Chapter 2

THE ALGEBRAIC STRUCTURE OF METAGRAPHS

In Chapter 1, the notion of a metagraph was introduced informally, using visual depictions and descriptions. In this chapter, the formal structure of a metagraph is defined, and its basic properties are identified.

1. FORMAL REPRESENTATION OF A METAGRAPH

DEFINITION 2.1. The *generating set* of a metagraph is the set of *elements* $X = \{x_1, x_2, \ldots, x_n\}$, which represent variables of interest, and which occur in the edges of the metagraph.

DEFINITION 2.2. An *edge* e in a metagraph is a pair $e = \langle V_e, W_e \rangle \in E$ (where E is the set of edges) consisting of an *invertex* $V_e \subset X$ and an *outvertex* $W_e \subset X$, each of which may contain any number of elements. The different elements in the invertex (outvertex) are *coinputs* (*cooutputs*) of each other.

DEFINITION 2.3. A *metagraph* $S = \langle X, E \rangle$ is then a graphical construct specified by its generating set X and a set of edges E defined on the generating set.

DEFINITION 2.4. A *simple path* $h(x, y)$ from an element x to an element y is a sequence of edges $\langle e_1, e_2, \ldots, e_n \rangle$ such that

$x \in invertex(e_1)$,
$y \in outvertex(e_n)$, and
for all $e_i, i = 1, \ldots, n-1, outvertex(e_i) \cap invertex(e_{i+1}) \neq \varnothing$.

The *coinput* of x in the path (denoted $coinput(x)$) is the set of all other invertex elements in the path's edges that are not also in the outvertex of any edges in the path, and the *cooutput* of y (denoted $cooutput(y)$) is the set of all outvertex elements other than y. The *length* of a simple path is the number of edges in the path.

EXAMPLE 2.1. The metagraph in Figure 2.1 can be represented as follows:

$S = \langle X, E \rangle$, where
$X = \{Exp, Notes, Prof, Rev, Pri, Vol, Wage\}$, and

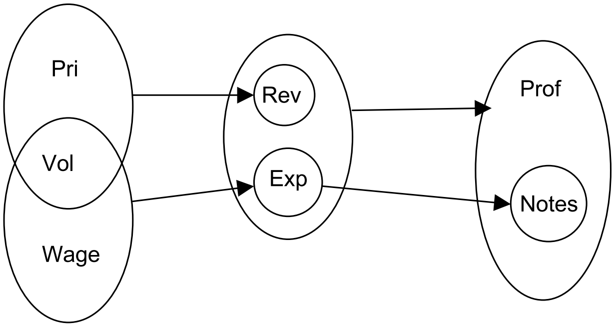

Figure 2.1. An example metagraph.

$$E = \{\langle\{Pri, Vol\}, \{Rev\}\rangle, \langle\{Vol, Wage\}, \{Exp\}\rangle, \langle\{Rev, Exp\},$$
$$\{Prof, Notes\}\rangle, \langle\{Exp\}, \{Notes\}\rangle\},$$
$$Invertex(\langle\{Rev, Exp\}, \{Prof, Notes\}\rangle) = \{Rev, Exp\},$$
$$Outvertex(\langle\{Rev, Exp\}, \{Prof, Notes\}\rangle) = \{Prof, Notes\},$$
$$Coinput(Rev, \langle\{Rev, Exp\}, \{Prof, Notes\}\rangle) = \{Exp\},$$
$$Cooutput(Prof, \langle\{Rev, Exp\}, \{Prof, Notes\}\rangle) = \{Notes\}.$$

The edges of S can be labeled, so that for example, $e_1 = \langle\{Rev, Exp\}, \{Prof, Notes\}\rangle$.

Note that a single metagraph edge is a singular metagraph. Also, note that an edge with a singular invertex and a singular outvertex is isomorphic with an edge in a directed graph.

Simple paths do not describe all of the connectivity properties of metagraphs. This is illustrated in the metagraph of Figure 2.2, in which there are two simple paths from x_1 to x_5, both of which have non-null coinputs. However, x_1 itself is sufficient to calculate x_5, if all three edges e_1, e_2, and e_3 are used. However, $\langle e_1, e_2, e_3\rangle$ does not represent a simple path, since there is no sequence of connected edges consisting of these edges. Rather, this metapath is the union of edges in two simple paths.

