SOME FUNDAMENTAL OPERATIONS ON MULTIMODAL NETWORKS IN BIOLOGY

ALLAN A. SIOSON
Department of Computer Science
Virginia Polytechnic Institute and State University
660 McBryde Hall, Blacksburg, VA 24061-0106, USA
e-mail: asioson@vt.edu

and

LENWOOD S. HEATH
Department of Computer Science
Virginia Polytechnic Institute and State University
660 McBryde Hall, Blacksburg, VA 24061-0106, USA
e-mail: heath@vt.edu

ABSTRACT

A multimodal network is a mathematical formalism based on hypergraphs that can model the informational, topological, and dynamic aspects of biological networks. In this paper, we discuss some fundamental operations on a multimodal network system. We also present how a multimodal network captures pertinent information from biology literature and how its structure and data can be stored in a relational database.

Keywords: multimodal networks, biological networks, hypergraphs, relational database.

1. Introduction

Biological networks arise from the study of biological phenomena such as those occuring in cells in response to biotic or abiotic stress. Various mathematical representations are employed to model both hypothetical and known components of biological networks including their dynamic and static behaviors. For example, the directed graph is the underlying mathematical representation of some popular computational frameworks used to model biological networks. Chief among these are Bayesian networks and Boolean networks. Pe'er, *et al.*[16] use a Bayesian network to infer finer structure of gene interactions using perturbed gene expression data of *Saccharomyces cerevisiae* while Albert and Othmer [1] use a Boolean network to model regulatory interactions to predict the expression pattern of segment polarity genes in *Drosophila melanogaster*. Similarly, directed hypergraphs are used to depict the metabolic network diagrams in popular metabolic network databases like KEGG [13] and MetaCyc [15]. However when depicting biological networks, specially those inferred from published works in biology, directed graphs and hypergraphs may be too limiting. Appropriate mathematical representations are crucial in computational frameworks designed to infer new information or hypothesis from the modeled biological network. In this paper, we offer the multimodal network formalism to model biological networks.

We note that our research on multimodal networks is a work in progress. Hence, there's a possibility that some definitions and concepts presented here will be updated appropriately.

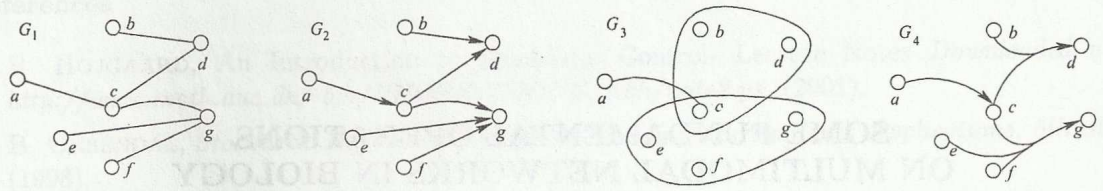

Figure 1: G_1 is a graph, G_2 is an directed graph, G_3 is a hypergraph, and G_4 is a directed hypergraph.

1.1. Basic Definitions

The following are graph theory definitions taken from Berge [4, 5]; Harris, Hirst and Mossinghoff [10]; and Gallo, Longo, and Pallotino [9].

Definition 1 *A graph is a tuple (V, E) with V as the set of vertices and E as the set of edges where each $e \in E$ is pair of vertices of V such that $e = (u, v) = (v, u)$ for $u, v \in V$.*

Definition 2 *A directed graph is a tuple (V, E) with V as the set of vertices and E as the set of directed edges where each $e \in E$ is an ordered pair of vertices of V. Let $e = (u, v)$ where $u, v \in V$. Vertices u and v are the tail and head of e respectively.*

Each vertex in both 'undirected' and directed graphs is represented by a circle. Each edge in a directed graph is represented by an arrow with the arrow head pointing to its head vertex. In an 'undirected' graph, edges are represented by a continuous curve that connects two vertices. In Figure 1, graphs G_1 and G_2 are 'undirected' and directed graphs respectively. Each has seven vertices and six edges.

