Constructing Adjacency Arrays from Incidence Arrays

Hayden Jananthan[1,2] Karia Dibert[2,3] Jeremy Kepner[2,3,4]

[1]Vanderbilt University Mathematics Department, [2]MIT Lincoln Laboratory Supercomputing Center,
[3]MIT Mathematics Department, [4]MIT Computer Science & AI Laboratory

May 28, 2017

Outline

Adjacency Matrices are a convenient way of representing directed graphs.

In practice Incidence Matrices are readily obtained from raw data.

The standard approach to getting an adjacency matrix from incidence matrices $\mathbf{E}_{\mathrm{out}}, \mathbf{E}_{\mathrm{in}}$ is via

$$\mathbf{A} = \mathbf{E}_{\mathrm{out}}^{\mathsf{T}} \mathbf{E}_{\mathrm{in}}$$

Real-world data need not be in the form of a matrix.

- The column and row indices need not be positive integers.

  $\implies$ They can be arbitrary key sets K_1 and K_2.

- The values need not be 0 and 1 nor do $+$ and $\times$ need be defined on those values.

  $\implies$ They can be taken from a set of values V with binary operations $\oplus$ and $\otimes$.

e.g. K_1 contains song identifiers, K_2 contains song properties (length, album, artist, etc), V contains strings with ordered lexicographically, $\oplus = \max$, $\otimes = \min$, $0 = \mathrm{a}$, $1 = \infty$.

Notation

- G is a directed multigraph with
  - vertex set $K_{\text{out}} \cup K_{\text{in}}$ where
  - K_{out} consists of vertices which are sources of edges and
  - K_{in} consists of vertices which are targets of edges, and
  - edge set K.

  Assume every vertex is either a source or a target, and that $K_{\text{out}} \cup K_{\text{in}}$ and K are finite and totally-ordered.

- V is some set with two binary operations $\oplus$ and $\otimes$ with (unequal) identity elements 0 and 1, resp.

Definitions

An associative array is a map $\mathbf{A} : K_1 \times K_2 \to V$.

Given associative arrays

$$\mathbf{A} : K_1 \times K_3 \to V$$
$$\mathbf{B} : K_3 \times K_2 \to V,$$

their array product

$$\mathbf{C} = \mathbf{A} \oplus . \otimes \mathbf{B} = \mathbf{AB}$$

is an associative array $K_1 \times K_2 \to V$ defined by

$$\mathbf{C}(k_1, k_2) = \bigoplus_{k_3 \in K_3} \mathbf{A}(k_1, k_3) \otimes \mathbf{B}(k_3, k_2)$$

Definitions

Given G directed multigraph,

- $\mathbf{A} : K_{\text{out}} \times K_{\text{in}} \to V$ is an adjacency array if $\mathbf{A}(x, y) \neq 0$ if and only if there is an edge from x to y.
- $\mathbf{E}_{\text{out}} : K \times K_{\text{out}} \to V$ is a source incidence array if $\mathbf{E}_{\text{out}}(k, x) \neq 0$ if and only if k is an edge from x.
- $\mathbf{E}_{\text{in}} : K \times K_{\text{in}} \to V$ is a target incidence array if $\mathbf{E}_{\text{in}}(k, y) \neq 0$ if and only if k is an edge into y.

(Note the use of "a"; incidence or adjacency arrays are not necessarily unique.)

Results

A natural question is whether $\mathbf{E}_{\text{out}}^{\mathsf{T}}\mathbf{E}_{\text{in}}$ is an adjacency array for general $\oplus, \otimes$.

Theorem

Suppose V a set with binary operations $\oplus, \otimes$ with identities $0, 1$, $0 \neq 1$. The following are equivalent:

1. *$\oplus$ and $\otimes$ satisfy the properties:*
    - *Zero-Sum-Free: $a \oplus b = 0$ if and only if $a = b = 0$.*
    - *No Zero Divisors: $a \otimes b = 0$ implies at least one of a, b is 0.*
    - *0 Annihilates: $a \otimes 0 = 0 \otimes a = 0$ for all a.*
2. *If G a directed multigraph with incidence arrays $\mathbf{E}_{\text{out}}, \mathbf{E}_{\text{in}}$ then $\mathbf{E}_{\text{out}}^{\mathsf{T}}\mathbf{E}_{\text{in}}$ is an adjacencey array.*

Outline

Proof Sketch: ❶ $\Longrightarrow$ ❷

Assume $\oplus, \otimes$ satisfy hypotheses of ❶ .