DEFINITION 2.5. Given a metagraph $S = \langle X, E\rangle$, a *metapath* $M(B, C)$ from a source $B \subset X$ to a target $C \subset X$ is a set of edges $E' \subseteq E$ such that (1) each $e' \in E'$ is on a simple path from some element in B to some element in C, (2) $[\bigcup_{e'} V_{e'} \backslash \bigcup_{e'} W_{e'}] \subseteq B$, and (3) $C \subseteq \bigcup_{e'} W_{e'}$.

There are three differences between simple paths and metapaths:

- First, a metapath is a set of edges and not a sequence of edges. For example, in Figure 2.2, one metapath from x_1 to x_5 is $M(\{x_1\}, \{x_5\}) = \{e_1, e_2, e_3\}$.

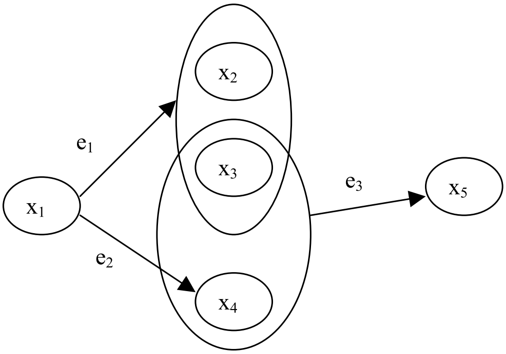

Figure 2.2. Metapath example.

- Second, the source and target of a metapath are sets, not elements, as in simple paths. Of course, these sets may sometimes be singleton sets, as is the case in Figure 2.2 (with $B = \{x_1\}$ and $C = \{x_5\}$).
- Third, the notion of a coinput does not apply to a metapath, since the source set includes all pure inputs.

2. THE INCIDENCE AND ADJACENCY MATRICES

In order to define an algebra for metagraph manipulation, two matrix representations of a metagraph are needed. These are the adjacency matrix and incidence matrix, respectively. It is worth noting that as with traditional graph structures, each of these matrices is a complete representation of a metagraph, and can be derived from the other.

DEFINITION 2.6. The *adjacency matrix* A for a metagraph $S = \langle X, E\rangle$ is an $I \times I$ matrix (where $I = |X|$), such that for all $i, j \in \{1, \ldots, I\}$,

$$a_{ij} = \bigcup_k (\alpha_{ij})_k,$$

where

$$(\alpha_{ij})_k = \begin{cases} \left\langle V_k \backslash \{x_i\}, W_k \backslash \{x\}, \langle E_k \rangle \right\rangle & \text{if } x_i \in V_k \wedge x_j \in W_k, \\ \phi \quad \text{otherwise.} \end{cases}$$

In other words, the adjacency matrix A of a metagraph is a square matrix with one row and one column for each element in the generating set X. The *ij*th element of A, denoted a_{ij}, is a set of triples, one for each edge e connecting x_i to x_j. Each triple is of the form $\langle CI_e, CO_e, e\rangle$, in which CI_e is the coinput of x_i in e and CO_e is the cooutput of x_j in e.

For example, the adjacency matrix for the metagraph in Figure 2.3 below is shown in Figure 2.4.

There is an algebra defined for metagraph adjacency matrices. Given adjacency matrices A_1 and A_2, defined for two metagraphs that have the same generating set, these matrices can be added and multiplied with the result in each case being another matrix over the same generating set. Intuitively, $A_1 + A_2$ represents the adjacency matrix of the union of the two metagraphs, while $A_1 * A_2$ represents all paths of length two, where the first edge is from the first metagraph and the second edge is from the second metagraph.