Graphs naturally capture binary relationships between vertices. To represent relationships that involve more than two vertices, we use hyperedges. A hyperedge is a non-empty subset of the set of vertices and naturally involve more than two vertices in a relationship. Intuitively, a collection of hyperedges forms a *hypergraph.*

Definition 3 *A hypergraph is a tuple (V, E) with V as the set of vertices and E a set of hyperedges or non-empty subsets of vertices in V such that for all $e \in E$,*

$$\bigcup_{e \in E} e = V \quad \text{where } e \neq \varnothing, \quad e \subseteq V.$$

Definition 4 *A directed hypergraph is a tuple (V, E) with V as the set of vertices and E as the set of directed hyperedges where each hyperedge $e \in E$ is an ordered pair of disjoint subsets of vertices in V. Let $e = (T, H)$ where $T, H \subseteq V$ and $T \cap H = \varnothing$. T and H are the tail and head of e respectively and cannot be both $\varnothing$.*

Vertices in a hypergraph are represented by circles. Hyperedges with more than two vertices are represented by a simple closed curve enclosing its vertices. Hyperedges with two vertices are represented by a continuous curve connecting its vertices. A directed hyperedge on the other hand is depicted by a 'merged' set of arrows that connects the tail vertices to the head vertices. In Figure 1, G_3 is a hypergraph while G_4 is a directed hypergraph. Each has seven vertices and three hyperedges.

Graphs are special hypergraphs. A hypergraph whose every edge has a cardinality of 2 is a graph. Similarly, directed graphs are special directed hypergraphs. If a directed hypergraph $G = (V, E)$ has $|T| = |H| = 1$ for all $e = (T, H) \in E$ where $T, H \subseteq V$, then G is also a directed graph.

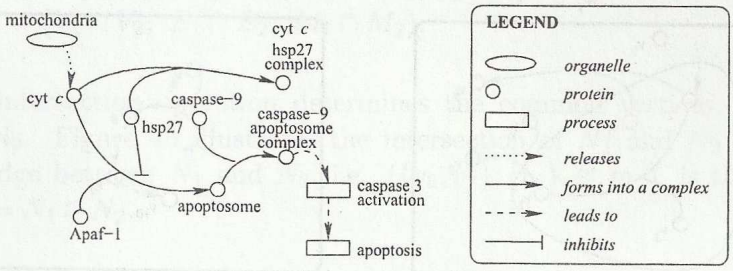

Figure 2: A portion of the proposed pathway for HSP mediated modulation of apoptosis by Xanthodakis and Nicholson [20].

1.2. A Sample Biological Network

Graphs and hypergraphs are frequently used to represent biological networks. The vertices may represent genes, proteins, enzymes, metabolites, small molecules, or even a biological process in a directed graph or directed hypergraph rendering of a biological network. The edges on the other hand represent relationships between biological entities.

Figure 2 is adapted from the proposed model of pathways for the heat shock protein (HSP) modulation of apoptosis by Xanthoudakis and Nicholson [20]. In their model, the proteins cytochrome c (cyt c) and Apaf-1 form a complex called apoptosome which in turn forms a complex with caspase-9 in order to activate the latter. This apoptosome-caspase-9 complex plays a role in caspase-3 activation which leads to cell death or apoptosis.

The complete Xanthoudakis and Nicholson model is a combination of experimental results of different researchers. For example, the result showing that hsp27 inhibits the formation of the apoptosome thereby putting a break in the cascade to apoptosis is due to Bruey, *et al.* [6]. Their results show that hsp27 forms a complex with the cyt c released from mitochondria. The capture of cyt c by hsp27 leaves Apaf-1 with nothing to pair with to form the apoptosome. This in turn prevents apoptosis from happening.

Observe that in Figure 2, the role of hsp27 in preventing apoptosome formation is linked to the hyperedge ({cyt c, Apaf-1}, {apoptosome}) and not to any of the vertices of the latter. This poses a difficulty in strictly using a directed hyperedge to depict this kind of relationship. The following sections show how to represent this relationship using the multimodal network formalism.

2. Multimodal Networks

2.1. Definition and Basic Operations

Definition 5 *A multimodal network (MMN) is a triple (V, E, M) where V is a set of vertices, E is a set of modal hyperedges, and M is a set of modes. A modal hyperedge $e \in E$ is an ordered 4-tuple consisting of three subsets of vertices in V and a mode in M. Let $e = (T, H, A, m)$ where $T, H, A \subseteq V$ and $m \in M$. The sets T and H, called the tail and head of e respectively, define the direction of the modal hyperedge. The set A, called the associate set in e, contains vertices that do not define the direction of e while m is the mode of e.*

