Symbolic Networks and Algebra

Vladimir Batagelj

IMFM Ljubljana, IAM UP Koper, NRU HSE Moscow

SDA 2022 – 8th Symbolic Data Analysis workshop

Caserta, Italy, 7-8. September 2022

Outline

Vladimir Batagelj: vladimir.batagelj@fmf.uni-lj.si

Last version of slides (September 8, 2022, 05 : 01): SDAnets.pdf

SDA and algebra

From Edwin's slides

Examples of symbols:

s = [min, max] , s = interquartile interval

s = cumulative distribution,

s = a barchart , a histogram etc.

THEY ARE CALLED SYMBOLS BECAUSE THEY CANNOT BE MANIPULATED AS NUMBERS

The Analysis of such data is called SYMBOLIC DATA ANALYSIS .

In this presentation, I would like to show that by imposing on a selected type of symbol an appropriate algebraic structure we can manipulate them in a similar way as numbers.

In recent years I used temporal quantities (a kind of symbolic object) in the analysis of temporal bibliographic [8] and bike-sharing [5] networks. I will put my experiences into a broader framework.

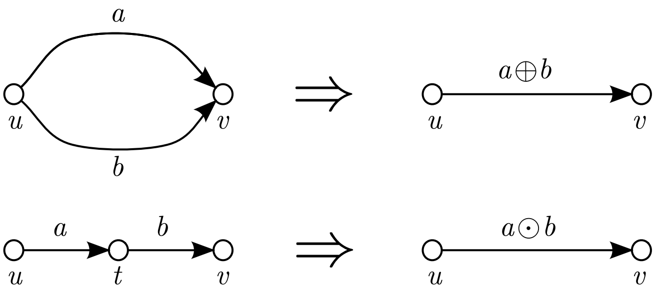
Symbolic networks

A *network* $\mathbf{N} = (V, L, P, W)$ consists of a *graph* $\mathbf{G} = (V, L)$, where

- V is the set of *nodes*
- $L = E \cup A$, $E \cap A = \emptyset$ is the set of *links* (A is the set of *arcs* – directed links, E is the set of *edges* – undirected links);
- P are node value functions / *properties*: $p : V \to R_p$; and
- W are link value functions / *weights*: $w : L \to R_w$.

A network $\mathbf{N}$ is *symbolic* if at least one of its value functions is symbolic – has as its range R a set of symbolic values.

Computing with link weights in networks



A *semiring* is a "natural" algebraic structure to formalize computations with link weights in networks. They allow us to extend the weights from links to nodes, walks (paths), and sets of walks.

Semiring

Let $\mathbb{A}$ be a set and a, b, c elements from $\mathbb{A}$. A *semiring* [11, 4, 16, 3, 1] is an algebraic structure $(\mathbb{A}, \oplus, \odot, 0, 1)$ with two binary operations (addition $\oplus$ and multiplication $\odot$) where:

$(\mathbb{A}, \oplus, 0)$ is an *abelian monoid* with the neutral element 0 (zero):
$a \oplus b = b \oplus a$ – commutativity
$(a \oplus b) \oplus c = a \oplus (b \oplus c)$ – associativity
$a \oplus 0 = a$ – existence of zero

$(\mathbb{A}, \odot, 1)$ is a *monoid* with the neutral element 1 (unit):
$(a \odot b) \odot c = a \odot (b \odot c)$ – associativity
$a \odot 1 = 1 \odot a = a$ – existence of a unit

Multiplication $\odot$ *distributes* over addition $\oplus$:
$a \odot (b \oplus c) = a \odot b \oplus a \odot c \qquad (b \oplus c) \odot a = b \odot a \oplus c \odot a$

In formulas we assume precedence of the multiplication over the addition.

Semiring

A semiring $(\mathbb{A}, \oplus, \odot, 0, 1)$ is *complete* iff the addition is well defined for countable sets of elements and the commutativity, associativity, and distributivity hold in the case of countable sets.

These properties are generalized in this case; for example, the distributivity takes the form

$$(\bigoplus_i a_i) \odot (\bigoplus_j b_j) = \bigoplus_i (\bigoplus_j (a_i \odot b_j)) = \bigoplus_{i,j} (a_i \odot b_j)$$

The addition is *idempotent* iff $a \oplus a = a$ for all $a \in \mathbb{A}$. In this case the semiring over a finite set $\mathbb{A}$ is complete.