$\mathbf{A} = \mathbf{E}_{\mathrm{out}}^{\mathsf{T}} \mathbf{E}_{\mathrm{in}}$ being an adjacency array is equivalent to

$$\bigoplus_{k \in K} \mathbf{E}_{\mathrm{out}}(k,x) \otimes \mathbf{E}_{\mathrm{in}}(k,y) = 0 \quad \Longleftrightarrow \quad \begin{array}{c} \text{for all } k \in K, \\ \mathbf{E}_{\mathrm{out}}(k,x) = 0 \text{ or } \mathbf{E}_{\mathrm{in}}(k,y) = 0 \end{array}$$

- Forward implication follows from "Zero-Sum-Free" followed by "No Zero Divisors".
- Backward implication follows from "0 Annihilates".

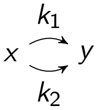
Proof Sketch: ❷ $\Longrightarrow$ ❶ (Zero-Sum-Free)

Assume that ❷ holds.

For sake of a contradiction, suppose v, w are non-zero and

$$v \oplus w = 0$$

Construct graph and incidence arrays such that $\mathbf{E}_{\text{out}}^{\mathsf{T}} \mathbf{E}_{\text{in}}$ is not an adjacency array.



$$\mathbf{E}_{\text{out}} = \begin{array}{c} \\ k_1 \\ k_2 \end{array} \begin{array}{c} x \\ \left[\begin{array}{c} v \\ w \end{array}\right] \end{array} \qquad \mathbf{E}_{\text{in}} = \begin{array}{c} \\ k_1 \\ k_2 \end{array} \begin{array}{c} y \\ \left[\begin{array}{c} 1 \\ 1 \end{array}\right] \end{array}$$

Then

$$\mathbf{E}_{\text{out}}^{\mathsf{T}} \mathbf{E}_{\text{in}}(x, y) = (v \otimes 1) \oplus (w \otimes 1) = v \oplus w = 0$$

a contradiction.

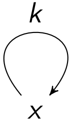
Proof Sketch: 2 $\Longrightarrow$ 1 (No Zero Divisors)

Assume that 2 holds.

For sake of a contradiction, suppose v, w are non-zero and

$$v \otimes w = 0$$

Construct graph and incidence arrays such that $\mathbf{E}_{\mathrm{out}}^{\mathsf{T}}\mathbf{E}_{\mathrm{in}}$ is not an adjacency array.



$$\mathbf{E}_{\mathrm{out}} = \begin{array}{cc} & x \\ k & \left[\begin{array}{c} v \end{array}\right] \end{array} \qquad \mathbf{E}_{\mathrm{in}} = \begin{array}{cc} & x \\ k & \left[\begin{array}{c} w \end{array}\right] \end{array}$$

Then

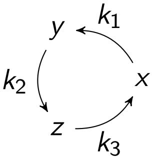
$$\mathbf{E}_{\mathrm{out}}^{\mathsf{T}}\mathbf{E}_{\mathrm{in}}(x, x) = v \otimes w = 0$$

a contradiction.

Proof Sketch: ❷ $\Longrightarrow$ ❶ (0 Annihilates)

Assume that ❷ holds.

Suppose $v \neq 0$. Construct graph and incidence arrays



$$\mathbf{E}_{\mathrm{out}} = \begin{array}{c} \\ k_1 \\ k_2 \\ k_3 \end{array} \begin{array}{c} \begin{array}{ccc} x & y & z \end{array} \\ \begin{bmatrix} v & 0 & 0 \\ 0 & v & 0 \\ 0 & 0 & v \end{bmatrix} \end{array} \qquad \mathbf{E}_{\mathrm{in}} = \begin{array}{c} \\ k_1 \\ k_2 \\ k_3 \end{array} \begin{array}{c} \begin{array}{ccc} x & y & z \end{array} \\ \begin{bmatrix} 0 & v & 0 \\ 0 & 0 & v \\ v & 0 & 0 \end{bmatrix} \end{array}$$

Then no edge from x to x means

$$\mathbf{E}_{\mathrm{out}}^{\mathsf{T}} \mathbf{E}_{\mathrm{in}}(x, x) = (v \otimes 0) \oplus (0 \otimes 0) \oplus (0 \otimes v) = 0$$

Zero-Sum-Free implies $v \otimes 0 = 0 \otimes 0 = 0 \otimes v = 0$.