We can view an MMN as a set of modal hyperedges. A *mode* is a type of hyperedge. Pertinent in the definition of an MMN is the set of modes M. When the associate set of a modal hyperedge is non-empty, we draw the implied arrow connecting the vertices of the head and tail sets and then enclose all vertices of the modal hyperedge in a closed curve. Whenever the associate set

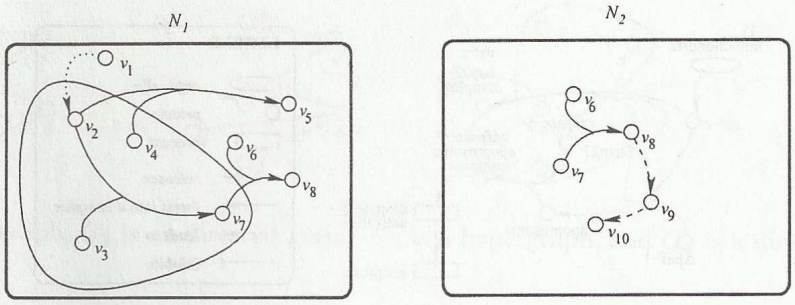

Figure 3: Depiction of multimodal networks N_1 and N_2. The dotted line, solid line, and dashed line represent modes m_1, m_2, and m_3 respectively.

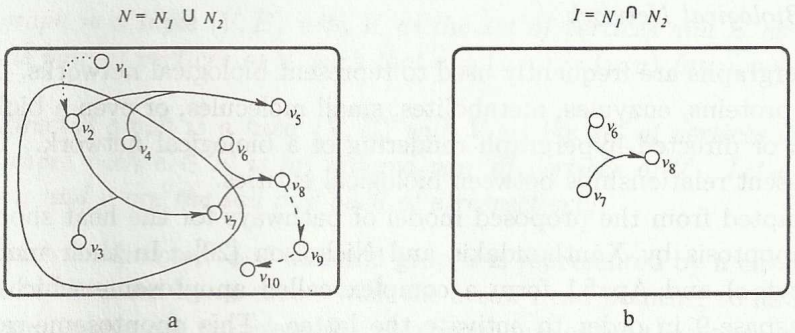

Figure 4: Resulting MMNs from the $\cup$ and $\cap$ operations

of a modal hyperedge is empty, we draw the modal hyperedge as we would draw a directed hyperedge.

In the subsequent discussions, we will use the term 'edge' to refer to either hyperedge or 'binary' edge for simplicity.

Example 1 Let $N_1 = (V_1, E_1, M_1)$ be an MMN where $V_1 = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8\}$, $E_1 = \{(\{v_1\}, \{v_2\}, \varnothing, m_1), (\{v_2, v_3\}, \{v_7\}, \{v_4\}, m_2), (\{v_2, v_4\}, \{v_5\}, \varnothing, m_2), (\{v_6, v_7\}, \{v_8\}, \varnothing, m_2)\}$, and $M_1 = \{m_1, m_2\}$.

Furthermore, let $N_2 = (V_2, E_2, M_2)$ be an MMN where $V_2 = \{v_6, v_7, v_8, v_9, v_{10}\}$, $E_2 = \{(\{v_6, v_7\}, \{v_8\}, \varnothing, m_2), (\{v_8\}, \{v_9\}, \varnothing, m_3), (\{v_9\}, \{v_{10}\}, \varnothing, m_3)\}$, and $M_2 = \{m_2, m_3\}$

Figure 3 shows how MMNs N_1 and N_2 are drawn. Observe that in one edge of N_1, specifically $(\{v_2, v_3\}, \{v_7\}, \{v_4\}, m_2)$, the vertex v_4 is part of the associate set and hence not connected by an arrow to any of the vertices v_2, v_3, and v_7. The other edges in N_1 and the edges in N_2 are not enclosed in closed curves since their respective associate sets are empty.

Definition 6 *Let $N_1 = (V_1, E_1, M_1)$ and $N_2 = (V_2, E_2, M_2)$ be MMNs. The union of N_1 and N_2 is $N_1 \cup N_2 = (V_1 \cup V_2,\ E_1 \cup E_2,\ M_1 \cup M_2)$*

Example 2 The union operation allows the 'integration' of two MMNs into a larger MMN. Figure 4a illustrates the union of N_1 and N_2 depicted in Figure 3. Observe that a common edge in N_1 and N_2, i.e. $(\{v_6, v_7\}, \{v_8\}, \varnothing, m_2)$, appears only once in $N = N_1 \cup N_2$.

