Shortest path length (timed)	5

TABLE VII
EFFECT OF TIME DEPENDENCY ON THE IL-2 MEDIATED SIGNALING PATHWAY

	IL-2 Mediated
Shortest path	CHEBI:15996 → RAF1 → MAP2K1 → STAT3
Shortest path length	4
Removed hypernode	MAP2K1
Shortest path (timed)	CHEBI:15996 → RAF1 → MAP2K2 → STAT3
Shortest path length (timed)	4

[6] S. Klamt *et al.*, "Hypergraphs and cellular networks," *PLOS Computational Biology*, vol. 5, no. 5, pp. 1–6, Jul. 2009.

[7] T. Hwang *et al.*, "Learning on weighted hypergraphs to integrate protein interactions and gene expressions for cancer outcome prediction," *Eighth IEEE International Conference on Data Mining*, pp. 293–302, Dec. 2008.

[8] Z. Tian *et al.*, "A hypergraph-based learning algorithm for classifying gene expression and arraycgh data with prior knowledge," *Bioinformatics*, vol. 25, no. 21, pp. 2831–2838, Jul. 2009.

[9] T. Aittokallio and B. Schwikowski, "Graph-based methods for analysing networks in cell biology," *Briefings in Bioinformatics*, vol. 7, no. 3, pp. 243–255, Jun. 2006.

[10] U. S. Bhalla and R. Iyengar, "Emergent properties of networks of biological signaling pathways," *Science*, vol. 283, pp. 381–387, Jan.

TABLE VIII
EFFECT OF TIME DEPENDENCY ON THE CERAMIDE SIGNALING PATHWAY

	Ceramide
Shortest path	CHEBI:57925 → SMPD3 → KSR1 → RAF1 → MAP2K1 → MAPK1
Shortest path length	6
Removed hypernode	MAP2K1
Shortest path (timed)	CHEBI:57925 → SMPD3 → KSR1 → RAF1 → MAP2K2 → MAPK1
Shortest path length (timed)	6

TABLE IX
EFFECT OF TIME DEPENDENCY ON THE IL2 SIGNALING PATHWAY

	IL2
Shortest path	IL2 → IL2RA → LCK → MAPK1 → GAB2 → GRB2
Shortest path length	6
Removed hypernodes	LCK
Shortest path (timed) (1)	IL2 → IL2RA → LYN → MAPK1 → GAB2 → GRB2
Shortest path (timed) (2)	IL2 → IL2RA → LYN → MAPK3 → GAB2 → GRB2
Shortest path length (timed)	6

TABLE X
EFFECT OF TIME DEPENDENCY ON THE CA2+ SIGNALING PATHWAY

	Ca2+
Shortest path	CHEBI:18348 → CHEBI:17815 → PRKCA
Shortest path length	3
Removed hypernode	CHEBI:17815
Shortest path (timed)	CHEBI:18348 → CHEBI:16595 → ITPR3 → PRKCA
Shortest path length (timed)	4

TABLE XI
EFFECT OF TIME DEPENDENCY ON THE REGULATION OF TELOME SIGNALING PATHWAY

	Regulation of Telome
Shortest path	IL2 → TERT
Shortest path length	2
Removed hypernode	TERT
Shortest path (timed)	None
Shortest path length (timed)	0

TABLE XII
EFFECT OF TIME DEPENDENCY ON THE PURINE CATABOLISM SIGNALING PATHWAY

	Purine catabolism
Shortest path	CHEBI:16039 → CHEBI:17202 → CHEBI:28997 → CHEBI:17368 → CHEBI:16240 → CHEBI:17858
Shortest path length	6
Removed hypernodes	CHEBI:28997, CHEBI:17596
Shortest path (timed)	CHEBI:16039 → CHEBI:17202 → CHEBI:17172 → PNP → XDH → GPX1 → CHEBI:17858
Shortest path length (timed)	7

TABLE XIII
EFFECT OF TIME DEPENDENCY ON THE GASTRIN CCK2R 240212 SIGNALING PATHWAY

	Gastrin CCK2R 240212
Shortest path	AP2M1 → GAST → PLCG1 → ITPR1 → PTK2B
Shortest path length	5
Removed hypernodes	PLCG1, PLA2G4A
Shortest path (timed)	AP2M1 → GAST → PTPN11 → NOS1 → GUCY2D → CD38 → TPCN2 → PTK2B
Shortest path length (timed)	8

1999.

[11] M. Y. Wu and C. S. Hill, "Tgf-b superfamily signaling in embryonic development and homeostasis," *Developmental Cell*, vol. 16, pp. 329–343, Mar. 2009.