Semiring

A semiring $(\mathbb{A}, \oplus, \odot, 0, 1)$ is *closed* iff for the additional (unary) *closure* operation $*$ it holds for all $a \in \mathbb{A}$:

$$a^* = 1 \oplus a \odot a^* = 1 \oplus a^* \odot a.$$

Different closures over the same semiring can exist. A complete semiring is always closed for the closure $a^* = \bigoplus_{i \in N} a^i$.

In a closed semiring we can also define a *strict closure* $\overline{a}$ as $\overline{a} = a \odot a^*$.

In a semiring $(\mathbb{A}, \oplus, \odot, 0, 1)$ the *absorption* law holds iff for all $a, b, c \in \mathbb{A}$:

$$a \odot b \oplus a \odot c \odot b = a \odot b.$$

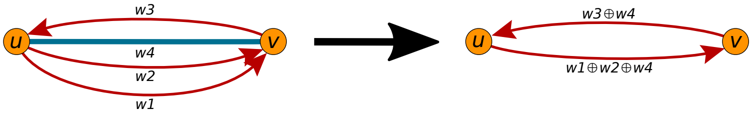
For the absorption law, because of the distributivity, it is sufficient to check the property $1 \oplus c = 1$ for all $c \in \mathbb{A}$.

Some semirings used in network data analysis

1. Combinatorial: $(\mathbb{N}, +, \cdot, 0, 1)$ or $(\mathbb{R}_0^+, +, \cdot, 0, 1)$
2. Reachability: $(\{0, 1\}, \vee, \wedge, 0, 1)$
3. Shortest paths [3]: $(\mathbb{R}_0^+, \min, +, \infty, 0)$
4. MaxMin (capacity): $(\mathbb{R}_0^+, \max, \min, 0, \infty)$
5. Pathfinder [21, 22, 23]: $(\overline{\mathbb{R}}_0^+, \min, \boxed{r}, \infty, 0)$ where $a \boxed{r} b = \sqrt[r]{a^r + b^r}$ (Minkowski)
6. Regular languages [24] – regular algebra (union, concatenation, and iteration of sets of strings over an alphabet A).

Simplification of weighted networks

For a subset $S \subseteq L$ of links, we define its *total* weight as $\Sigma(S) = \bigoplus_{\ell \in S} w(\ell)$.



A *simplification* of a weighted network $\mathbf{N} = (V, L, w)$ over an abelian monoid $(\mathbb{A}, \oplus, 0)$ is its transformation into a *simple* directed weighted network $\widehat{\mathbf{N}} = (V, A, \widehat{w})$ such that for

$$L(u, v) = \{\ell \in L : \text{init}(\ell) = u \wedge \text{term}(\ell) = v\}$$

it holds $L(u, v) \neq \emptyset \Rightarrow (u, v) \in A \wedge \widehat{w}(u, v) = \Sigma(L(u, v))$.

For undirected weighted networks, a special undirected simplification can be defined similarly.

Activity of nodes

$$\text{inStar}(v) = \{\ell \in L : \text{term}(\ell) = v\}, \quad \text{outStar}(v) = \{\ell \in L : \text{init}(\ell) = v\}$$

Using the link weight w of the network $\mathbf{N}$ we can compute some interesting node properties such as *in-sum* (weighted indegree):

$$inS(\mathbf{N}, v) = \Sigma(\text{inStar}(v))$$

and *out-sum* (weighted outdegree):

$$outS(\mathbf{N}, v) = \Sigma(\text{outStar}(v))$$

of a node v. They are describing the *activity* of the node v.