Definition 7 *Let $N_1 = (V_1, E_1, M_1)$ and $N_2 = (V_2, E_2, M_2)$ be MMNs. The intersection of N_1 and N_2 is $N_1 \cap N_2 = (V_1 \cap V_2,\ E_1 \cap E_2,\ M_1 \cap M_2)$.*

Example 3 The intersection operation determines the common vertices, edges, and modes between two MMNs. Figure 4b illustrates the intersection of N_1 and N_2 depicted in Figure 3. The common edge between N_1 and N_2, i.e. $(\{v_6, v_7\}, \{v_8\}, \varnothing, m_2)$, is the only edge in the resulting MMN $I = N_1 \cap N_2$.

2.2. Using MMNs to Model Biological Networks

Biological networks typically show multiple types of relationships or interactions of biological entities in a single diagram. These relationships may not be binary at all. Furthermore, the role of each biological entity in various relationships may not be the same.

These nuances of biological networks can be represented in an MMN. In the latter, the biological entities can be depicted as vertices and the relationships or interactions as modal hyperedges. The modal hyperedges involving a specific biological entity may have different modes. Note that we can impose a one to one correspondence between each mode and relationship or interaction type. Furthermore, the possibly distinct roles of each biological entity in various relationships may be associated with the modal hyperedges.

If biological networks are represented as MMNs, then the union operation is a tool to integrate these separate biological networks into a larger MMN system. The intersection operation on the other hand allows identification of common interaction structures between two distinct MMNs. For example, suppose MMNs X and Y represent all biological pathway information about *Arabidopsis thaliana* (a widely used model plant) and *Saccharomyces cerevisiae* (budding yeast) respectively. The intersection operation would allow identification of common pathways (perhaps conserved pathways) between *Arabidopsis thaliana* and *Saccaromyces cerevisiae.*

3. Labeled Multimodal Network and Relational Database

3.1. Labeled Multimodal Network

A labeled MMN (V, E, M) is an MMN where each $v \in V$, $e \in E$, $m \in M$, and $(v, e, m) \in V \times E \times M$ have named attributes. We call the set $V \times E \times M$ as the *incidence set.* Attributes add meaning to the components of an MMN. Furthermore, attributes elucidate the respective roles of each vertex in different relationships.

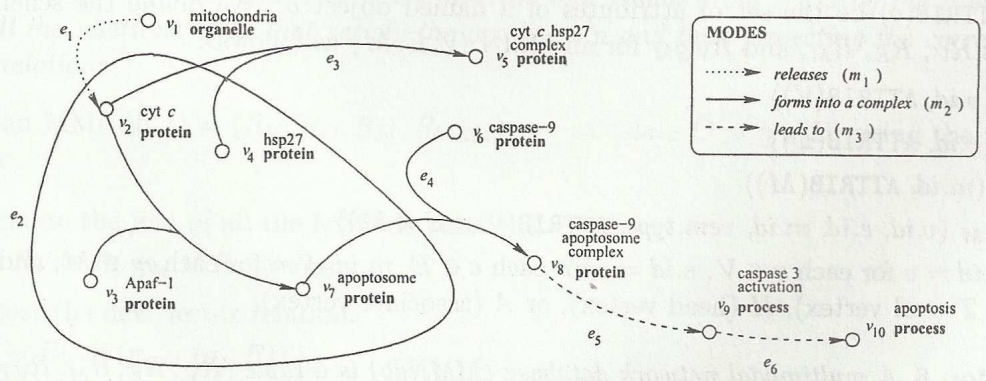

Figure 5: The labeled multimodal network N

Consider the MMN $N = N_1 \cup N_2$ depicted in Figure 4a. N is in fact a representation of the partial pathway of heat shock protein modulation of apoptosis shown in Figure 2. Figure 5 shows that this is the case. In the figure, each vertex has an ID, a name, and a type; each mode has an ID and a name; and each edge has an ID. It is difficult however to show the role of each vertex in a particular edge. In Figure 2, hsp27 is an 'inhibitor' of edge e_2 (formation of apoptosome from cyt c and Apaf-1) while it is an 'input' in edge e_3 (formation of *cyt c* - hsp27 complex). The role of each vertex in a particular edge with a specific mode is an attribute of the incidence set.