Further Results

Corollary

Suppose V a set with binary operations $\oplus, \otimes$ with identities $0, 1$, $0 \neq 1$. The following are equivalent:

1. *$\oplus$ and $\otimes$ satisfy the properties:*
   - *Zero-Sum-Free: $a \oplus b = 0$ if and only if $a = b = 0$.*
   - *No Zero Divisors: $a \otimes b = 0$ implies at least one of a, b is 0.*
   - *0 Annihilates: $a \otimes 0 = 0 \otimes a = 0$ for all a.*
2. *If G a directed graph with incidence arrays $\mathbf{E}_{\text{out}}, \mathbf{E}_{\text{in}}$ then $\mathbf{E}_{\text{in}}^{\mathsf{T}}\mathbf{E}_{\text{out}}$ is an adjacencey array of the reverse of G.*

Proof Sketch.

Taking the reverse switches the roles of $\mathbf{E}_{\text{out}}$ and $\mathbf{E}_{\text{in}}$. □

Outline

Consider the following associative array **E** of data from a music database with selected subarrays $\mathbf{E}_1$ and $\mathbf{E}_2$.

E	Artist|Bandayde	Artist|Kastle	Artist|Kitten	Date|2010-06-30	Date|2012-08-28	Date|2012-09-16	Date|2013-05-30	Date|2013-09-30	Date|2013-10-03	Genre|Electronic	Genre|Pop	Genre|Rock	Label|Atlantic	Label|Elektra Records	Label|Free	Label|The Control Group	Release|Cut It Out	Release|Cut It Out Remixes	Release|Cut It Out/Sugar	Release|Japanese Eyes	Release|Kill The Light	Release|Like A Stranger	Release|Yesterday	Type|EP	Type|LP	Type|Single	Writer|Barrett Rich	Writer|Chad Anderson	Writer|Chloe Chaidez	Writer|Julian Chaidez	Writer|Nicholas Johns
031013ktnA1			1						1			1			1								1			1		1	1	1	
053013ktnA1	1						1			1					1			1								1				1	1
053013ktnA2		1					1			1					1			1								1	1				
063012ktnA1			1	1								1	1						1					1				1	1		
063012ktnA2			1	1								1	1						1					1				1	1		
063012ktnA3			1	1								1				1					1			1				1	1		
063012ktnA4			1	1								1				1					1			1				1	1		
063012ktnA5			1	1								1				1					1			1				1	1		
082812ktnA1			1		1						1		1				1							1				1	1	1	
082812ktnA2			1		1						1		1				1							1				1	1		
082812ktnA3			1		1						1		1				1							1				1	1		
082812ktnA4			1		1						1		1				1							1				1	1		
082812ktnA5			1		1						1		1				1							1				1	1	1	
082812ktnA6			1		1						1		1				1							1				1	1		
093012ktnA1			1					1		1	1			1								1			1			1	1		
093012ktnA2			1					1		1	1			1								1			1			1	1		
093012ktnA3			1					1		1	1			1								1			1			1	1	1	
093012ktnA4			1					1		1	1			1								1			1			1	1		
093012ktnA5			1					1		1	1			1								1			1			1	1		
093012ktnA6			1					1		1	1			1								1			1			1	1		
093012ktnA7			1					1		1	1			1								1			1			1	1		
093012ktnA8			1			1														1						1		1	1		

$\mathbf{E}_1 = \mathbf{E}(:,\text{'Genre|A : Genre|Z '})$

$\mathbf{E}_2 = \mathbf{E}(:,\text{'Writer|A : Write|Z '})$

$\mathbf{E}_1$ and $\mathbf{E}_2$ can be considered as source and target incidence arrays, respectively.