Activity of sets

For a simple weighted network $\mathbf{N} = (V, A, w)$ and a subset of nodes $C \subseteq V$ we can define the *boundaries* of C:

$$\mathrm{inBd}(C) = \{\ell \in A : \mathrm{init}(\ell) \notin C \wedge \mathrm{term}(\ell) \in C\}$$
$$\mathrm{outBd}(C) = \{\ell \in A : \mathrm{init}(\ell) \in C \wedge \mathrm{term}(\ell) \notin C\}$$
$$\mathrm{Bd}(C) = \mathrm{inBd}(C) \cup \mathrm{outBd}(C)$$

the *interior* of C:

$$\mathrm{Int}(C) = \{\ell \in A : \mathrm{init}(\ell) \in C \wedge \mathrm{term}(\ell) \in C\}$$

and the corresponding *activities*:

$$\mathrm{inPut}(C) = \Sigma(\mathrm{inBd}(C)) \quad \text{and} \quad \mathrm{outPut}(C) = \Sigma(\mathrm{outBd}(C))$$
$$\mathrm{Act}(C) = \Sigma(\mathrm{Int}(C))$$

Extension to walks and sets of walks

We can extend the weight w to walks and sets of walks on **N** by the following rules:

- Let $\sigma_v = (v)$ be a null walk in the node $v \in V$; then $w(\sigma_v) = 1$.
- Let $\sigma = (u_0, u_1, u_2, \ldots, u_{p-1}, u_p)$ be a walk of length $p \geq 1$ on **N** then

$$w(\sigma) = \bigodot_{i=1}^{p} w(u_{i-1}, u_i).$$

- For the empty set of walks $\emptyset$ it holds $w(\emptyset) = 0$.
- Let $S = \{\sigma_1, \sigma_2, \ldots\}$ be a set of walks in **N** then

$$w(S) = \bigoplus_{\sigma \in S} w(\sigma).$$

Two-mode network matrix

Let $(\mathbb{A}, \oplus, \odot, 0, 1)$ be a semiring then also $(\mathbb{A} \cup \{\square\}, \oplus, \odot, \square, 1)$, where for every $a \in \mathbb{A} \cup \{\square\}$ hold $\square \oplus a = a$ and $\square \odot a = a \odot \square = \square$, is a semiring.

The network *matrix* of a two-mode network $\mathbf{N} = ((U, V), L, w)$, $w : L \to \mathbb{A}$ is defined as

$$\mathbf{W}[u, v] = \begin{cases} w(u, v) & (u, v) \in L \\ \square & \text{otherwise} \end{cases}$$

We distinguish between 0 (weight) and $\square$ (no link).

In programming $\square$ can be represented by `NA` or `none`.

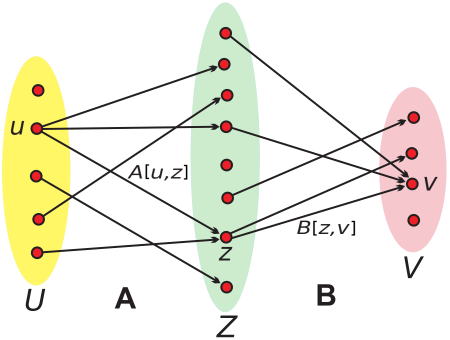
Multiplication of networks

The *product* $\mathbf{C} = \mathbf{A} \odot \mathbf{B}$ of two compatible matrices $\mathbf{A}_{U\times Z}$ and $\mathbf{B}_{Z\times V}$ over a semiring $(\mathbb{A} \cup \{\square\}, \oplus, \odot, \square, 1)$ is defined in the standard way

$$C[u,v] = \bigoplus_{z\in Z} A[u,z] \odot B[z,v]$$

The product of two compatible networks $\mathbf{N}_A = ((U,Z), L_A, a)$ and $\mathbf{N}_B = ((Z,V), L_B, b)$ is the network $\mathbf{N}_C = ((U,V), L_C, c)$ where $L_C = \{(u,v) : c[u,v] \neq \square\}$ and the weight c is determined by the matrix $\mathbf{C}$, $c(u,v) = C[u,v]$.

Multiplication of networks



For a combinatorial semiring, in binary networks $\mathbf{N}_A$ and $\mathbf{N}_B$, the value of $C[u,v]$ of $\mathbf{C} = \mathbf{A} \cdot \mathbf{B}$ counts the number of ways we can go from the node $u \in U$ to the node $v \in V$ passing through Z, $C[u,v] = |N_A(u) \cap N_B(v)|$.

$$C[u,v] = \bigoplus_{z \in N_A(u) \cap N_B(v)} A[u,z] \odot B[z,v]$$

Semirings

In our article [12] we made an overview of semirings used in network data analysis, and results on network matrices and vectors over a semiring (addition, multiplication, power, closure) and sets of walks in networks.