3.2. Some Relational Database Concepts

Information associated with the components of a multimodal network can be stored in a relational database. Relational databases are built on top of the relational theory that Codd [7] developed in 1970. Modern discussions on relational theory and relational database can be read in Silberschatz, *et al.* [17]; Lewis, *et al.* [14]; and other recent books on database design.

A *relation* is composed of a *schema* and a *relation instance*. A schema consists of a unique name of the relation across the database, a set of attribute names with their associated domain names, and a set of constraints on which tuples can appear as an instance of the relation. A relation instance on the other hand is a table with rows (tuples) and named columns (attributes). Each row has the same number of columns (called the *arity* of the relation). The *cardinality* of a relation is the number of tuples of a relation. In a relation, tuples are not duplicated.

The operations σ (selection), π (projection), and $\bowtie$ (natural join) are some of the operators that can be used to manipulate relations. The σ operator is used to select tuples of a relation R subject to a specific boolean predicate p. For example, $\sigma_p R$ is a relation consisting of tuples of R that satisfy p. The π operator on the otherhand project a relation by selecting columns from a given relation. For example, suppose a, b, and c are attributes of relation R. Then, $\pi_{a,c} R$ is a relation that results from removing attribute 'b' from R. The $\bowtie$ operator takes two relations R and Q and produce a relation with tuples formed by concatenating tuples from R and Q whose values in their common attributes match up. If R and Q have no common set of attributes, $R \bowtie Q$ is empty.

3.3. Multimodal Network Database

Information about the components of an MMN can be stored in four relations. Data on vertices, edges, and modes can be stored in a vertex relation R_V, edge relation R_E, and mode relation R_M respectively. Furthermore, the details on the participation of a vertex in a particular edge with a specific mode may be kept in an incidence relation R_{VEM}. This relation should indicate whether a vertex is a head vertex, a tail vertex, or an associate vertex.

Let $\texttt{ATTRIB}(o)$ be the set of attributes of a named object o. We define the schema of the relations R_V, R_E, R_M, and R_{VEM} for an MMN (V, E, M) as follows:

R_V (*v.id*, $\texttt{ATTRIB}(V)$)

R_E (*e.id*, $\texttt{ATTRIB}(E)$)

R_M (*m.id*, $\texttt{ATTRIB}(M)$)

R_{VEM} (*v.id*, *e.id*, *m.id*, *vem.type*, $\texttt{ATTRIB}(V \times E \times M)$)

where $v.id = v$ for each $v \in V$, $e.id = e$ for each $e \in E$, $m.id = m$ for each $m \in M$, and *vem.type* is either T (tail vertex), H (head vertex), or A (associate vertex).

Definition 8 *A multimodal network database (MMNdb) is a tuple* (R_V, R_E, R_M, R_{VEM}) *where* R_V, R_E, R_M, *and* R_{VEM} *are the vertex relation, edge relation, mode relation, and incidence relation, respectively derived from the multimodal network* (V, E, M).

Example 4 Let $N = (V, E, M)$ be the labeled MMN depicted in Figure 5 and let D_N be the database that corresponds to N. The database tables of D_N follow.

$R_V =$

v.id	v.name	v.type
v_1	mitochondrion	organelle
v_2	cyt *c*	protein
v_3	Apaf-1	protein
v_4	hsp 27	protein
v_5	hsp 27, cyt *c* complex	protein
v_6	caspase-9	protein
v_7	apoptosome	protein
v_8	apoptosome caspase-9 complex	protein
v_9	caspase-3 activation	process
v_{10}	apoptosis	process

$R_E =$

e.id
e_1
e_2
e_3
e_4
e_5
e_6

$R_M =$

m.id	m.name
m_1	release
m_2	forms into a complex
m_3	leads to

$R_{VEM} =$

v.id	*e.id*	*vem.type*	*m.id*	vem.role
v_1	e_1	T	m_1	input
v_2	e_1	H	m_1	output
v_2	e_2	T	m_2	input
v_3	e_2	T	m_2	input
v_4	e_2	A	m_2	inhibitor
v_7	e_2	H	m_2	output
v_2	e_3	T	m_2	input
v_4	e_3	T	m_2	input
v_5	e_3	H	m_2	output
v_6	e_4	T	m_2	input
v_7	e_4	T	m_2	input
v_8	e_4	H	m_2	output
v_8	e_5	T	m_3	input
v_9	e_5	H	m_3	output
v_9	e_6	T	m_3	input
v_{10}	e_6	H	m_3	output

In this example, the role of a vertex in an edge is set to 'input' or 'output' when its type is 'T' or 'H' respectively. Vertex v_4 (hsp27) is an 'inhibitor' in edge e_2.