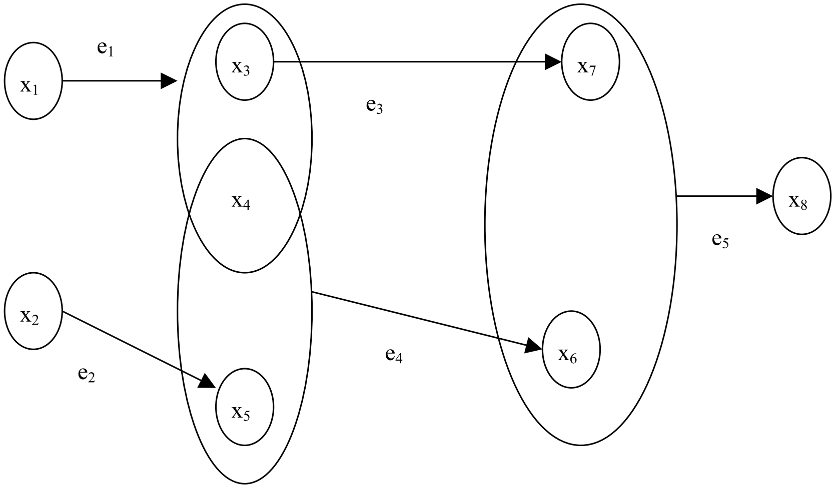

Figure 2.3. An example metagraph.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	$\varnothing$	$\varnothing$	$\langle \phi, \{x_4\}, e_1 \rangle$	$\langle \phi, \{x_3\}, e_1 \rangle$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$
x_2	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\langle \phi, \phi, e_2 \rangle$	$\varnothing$	$\varnothing$	$\varnothing$
x_3	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\langle \phi, \phi, e_3 \rangle$	$\varnothing$
x_4	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\langle \{x_5\}, \phi, e_4 \rangle$	$\varnothing$	$\varnothing$
x_5	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\langle \{x_4\}, \phi, e_4 \rangle$	$\varnothing$	$\varnothing$
x_6	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\langle \{x_7\}, \phi, e_5 \rangle$
x_7	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\langle \{x_6\}, \phi, e_5 \rangle$
x_8	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$

Figure 2.4. The adjacency matrix for the metagraph in Figure 2.3.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$
x_2	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$
x_3	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\langle\{x_4\},\phi,e_6\rangle$	$\varnothing$	$\varnothing$
x_4	$\langle\phi,\phi,e_7\rangle$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\langle\{x_3\},\phi,e_6\rangle$	$\varnothing$	$\varnothing$
x_5	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$
x_6	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$
x_7	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$
x_8	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$

Figure 2.5. Adjacency matrix of additional metagraph.

DEFINITION 2.7. Given a generating set X and two metagraphs $S_1 = \langle X, E_1\rangle$ and $S_2 = \langle X, E_2\rangle$ with adjacency matrices A_1 and A_2 respectively, then the *sum* of the two adjacency matrices is the adjacency matrix of the metagraph $S_3 = \langle X, E_1 \cup E_2\rangle$ with components

$$(A_1 + A_2)_{ij} = a^1_{ij} \cup a^2_{ij}.$$

Note that the two matrices must be defined on the same generating set. However, this is not a restrictive requirement. If the generating sets of the two metagraphs are overlapping but not identical, each metagraph can be defined over a new generating set which is the union of the two generating sets, and then the above definition can be applied.

As an example, consider the metagraph in Figure 2.3 combined with a metagraph consisting of two edges, $e_6 = \langle\{x_3, x_4\}, \{x_6\}\rangle$ and $e_7 = \langle\{x_4\}, \{x_1\}\rangle$, which has the adjacency matrix shown in Figure 2.5.

The result of adding the two adjacency matrices gives the adjacency matrix of the union of the two metagraphs, and this is shown in Figure 2.6.

The definition of multiplication of adjacency matrices is computationally more complex, since the result is not an adjacency matrix, but rather a matrix that identifies paths of length two between elements, as mentioned above. In order to define this operator, a number of preliminary concepts need to be specified.

DEFINITION 2.8. The *components* of an ordered triple R are $\alpha(R)$, $\beta(R)$ and $\gamma(R)$ respectively (i.e., $R = \langle\alpha(R), \beta(R), \gamma(R)\rangle$).

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	$\varnothing$	$\varnothing$	$\langle\phi,\{x_4\},e_1\rangle$	$\langle\phi,\{x_3\},e_1\rangle$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$
x_2	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\langle\phi,\phi,e_2\rangle$	$\varnothing$	$\varnothing$	$\varnothing$
x_3	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\langle\{x_4\},\phi,e_6\rangle$	$\langle\phi,\phi,e_3\rangle$	$\varnothing$
x_4	$\langle\phi,\phi,e_7\rangle$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\langle\{x_5\},\phi,e_4\rangle$, $\langle\{x_3\},\phi,e_6\rangle$	$\varnothing$	$\varnothing$
x_5	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\langle\{x_4\},\phi,e_4\rangle$	$\varnothing$	$\varnothing$
x_6	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\langle\{x_7\},\phi,e_5\rangle$
x_7	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\langle\{x_6\},\phi,e_5\rangle$
x_8	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$

Figure 2.6. The adjacency matrix for the combined metagraph.

DEFINITION 2.9. The operator $Cat(A, B)$ represents the concatenation of two ordered lists A and B.

For example, $Cat(\langle q, r\rangle, \langle q, s, t\rangle) = \langle q, r, q, s, t\rangle$.