[12] B. Thisse and C. Thisse, "Functions and regulations of fibroblast growth factor signaling during embryonic development," *Developmental*

Biology, vol. 287, pp. 390–402, Oct. 2005.
[13] D. M. Ornitz and P. J. Marie, "Fgf signaling pathways in endochondral and intramembranous bone development and human genetic disease," *Genes and Development*, vol. 16, pp. 1446–1465.
[14] H. Clevers, "Wnt/b-catenin signaling in development and disease," *Cell*, vol. 127, pp. 469–480, Nov. 2006.
[15] L. Hennighausen and G. W. Robinson, "Signaling pathways in mammary gland development," *Developmental Cell*, vol. 1, pp. 467–475, 2001.
[16] N. Pearcy *et al.*, "Complexity and robustness in hypernetwork models of metabolism," *Journal of Theoretical Biology*, vol. 406, pp. 99–104, Oct. 2016.
[17] Yoshida *et al.*, "A time-dependent phase shift in the mammalian unfolded protein response," *Developmental Cell*, vol. 4, pp. 265–271, Feb. 2003.
[18] B. Beutler, "Inferences, questions and possibilities in toll-like receptor signalling," *Nature*, vol. 430, pp. 2592–2618, Jul. 2004.
[19] Sachs *et al.*, "Causal protein-signaling networks derived from multiparamter single-cell data," *Science*, vol. 308, pp. 523–529, Apr. 2005.
[20] R. A. Segal and M. E. Greenberg, "Intracellular signaling pathways activated by neurotrophic factors," *Annual Reviews in Neuroscience*, vol. 19, pp. 463–489, 1996.
[21] M. Schnare *et al.*, "Toll-like receptors control activation of adaptive immune responses," *Nature Immunology*, vol. 2, no. 10, pp. 947–950, Oct. 2001.
[22] S. Klamt *et al.*, "A methodology for the structural and functional analysis of signaling and regulatory networks," *BMC Bioinformatics*, Feb. 2006.
[23] H. Einat *et al.*, "The role of the extracellular signal-regulated kinase signaling pathway in mood modulation," *The Journal of Neuroscience*, vol. 23, no. 19, pp. 7311–7316, Aug. 2003.
[24] Y. Sun *et al.*, "Discovering time-dependent shortest path on traffic graph for drivers towards green driving," *Journal of Network and Computer Applications*, vol. 83, pp. 204–212, Oct. 2015.
[25] X. Hu *et al.*, "A constrained time-dependent k shortest paths algorithm addressing overlap and travel time deviation," *International Journal of Transportation Science and Technology*, vol. 4, no. 4, pp. 371–394, 2015.
[26] I. Chabini, "Discrete dynamic shortest path problems in transportation applications: Complexity and algorithms with optimal run time," *Transportation Research Records*, vol. 1645, pp. 170–175, 1998.
[27] A. Idri *et al.*, "A new time-dependent shortest path algorithm for multimodal transportation network," *Procedia Computer Science*, vol. 109, pp. 692–697, 2017.
[28] A. A. Prakash and K. K. Srinivasan, "Finding the most reliable strategy on stochastic and time-dependent transportation networks: A hypergraph based formulation," *Networks and Spatial Economics*, vol. 17, no. 3, pp. 809–840, Apr. 2017.
[29] G. Ausiello and L. Laura, "Directed hypergraphs: Introduction and fundamental algorithms- a survey," *Theoretical Computer Science*, vol. 658, pp. 293–306, Jan. 2017.
[30] G. S. Brodal and R. Jacob, "Time-dependent networks as models to achieve fast exact time-table queries," *Electronic Notes in Theoretical Computer Science*, vol. 92, pp. 3–15, 2004.
[31] W. Jigang *et al.*, "Algorithm for time-dependent shortest safe path on transportation networks," *Procedia Computer Science*, vol. 4, pp. 958–966, 2011.
[32] Y. Pan *et al.*, "Finding reliable shortest path in stochastic time-dependent network," *Procedia - Social and Behavioral Sciences*, vol. 96, pp. 451–460, Nov. 2013.
[33] H. Aljazzar and S. Leue, "K*: A heuristic search algorithm for finding the k shortest paths," *Artificial Intelligence*, vol. 175, pp. 2129–2154, Jul. 2011.
[34] X. Cai *et al.*, "Time-varying shortest path problems with constraints," *Networks*, vol. 29, no. 3, pp. 141–150, May 1998.
[35] D. Pretolani, "Finding hypernetworks in directed hypergraphs," *European Journal of Operational Research*, vol. 230, no. 2, pp. 226–230, Oct. 2013.
[36] D. Pretolani, "A directed hypergraph model for random time dependent shortest paths," *European Journal of Operational Research*, vol. 123, pp. 315–324, Sep. 2000.
[37] L. R. Nielsen *et al.*, "K shortest paths in stochastic time-dependent networks," vol. 5, pp. 1–26, Nov. 2004.
[38] A. Orda and R. Rom, "Shortest-path and minimum-delay algorithms in networks with time-dependent edge-length," *Journal of the Association for Computing Machinery*, vol. 37, no. 3, pp. 607–625, Jul. 1990.
[39] U. Dogrusoz *et al.*, "Algorithms for effective querying of compound graph-based pathway databases," *BMC Bioinformatics*, vol. 10, no. 1, p. 376, Nov. 2009.