3.4. Computations on a Multimodal Network Database

Given an MMNdb (R_V, R_E, R_M, R_{VEM}), the components of a corresponding MMN (V, E, M) can be computed using the σ and π operators. The steps to perform the calculation follows.

1. Project the elements of V from R_V.
   $V = \pi_{v.id}\ R_V$
2. Project the elements of M from R_M.
   $M = \pi_{m.id}\ R_M$
3. Extract the edges from R_E and R_{VEM}.
   $E = \{(T_j, H_j, A_j, m_j) \mid e_j \in \pi_{e.id}\ R_E\}$, where
   $T_j = \pi_{v.id}\ (\sigma_{vem.type='T'\ \text{AND}\ e.id=e_j}\ R_{VEM})$,
   $H_j = \pi_{v.id}\ (\sigma_{vem.type='H'\ \text{AND}\ e.id=e_j}\ R_{VEM})$,
   $A_j = \pi_{v.id}\ (\sigma_{vem.type='A'\ \text{AND}\ e.id=e_j}\ R_{VEM})$, and
   $m_j = \pi_{m.id}\ (\sigma_{e.id=e_j}\ R_{VEM})$.

Definition 9 *Let $D = (R_V, R_E, R_M, R_{VEM})$ be an MMNdb. We define the MMNdb selection function γ as $O = \gamma_p(D)$ where O is an MMNdb that results from selecting the tuples of the join of all the relations of D that satisfy the condition p and then projecting the corresponding MMNdb relations.*

Given an MMNdb $D = (R_V, R_E, R_M, R_{VEM})$, we calculate $O = \gamma_p(D)$ using the following procedure.

1. Calculate the join of all the MMNdb relations.
   $R = R_V \bowtie R_{VEM} \bowtie R_E \bowtie R_M$
2. Project the new vertex relation.
   $R'_V = R_V \bowtie (\pi_{v.id}\ (\sigma_p\ R))$
3. Project the new edge relation.
   $R'_E = R_E \bowtie (\pi_{e.id}\ (\sigma_p\ R))$

4. Project the new mode relation.
   $R'_M = R_M \bowtie (\pi_{m.id}\ (\sigma_p\ R))$
5. Project the new incidence relation.
   $R'_{VEM} = R_{VEM} \bowtie (\pi_{v.id,e.id,vem.type,m.id}\ (\sigma_p\ R))$

The desired MMNdb is $O = (R'_V, R'_E, R'_M, R'_{VEM})$.

Example 5 We illustrate the γ operator using the MMNdb D_N defined in example 4. Suppose we want the portion of D_N that involves only the proteins. Let this portion be called O. Formally, this is accomplished by the statement $O = \gamma_p D_N$ where p is {*v.type*='protein'}. To compute O, the join of all relations in D_N which we call R is calculated first.

$$R = R_V \bowtie R_{VEM} \bowtie R_E \bowtie R_M$$

$=$

v.id	v.name	v.type	*e.id*	*vem.type*	vem.role	*m.id*	m.name
v_1	mitochondrion	organelle	e_1	T	input	m_1	releases
v_2	cyt *c*	protein	e_1	H	output	m_1	releases
v_2	cyt *c*	protein	e_2	T	input	m_2	forms into a complex
v_3	Apaf-1	protein	e_2	T	input	m_2	forms into a complex
v_4	hsp 27	protein	e_2	A	inhibitor	m_2	forms into a complex
v_7	apoptosome	protein	e_2	H	output	m_2	forms into a complex
v_2	cyt *c*	protein	e_3	T	input	m_2	forms into a complex
v_4	hsp 27	protein	e_3	T	input	m_2	forms into a complex
v_5	hsp 27, cyt *c* complex	protein	e_3	H	output	m_2	forms into a complex
v_6	caspase-9	protein	e_4	T	input	m_2	forms into a complex
v_7	apoptosome	protein	e_4	T	input	m_2	forms into a complex
v_8	apoptosome caspase-9 complex	protein	e_4	H	output	m_2	forms into a complex
v_8	apoptosome caspase-9 complex	protein	e_5	T	input	m_3	leads to
v_9	caspase-3 activation	process	e_5	H	output	m_3	leads to
v_9	caspase-3 activation	process	e_6	T	input	m_3	leads to
v_{10}	apoptosis	process	e_6	H	output	m_3	leads to

Next we calculate $\sigma_p R$.