DEFINITION 2.10. The $Trnc(.)$ operator truncates a list when it encounters a duplicate element.

For example, $Trnc(\langle a_n,\ n = 1, \ldots, N\rangle) = \langle a_n, n = 1, \ldots, M\rangle$, where $Q = \{a_n,\ n = 1, \ldots, M\}$ is a set of distinct elements and $a_{M+1} \in Q$.

DEFINITION 2.11. Let X be a generating set and let two metagraphs with adjacency matrices A and B respectively be defined on this generating set. Each cell in these matrices is a list of triples, with the nth triple in a_{ik} and the mth triple in b_{kj} denoted as $(a_{ik})_n$ and $(b_{kj})_m$ respectively. Then the '$\circ$' operator defines either an ordered triple or a null set, as follows:

(1) If $((a_{ik})_n \neq \phi) \wedge ((b_{kj})_m \neq \phi)$ then $(a_{ik})_n \circ (b_{kj})_m$ is a triple R specified as follows:
   (a) $\alpha(R) = (\alpha((a_{ik})_n) \cup \alpha((b_{kj})_m)) \backslash (\beta((a_{ik})_n) \cup \{x_i\})$,
   (b) $\beta(R) = (\beta((a_{ik})_n) \cup \beta((b_{kj})_m) \cup \{x_k\}) \backslash \{x_j\}$,

(c) $\gamma(R) = Trnc(Cat(\gamma(a_{ik})_n, \gamma(b_{kj})_m))$;

(2) Else $(a_{ik})_n \circ (b_{kj})_m = \phi$.

DEFINITION 2.12. Given a generating set X and two metagraphs $S_1 = \langle X, E_1\rangle$ and $S_2 = \langle X, E_2\rangle$ with adjacency matrices A and B respectively, let $(a_{ij})_n$ and $(b_{ij})_n$ be the ordered triples in a_{ij} and b_{ij} such that:

$$a_{ik} = \big\{(a_{ik})_n,\ n = 1, \ldots, N\big\} \quad \text{and} \quad b_{kj} = \big\{(b_{kj})_m,\ m = 1, \ldots, M\big\}.$$

Then the *product* of the two adjacency matrices A and B is denoted $A \times B$ with components

$$(A \times B)_{ij} = \bigcup_{k=1}^{K} \bigcup_{n=1}^{N} \bigcup_{m=1}^{M} \big((a_{ik})_n \circ (b_{kj})_m\big).$$

EXAMPLE 2.2. Given $a_{ik} = \langle \phi, \{x_4\}, e_1\rangle$ and $b_{kj} = \{\langle\{x_2\}, \{x_5\}, e_2\rangle, \langle\{x_4\}, \phi, e_3\rangle\}$, consider the first combination of $(a_{ik})_1 \circ (b_{kj})_1$. Since neither of them is null, we get a triple as follows:

$$\alpha\big((a_{ik})_1 \circ (b_{kj})_1\big) = \big(\phi \cup \{x_2\}\big)\backslash\big(\{x_4\} \cup \{x_1\}\big) = \{x_2\},$$
$$\beta\big((a_{ik})_1 \circ (b_{kj})_1\big) = \big(\{x_4\} \cup \{x_5\} \cup \{x_3\}\big)\backslash\{x_6\} = \{x_3, x_4, x_5\},$$
$$\gamma\big((a_{ik})_1 \circ (b_{kj})_1\big) = Trnc\big(Cat\big(\langle e_1\rangle, \langle e_2\rangle\big)\big) = \langle e_1, e_2\rangle.$$

Similarly, $(a_{ik})_1 \circ (b_{kj})_2 = \langle \phi, \{x_3, x_4\}, \langle e_1, e_3\rangle\rangle$.

Using multiplication, the powers of an adjacency matrix can also be computed. The nth power of A is denoted A^n. The *ij*th element of A^n, denoted a^n_{ij}, is a set of triples, one for each simple path $h(x_i, x_j)$ of length n connecting x_i to x_j. Each triple is of the form $\langle CI_h, CO_h, h\rangle$, in which h denotes the sequence of edges comprising the path, CI_h is the coinput of x_i in h and CO_h is the cooutput of x_j in h. The *closure* of A, denoted $A^* = A + A^2 + \cdots$, represents all simple paths of any length in the metagraph. The *ij*th element of A^*, denoted a^*_{ij}, is a set of triples, one for each simple path $h(x_i, x_j)$ of any length connecting x_i to x_j. Note that the multiplication operator allows any cycle to be traversed only once. Figure 2.7 shows the closure of the adjacency matrix in Figure 2.4.