[40] N. Novere *et al.*, "The systems biology graphical notation," *Nature Biotechnology*, vol. 27, no. 1, pp. 735–741, Aug. 2009.
[41] C. Chiu and J. Zhan, "Deep learning for link prediction in dynamic networks using weak estimators," *IEEE Access*, vol. 6, pp. 35 937 – 35 945, 2018.
[42] M. Bhaduri and J. Zhan, "Using empirical recurrences rates ratio for time series data similarity," *IEEE Access*, vol. 6, pp. 30 855 – 30 864, 2018.
[43] J. Z. Jimmy Ming-Tai Wu and S. Chobe, "Mining association rules for low frequency itemsets," *PLoS ONE*, vol. 13, p. e0198066, 2018.
[44] J. M.-T. W. Payam Ezatpoor, Justin Zhan and C. Chiu, "Finding top-k dominance on incomplete big data using mapreduce framework," *IEEE Access*, vol. 6, pp. 7872–7887, 2018.
[45] P. Chopade and J. Zhan, "Towards a framework for community detection in large networks using game-theoretic modeling," *IEEE Transactions on Big Data*, vol. 3, pp. 276–288, 2017.
[46] J. Z. Moinak Bhaduri and C. Chiu, "A weak estimator for dynamic systems," *IEEE Access*, vol. 5, pp. 27 354–27 365, 2017.
[47] M. Pirouz and J. Zhan, "Toward efficient hub-less real time personalized pagerank," *IEEE Access*, vol. 5, pp. 26 364–26 375, 2017.
[48] C. C. Moinak Bhaduri, Justin Zhan and F. Zhan, "A novel online and non-parametric approach for drift detection in big data," *IEEE Access*, vol. 5, pp. 15 883–15 892, 2017.
[49] J. Z. Carter Chiu and F. Zhan, "Uncovering suspicious activity from partially paired and incomplete multimodal data," *IEEE Access*, vol. 5, pp. 13 689–13 698, 2017.
[50] R. Ahn and J. Zhan, "Using proxies for node immunization identification on large graphs," *IEEE Access*, vol. 5, pp. 13 046–13 053, 2017.
[51] J. Z. Jimmy Ming-Tai Wu and J. C.-W. Lin, "Ant colony system sanitization approach to hiding sensitive itemsets," *IEEE Access*, vol. 5, pp. 10 024–10 039, 2017.
[52] J. Zhan and B. Dahal, "Using deep learning for short text understanding," *Journal of Big Data*, vol. 4, 2017.
[53] S. G. Justin Zhan and S. P. K. Parsa, "Identification of top-k nodes in large networks using katz centrality," *Journal of Big Data*, vol. 4, 2017.
[54] G. S.-E. V. Justin Zhan, Timothy Rafalski, "Vaccination allocation in large dynamic networks," *Journal of Big Data*, vol. 4, 2016.
[55] H. Selim and J. Zhan, "Towards shortest path identification on large networks," *Journal of Big Data*, vol. 3, 2016.
[56] G. Ausiello *et al.*, "Optimal transversal of directed hypergraphs," *ICSI Technical Report TR-92-073*, Sep. 1992.
[57] M. Thakur and R. Tripathi, "Linear connectivity problems in directed hypergraphs," *Theoretical Computer Science*, vol. 410, pp. 2592–2618, Feb. 2009.
[58] R. Cambini *et al.*, "Flows on hypergraphs," *Mathematical Programming*, vol. 78, no. 2, pp. 195–217, Aug. 1993.
[59] C. Schaefer *et al.*, "Pid: The pathway interaction database," *Nucleic Acids Res.*, vol. 37, 2009.
[60] A. Fabregat *et al.*, "The reactome pathway knowledgebase," *Nucleic Acids Research*, vol. 46, no. 1, pp. 649–655, Jan. 2018.
[61] A. Zhang *et al.*, "Deletion of the telomerase reverse transcriptase gene and haploinsufficiency of telomere maintenance in cri du chat syndrome," *American Journal of Human Genetics*, vol. 72, no. 4, pp. 940–948, Apr. 2003.
[62] H. Lei and A. Kazlauskas, "A reactive oxygen species-mediated, self-perpetuating loop persistently activates platelet-derived growth factor receptor ?" *Molecular and Cellular Biology*, vol. 34, no. 1, pp. 110–122, Jan. 2014.
[63] E. Demir *et al.*, "The biopax community standard for pathway data sharing," *Nature Biotechnology*, vol. 28, no. 9, 935-942.
[64] E. Demir *et al.*, "Using biological pathway data with paxtools," *PLOS Computational Biology*, vol. 9, no. 9, 2013.
[65] A. Gresset *et al.*, "Mechanism of phosphorylation-induced activation of phospholipase c-? isozymes," *The Journal of Biological Chemistry*, vol. 285, no. 4, pp. 35 836–35 847, Aug. 2010.
[66] B. Henglein *et al.*, "Structure and cell cycle-regulated transcription of the human cyclin a gene," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 91, no. 12, Jun. 1994.
[67] T. Matsuki *et al.*, "Reduction of crk and crkl expression blocks reelin-induced dendritogenesis," *Journal of Cell Science*, vol. 121, no. 11, pp. 1869–1875, Jun. 2008.
[68] S. Tanaka *et al.*, "C3g, a guanine nucleotide-releasing protein expressed ubiquitously, binds to the src homology 3 domains of crk and grb2/ash proteins," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 91, no. 8, pp. 3443–3447, Apr. 1994.
[69] P. Holland *et al.*, "Mkk7 is a stress-activated mitogen-activated protein kinase functionally related to hemipterous," *The Journal of Biological Chemistry*, vol. 272, no. 40, pp. 24 994–24 998, Oct. 1997.