$\mathbf{E}_1$

	Genre|Electronic	Genre|Pop	Genre|Rock
031013ktnA1			1
053013ktnA1	1		
053013ktnA2	1		
063012ktnA1			1
063012ktnA2			1
063012ktnA3			1
063012ktnA4			1
063012ktnA5			1
082812ktnA1		1	
082812ktnA2		1	
082812ktnA3		1	
082812ktnA4		1	
082812ktnA5		1	
082812ktnA6		1	
093012ktnA1	1	1	
093012ktnA2	1	1	
093012ktnA3	1	1	
093012ktnA4	1	1	
093012ktnA5	1	1	
093012ktnA6	1	1	
093012ktnA7	1	1	

$\mathbf{E}_2$

	Writer|Barrett Rich	Writer|Chad Anderson	Writer|Chloe Chaidez	Writer|Julian Chaidez	Writer|Nicholas Johns
031013ktnA1		1	1	1	
053013ktnA1				1	1
053013ktnA2	1				
063012ktnA1		1	1		
063012ktnA2		1	1		
063012ktnA3		1	1		
063012ktnA4		1	1		
063012ktnA5		1	1		
082812ktnA1		1	1	1	
082812ktnA2		1	1		
082812ktnA3		1	1		
082812ktnA4		1	1		
082812ktnA5		1	1	1	
082812ktnA6		1	1		
093012ktnA1		1	1		
093012ktnA2		1	1		
093012ktnA3		1	1	1	
093012ktnA4		1	1		
093012ktnA5		1	1		
093012ktnA6		1	1		
093012ktnA7		1	1		
093012ktnA8		1	1		

Values of $\mathbf{E}_1, \mathbf{E}_2$ are non-negative reals (with possibly ∞).
$\mathbf{E}_1^\mathsf{T}\mathbf{E}_2$ calculated under several pairs of operations $\oplus, \otimes$:

$\mathbf{E}_1^\mathsf{T}$ +.× $\mathbf{E}_2$ =

	Writer|Barrett Rich	Writer|Chad Anderson	Writer|Chloe Chaidez	Writer|Julian Chaidez	Writer|Nicholas Johns
Genre|Electronic	1	7	7	2	1
Genre|Pop		13	13	3	
Genre|Rock		6	6	1	

$\mathbf{E}_1^\mathsf{T}$ max.+ / min.+ $\mathbf{E}_2$ =

	Writer|Barrett Rich	Writer|Chad Anderson	Writer|Chloe Chaidez	Writer|Julian Chaidez	Writer|Nicholas Johns
Genre|Electronic	2	2	2	2	2
Genre|Pop		2	2	2	
Genre|Rock		2	2	2	

$\mathbf{E}_1^\mathsf{T}$ max.min $\mathbf{E}_2$ =

	Writer|Barrett Rich	Writer|Chad Anderson	Writer|Chloe Chaidez	Writer|Julian Chaidez	Writer|Nicholas Johns
Genre|Electronic	1	1	1	1	1
Genre|Pop		1	1	1	
Genre|Rock		1	1	1	

$\mathbf{E}_1^\mathsf{T}$ min.max $\mathbf{E}_2$ =

	Writer|Barrett Rich	Writer|Chad Anderson	Writer|Chloe Chaidez	Writer|Julian Chaidez	Writer|Nicholas Johns
Genre|Electronic	1	1	1	1	1
Genre|Pop		1	1	1	
Genre|Rock		1	1	1	

$\mathbf{E}_1^\mathsf{T}$ max.× / min.× $\mathbf{E}_2$ =

	Writer|Barrett Rich	Writer|Chad Anderson	Writer|Chloe Chaidez	Writer|Julian Chaidez	Writer|Nicholas Johns
Genre|Electronic	1	1	1	1	1
Genre|Pop		1	1	1	
Genre|Rock		1	1	1	

Outline

What algebraic properties are necessary and sufficient when restricting to smaller classes of directed graphs?

- For example, directed graphs without multiple edges:

Conjecture

Suppose $\oplus, \otimes$ are binary operations on V with additive identity 0. Then the following are equivalent:

1. *V has no zero divisors and 0 is an annihilator.*
2. *If G is a directed graph (without multiple edges) and $\mathbf{E}_{\mathrm{out}}, \mathbf{E}_{\mathrm{in}}$ are incidence arrays, then $\mathbf{E}_{\mathrm{out}}^{\mathsf{T}} \mathbf{E}_{\mathrm{in}}$ is an adjacency array.*

- What about undirected graphs?

Is the existence of 1 necessary?

What properties does 0 necessarily have without assuming it is the additive identity?

Any way to reflect other structures of the graph (weights, counting of multiple edges, etc) and still be compatible with forming the product $\mathbf{E}_{\text{out}}^{\mathsf{T}}\mathbf{E}_{\text{in}}$?