The addition and multiplication operators on adjacency matrices of metagraphs also support the properties of associativity and distributivity, as shown below:

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	$\varnothing$	$\varnothing$	$\langle\phi,\{x_4\},e_1\rangle$	$\langle\phi,\{x_3\},e_1\rangle$	$\varnothing$	$\langle\{x_5\},\{x_3,x_4\},\langle e_1,e_3\rangle\rangle$	$\langle\phi,\{x_3,x_4\},\langle e_1,e_3\rangle\rangle$	$\{\langle x_6,\{x_3,x_4,x_7\},\langle e_1,e_3,e_5\rangle\rangle,$ $\langle\{x_5,x_7\},\{x_3,x_4,x_6\},\langle e_1,e_4,e_5\rangle\rangle\}$
x_2	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\langle\phi,\phi,e_2\rangle$	$\{\langle\{x_4\},\{x_5\},\langle e_2,e_4\rangle\rangle$	$\varnothing$	$\{\langle\{x_4,x_7\},\{x_5,x_6\},\langle e_2,e_4,e_5\rangle\rangle\}$
x_3	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\langle\phi,\phi,e_2\rangle$	$\{\langle\{x_6\},\{x_7\},\langle e_3,e_5\rangle\rangle\}$
x_4	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\langle\{x_5\},\phi,e_4\rangle$	$\varnothing$	$\{\langle\{x_5,x_7\},\{x_6\},\langle e_4,e_5\rangle\rangle\}$
x_5	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\langle\{x_4\},\phi,e_4\rangle$	$\varnothing$	$\{\langle\{x_4,x_7\},\{x_6\},\langle e_4,e_5\rangle\rangle\}$
x_6	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\langle\{x_7\},\phi,e_5\rangle$
x_7	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\langle\{x_6\},\phi,e_5\rangle$
x_8	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$

Figure 2.7. The closure of the adjacency matrix in Figure 2.4.

THEOREM 2.1. *Given a generating set X and three metagraphs defined on this set with adjacency matrices A, B, and C respectively, then*

(1) $A \times (B \times C) = (A \times B) \times C$,
(2) $A + (B \times C) = (A \times C) + (B \times C)$.

PROOF. Since the multiplication operation identifies all paths made up of edges in the first metagraph followed by an edge in the second metagraph, $A \times (B \times C)$ identifies all paths of length three consisting of an edge from A followed by an edge from B and then an edge from C respectively. This is the same as in $(A \times B) \times C$, which proves associativity.

To prove the distributive property, if $D = (A \times C) + (B \times C)$, it suffices to show that for any i, j, $d_{ij} = ((A + B) \times C)_{ij}$. In the following, the notation $(a_{ij})_n$ refers to the nth triple in a_{ij}, while a^n_{ij} refers to the entry in the ith row and jth column of A^n, and $(a^n_{ij})_m$ refers to the mth element of a^n_{ij}.

Let $|a_{ij}| = M_1$, $|b_{ij}| = M_2$, and $|c_{ij}| = N$. Also, let $Y = A + B$ (i.e., $\forall i, j, y_{ij} = a_{ij} \cup b_{ij}$. Reorganize b_{ij} so that for $q \leq Q$, $(b_{ij})_q \notin a_{ij}$ and for all $q > Q$, $(b_{ij})_q \in a_{ij}$. Thus, $|y_{ij}| = M_1 + Q$. Then x_{ij} can be partitioned into the following sets:

$$(y_{ij})_p = (a_{ij})_{m_1} \quad \text{for } p = 1, \ldots, M_1,$$

$$(y_{ij})_p = (b_{ij})_{p-M_1} \quad \text{for } p = (M_1 + 1), \ldots, (M_1 + Q).$$

Then,

$$
\begin{aligned}
d_{ij} &= \left(\bigcup_{k,m_1,n} (a_{ik})_{m_1} \circ (c_{kj})_n \right) \cup \left(\bigcup_{k,m_2,n} (b_{ik})_{m_2} \circ (c_{kj})_n \right) \\
&= \bigcup_{k,n} \left(\left(\bigcup_{m_1} (a_{ik})_{m_1} \circ (c_{kj})_n \right) \cup \left(\bigcup_{m_2} (b_{ik})_{m_2} \circ (c_{kj})_n \right) \right) \\
&= \bigcup_{k,n} \left(\left(\bigcup_{p=1}^{M_1} (y_{ik})_p \circ (c_{kj})_n \right) \cup \left(\bigcup_{p=M_1+1}^{M_1+Q} (y_{ik})_p \circ (c_{kj})_n \right) \right) \\
&= \bigcup_{k,n} \left(\bigcup_{p=1}^{M_1+Q} (y_{ik})_p \circ (c_{kj})_n \right) \\
&= \bigcup_{k,p,n} \big((a_{ik} \cup b_{ik})_p \circ (c_{kj})_n \big).
\end{aligned}
$$