$\sigma_p R =$

v.id	v.name	v.type	*e.id*	*vem.type*	vem.role	*m.id*	m.name
v_2	cyt *c*	protein	e_1	H	output	m_1	releases
v_2	cyt *c*	protein	e_2	T	input	m_2	forms into a complex
v_3	Apaf-1	protein	e_2	T	input	m_2	forms into a complex
v_4	hsp 27	protein	e_2	A	inhibitor	m_2	forms into a complex
v_7	apoptosome	protein	e_2	H	output	m_2	forms into a complex
v_2	cyt *c*	protein	e_3	T	input	m_2	forms into a complex
v_4	hsp 27	protein	e_3	T	input	m_2	forms into a complex
v_5	hsp 27, cyt *c* complex	protein	e_3	H	output	m_2	forms into a complex
v_6	caspase-9	protein	e_4	T	input	m_2	forms into a complex
v_7	apoptosome	protein	e_4	T	input	m_2	forms into a complex
v_8	apoptosome caspase-9 complex	protein	e_4	H	output	m_2	forms into a complex
v_8	apoptosome caspase-9 complex	protein	e_5	T	input	m_3	leads to

Using the procedure described earlier, we can then project the corresponding MMNdb relations from the relation $\sigma_p R$. These relations shown below comprise the MMNdb O.

$R'_V =$

v.id	v.name	v.type
v_2	cyt *c*	protein
v_3	Apaf-1	protein
v_4	hsp 27	protein
v_5	hsp 27, cyt *c* complex	protein
v_6	caspase-9	protein
v_7	apoptosome	protein
v_8	apoptosome caspase-9 complex	protein

$R'_E =$

e.id
e_1
e_2
e_3
e_4
e_5

$R'_M =$

m.id	m.name
m_1	release
m_2	forms into a complex
m_3	leads to

$R'_{VEM} =$

v.id	*e.id*	*vem.type*	*m.id*	vem.role
v_2	e_1	H	m_1	output
v_2	e_2	T	m_2	input
v_3	e_2	T	m_2	input
v_4	e_2	A	m_2	inhibitor
v_7	e_2	H	m_2	output
v_2	e_3	T	m_2	input
v_4	e_3	T	m_2	input
v_5	e_3	H	m_2	output
v_6	e_4	T	m_2	input
v_7	e_4	T	m_2	input
v_8	e_4	H	m_2	output
v_8	e_5	T	m_3	input

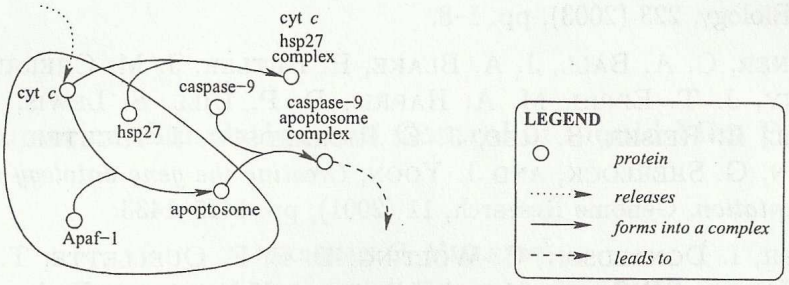

Figure 6: The MMN that corresponds to MMNdb O

The operator γ provides a way to focus on a selected portion of an MMN indirectly through its corresponding MMNdb. For example, the corresponding MMN for MMNdb O is shown in Figure 6. We note that there are edges with an empty tail or empty head in it. These empty sets depict unknown entities in the relationship that the edge represents.

4. Conclusions

Published models based on directed graphs and directed hypergraphs could be viewed as special multimodal networks. An MMN $N = (V, E, M)$ with $|M| = 1$ and modal hyperedges $e = (T, H, A, m) \in E$ such that $A = \varnothing$, is a directed hypergraph. Furthermore, if $|H| = |T| = 1$, we have a directed graph. It is possible therefore to integrate such models in a computational framework based on multimodal networks. For example, the network model of HSP mediated regulation of apoptosis by Xanthoudakis and Nicholson [20] can be represented as an MMN.