[70] C.-F. Zheng and K.-L. Guan, "Activation of mek family kinases requires phosphorylation of two conserved ser/thr residues," *The EMBO Journal*, vol. 13, no. 5, pp. 1123–1131, Mar. 1994.
[71] S. Bhagwat, "Kinase inhibitors for the treatment of inflammatory and autoimmune disorders," *Purinergic Signalling*, vol. 5, no. 1, pp. 107–115, Jun. 2008.
[72] J. Fantini and N. Yahi, *Brain lipids in synaptic function and neurological disease: clues to innovative therapeutic strategies for brain disorders*. Academic Press, 2015.
[73] L. Miller and R. Wells, "Deoxyribonucleoside triphosphate stimulation of exonucleolytic activity of the micrococcus luteus deoxyribonucleic acid polymerase," *The Journal of Biological Chemistry*, vol. 247, no. 9, pp. 2675–2681, May 1972.
[74] N. Menzel *et al.*, "Map-kinase regulated cytosolic phospholipase a2 activity is essential for production of infectious hepatitis c virus particles," *PLoS Pathogens*, vol. 8, no. 7, Jul. 2012.

Michael R. Schwob is an undergraduate student at the University of Nevada, Las Vegas, majoring in mathematics and economics. Due to his diverse research experience in astrophysics, bioinformatics, big data, and economics, Michael was named a 2019 Goldwater Scholar. As a research assistant under the advisement of Dr. Justin Zhan in UNLV's Big Data Hub, he explores his research interests in game theory, Bayesian analysis, biostatistics, hypergraph theory, machine learning, and genomics.

Dr. Justin Zhan is an ARA Scholar and Professor of Data Science, Department of Computer Science and Engineering, University of Arkansas. This work has been mainly conducted while he was at University of Nevada Las Vegas, where he was the director of the Big Data Hub and a professor in the Department of Computer Science, College of Engineering, University of Nevada, Las Vegas. His research interests include Big Data, Information Assurance, Social Computing, and Health Sciences. He was a steering chair of ASE/IEEE International Conference on Social Computing (SocialCom), ASE/IEEE International Conference on Privacy, Security, Risk and Trust (PASSAT), and ASE/IEEE International Conference on BioMedical Computing (BioMedCom). He has been an editor-in-chief of International Journal of Privacy, Security and Integrity, and International Journal of Social Computing and Cyber-Physical Systems. He has served as a conference general chair, a program chair, a publicity chair, a workshop chair, and a program committee member for 200 international conferences; he has also served as an editor-in-chief, editor, associate editor, guest editor, editorial advisory board member, and editorial board member for thirty journals. He has published 230 articles in peer-reviewed journals and conferences and delivered more than thirty keynote speeches and invited talks. He has been involved more than 50 projects as a Principal Investigator (PI) or a Co-PI, which were funded by the National Science Foundation, Department of Defense, National Institute of Health, etc.

Aeren Dempsey is an undergraduate student majoring in Computer Science at the University of Nevada, Las Vegas. He is currently working as a research assistant at the Big Data Hub at UNLV, under the supervision of Dr. Justin Zhan. His main research interests are machine learning, data science and deep learning.