Thus, $D = (A + B) \times C$, which is the desired result. □

Also, note that the null matrix D (with $d_{ij} = \phi\ \forall i, j$) is a left and right identity under addition (i.e., $A + D = D + A = A$). This implies that the set of all adjacency matrices defined on the same generating set forms a commutative idempotent monoid under addition, while the set of all non-null adjacency matrices forms a semi-group under multiplication.

DEFINITION 2.13. The *incidence matrix* G of a metagraph has one row for each element in the generating set and one column for each edge. The *ij*th component of G, g_{ij}, is -1 if x_i is in the invertex of e_j, it is $+1$ if x_i is in the outvertex of e_j, and it is Ø otherwise.

The incidence matrix for the metagraph in Figure 2.3 is shown in Figure 2.8 below.

Once the closure A^* of a metagraph's adjacency matrix has been constructed, it can be used to identify a variety of connectivity features of that metagraph, as discussed in the next chapter.

3. IDENTIFYING METAPATHS

The adjacency matrix and its closure can be used to find paths and metapaths. One of the benefits of the metagraph representation (versus simpler graph representations) is that searches for metapaths can be limited to only

	e_1	e_2	e_3	e_4	e_5
x_1	-1				
x_2		-1			
x_3	+1		-1		
x_4	+1			-1	
x_5		+1		-1	
x_6				+1	-1
x_7			+1		-1
x_8					+1

Figure 2.8. The incidence matrix for the metagraph in Figure 2.3.

those portions of the A^* matrix that deals with the elements in the source and target sets, which can substantially reduce the search space. This is because every metapath from B to C must consist of edges based on a combination of triples from cells a^*_{ij} such that $x_i \in B$ and $x_j \in C$. Furthermore, the efficiency of the search procedure now becomes a function of the number of simple paths between B and C (each corresponding to a triple in the candidate set), rather than the entire closure matrix.

Another useful observation that can be exploited is that if there is a metapath from B to C, then there should be triples composed of these edges in A^* in every column j such that $x_j \in C$. Also, in using the closure matrix to find metapaths $M(B, C)$, even though there is at least one triple in every column of A^* corresponding to elements of C, it is not always necessary to examine each triple explicitly, because the triples include the co-inputs and co-outputs for the path that they represent. Furthermore, if we use a conservative approach that always considers a minimal number of rows, then the metapaths obtained are all input-dominant.

DEFINITION 2.14. Given a metagraph $S = \langle X, E \rangle$, for any two sets of elements B and C in S, a metapath $M(B, C)$ is said to be *input-dominant* if there is no metapath $M'(B', C)$ such that $B' \subset B$.

Based on the above observations, the procedure to find metapaths can be described as follows:

1. Select a candidate set of input rows I in A^* such that $x_i \in B, \forall i \in I, B = \bigcup_i x_i$. Start with single rows, and repeat with larger sets progressively in successive iterations.
2. If $\exists x_j \in B$ such that $a^*_{ij} = \phi, \forall i \in I$, then there is no metapath from $\{x_i \mid i \in I\}$ to C. Return to step 1 and repeat with another set of rows.
3. Find a candidate set of triples in cells a^*_{ij} such that $i \in I, x_j \in C$ that forms a cover for C (where a cover for C is a set of triples T such that $C \subseteq \bigcup_{t \in T} output(t)$). If such a cover is found, then $\bigcup_{t \in T} path(t)$ comprises an input dominant metapath from $B(= \{x_i \mid i \in I\})$ to C.
4. Otherwise, return to step 1 and use an alternative candidate set I.

The stopping criterion for the procedure (in step 3 after a metapath is found, or in step 1 if there are no more candidate sets) depends upon whether the desired outcome is one metapath or every metapath.

http://www.springer.com/978-0-387-37233-4

Metagraphs and Their Applications
Basu, A.; Blanning, R.W.
2007, VIII, 172 p. 86 illus., Hardcover
ISBN: 978-0-387-37233-4