Each vertex and modal hyperedge in an MMN has named attributes. Information on each biological entity could be kept in an MMNdb system. Since every MMN has a corresponding MMNdb, data such as reference citations and experiment details kept in relational databases can be integrated. The stochastic parameters of possibly subsumed Bayesian or Boolean network models can also be part of the attributes of the vertices in an MMN. Furthermore, this provides an opportunity to include information provided in online databases such as metabolic pathway databases KEGG [11, 13] and MetaCyc [12, 15]; controlled vocabularies like Gene Ontology [2, 19]; protein protein interaction like BIND [3, 18]; pathway models from the literature; and experimental data in an MMN system.

The MMN basic operators $\cup$ and $\cap$ and MMNdb selection operator γ are tools to enrich an MMN system given the information from several MMNs.

Acknowledgements

The authors gratefully acknowledge the support of the United States National Science Foundation (NSF), Grant ITR-0219322. This work on multimodal networks is one of the NSF funded research areas of the Expresso research group [8] at Virginia Tech. We also thank John Paul Vergara for helpful comments that improved the exposition.

References

[1] R. Albert and H. G. Othmer, *The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in drosophila melanogaster*, Journal of

Theoretical Biology, 223 (2003), pp. 1–8.

[2] M. Ashburner, C. A. Ball, J. A. Blake, H. Butler, J. M. Cherry, J. Corradi, K. Dolinsky, J. T. Eppig, M. A. Harris, D. P. Hill, S. Lewis, B. Marshall, C. Mungall, L. Reiser, S. Ree, J. E. Richardson, J. Richter, M. Ringwald, G. M. Rubin, G. Sherlock, and J. Yoon, *Creating the gene ontology resource: design and implementation*, Genome Research, 11 (2001), pp. 1425–1433.

[3] G. D. Bader, I. Donaldson, C. Wolting, B. F. F. Ouellette, T. Pawson, and C. W. V. Hogue, *BIND - the biomolecular interaction network database*, Nucleic Acids Research, 29 (2001), pp. 242–245.

[4] C. Berge, *Graphs*, North-Holland, 2nd revised edition ed., 1985.

[5] ——, *Hypergraphs*, North-Holland, 1989.

[6] J.-M. Bruey, C. Ducasse, P. Bonniaud, L. Ravagnan, S. A. Susin, C. Diaz-Latoud, S. Gurbuxani, A.-P. Arrigo, G. Kroemer, E. Solary, and C. Garrido, *HSP27 negatively regulates cell death by interacting with cytochrome c*, Nature Cell Biology, 2 (2000), pp. 645–652.

[7] E. F. Codd, *A relational model of data for large shared data banks*, Communications of the ACM, 13 (1970), pp. 377–387.

[8] Expresso Bioinformatics Research Group website. `http://bioinformatics.cs.vt.edu/expresso/`.

[9] G. Gallo, G. Longo, S. Pallottino, and S. Nguyen, *Directed hypergraphs and applications*, Discrete Applied Mathematics, 42 (1993), pp. 177–201.

[10] J. M. Harris, J. L. Hirst, and M. J. Mossinghoff, *Combinatorics and Graph Theory*, Springer, 2000.

[11] M. Kanehisa, S. Goto, S. Kawashima, and A. Nakaya, *The KEGG databases at GenomeNet*, Nucleic Acids Research, 30 (2002), pp. 42–46.

[12] P. D. Karp, M. Riley, S. M. Paley, and A. Pellegrini-Toole, *The MetaCyc database*, Nucleic Acids Research, 30 (2002), pp. 59–61.

[13] Kyoto Encyclopedia of Genes and Genomes (KEGG) website. `http://www.kegg.org/`.

[14] P. M. Lewis, M. Kifer, and A. J. Bernstein, *Database and transaction processing*, Pearson Addison Wesley, 2001.

[15] MetaCyc website. `http://metacyc.org/`.

[16] D. Pe'er, A. Regev, G. Elidan, and N. Friedman, *Inferring subnetworks from perturbed expression profiles*, Bioinformatics, 17 (2001).

[17] A. Silberschatz, H. F. Korth, and S. Sudarshan, *Database system concepts*, McGraw-Hill Science/Engineering/Math, 4th ed., 2001.

[18] The Biomolecular Interaction Network Database website. `http://bind.ca/`.

[19] The Gene Ontology (GO) Consortium website. `http://www.geneontology.org/`.

[20] S. Xanthoudakis and D. W. Nicholson, *Heat shock proteins as death determinants*, Nature Cell Biology, 2 (2000), pp. E163–E165.