Some books about semirings: Gondran and Minoux [16], Glazek [15], Ostoic [19].

Geodetic semiring

7 The *geodetic semiring* $(\overline{\mathbb{N}}^2, \oplus, \odot, (\infty, 0), (0, 1))$ [4], where $\overline{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$ and we define *addition* $\oplus$ with:

$$(a, i) \oplus (b, j) = (\min(a, b), \begin{cases} i & a < b \\ i + j & a = b \\ j & a > b \end{cases})$$

and *multiplication* $\odot$ with:

$$(a, i) \odot (b, j) = (a + b, i \cdot j).$$

(a combination of the combinatorial and the shortest paths semirings; counting the number of shortest paths). It was used in an algorithm for computing the betweenness in a network. In 2001 Brandes proposed a faster algorithm [9].

Sets and multisets

8. Sets. Let A be a finite set, $|A| < \infty$: $(2^A, \cup, \cap, \emptyset, A)$.

9. Multisets. Let $(R, +, \cdot, 0, 1)$ be a semiring and $a = \{(a_i, n_i) : i \in I\}$, $a_i \in A$, $n_i \in R$ a multiset.

$$a \oplus b = \left\{ \begin{array}{lcl} (a_i, n_i + n_j) & : & a_i, b_j \in a \cap b \wedge a_i = b_j \\ (a_i, n_i) & : & a_i \notin b \\ (b_j, n_j) & : & b_j \notin a \end{array} \right\}$$

$$a \odot b = \{(a_i b_j, n_i \cdot n_j) : a_i \in a \wedge b_j \in b\}$$

Intervals

10 Interval 1 [18, 2]: $[a, A], [b, B] \subset \mathbb{R}_0^+$
$[a, A] \oplus [b, B] = [a + b, A + B]$ and
$[a, A] \odot [b, B] = [a \cdot b, A \cdot B]$

11 Interval 2: $[a, A], [b, B] \subset \mathbb{R}$
$[a, A] \oplus [b, B] = [\min(a, b), \max(A, B)]$ and
$[a, A] \odot [b, B] = [a + b, A + B]$

12 Interval 3:

$$[a, A] \oplus [b, B] = \begin{cases} [a, A] & A < B \\ [b, B] & B < A \\ [\min(a, b), A] & A = B \end{cases}$$

Histograms

13 Let the set of bins $\mathbf{B} = \{B_1, B_2, \ldots, B_k\}$ be a partition of the set B such that $(\mathbf{B}, \circ)$ is a semigroup. A *histogram* $h : \mathbf{B} \to \mathbb{N}$
$h_i = h(B_i) = |\{X : v(X) \in B_i\}|$

$$h \oplus g = h + g \quad (h \oplus g)(i) = h(i) + g(i)$$

$h \odot g = h * g$ convolution [13, 10]

$$(h * g)(i) = \sum_{p \circ q = i} h(p) \cdot g(q)$$

Temporal quantities

14 A temporal quantity (TQ) a is a function $a : \mathcal{T} \to A \cup \{\maltese\}$ where $\maltese$ denotes the value *undefined*. $(A, +, \cdot, 0, 1)$ is a semiring. The *activity time set* T_a of a consists of instants $t \in T_a$ in which a is defined $T_a = \{t \in \mathcal{T} : a(t) \in A\}$.

We can extend both operations to the set $A_\maltese = A \cup \{\maltese\}$ by requiring that for all $a \in A_\maltese$ it holds $a + \maltese = \maltese + a = a$ and $a \cdot \maltese = \maltese \cdot a = \maltese$.
The structure $(A_\maltese, +, \cdot, \maltese, 1)$ is also a semiring.

Let $A_\maltese(\mathcal{T})$ denote the set of all TQs over $A_\maltese$ in time $\mathcal{T}$. To extend the operations to networks and their matrices we first define the *sum* (parallel links) $a + b$ as

$$(a + b)(t) = a(t) + b(t) \quad \text{and} \quad T_{a+b} = T_a \cup T_b.$$

The *product* (sequential links) $a \cdot b$ is defined as

$$(a \cdot b)(t) = a(t) \cdot b(t) \quad \text{and} \quad T_{a \cdot b} = T_a \cap T_b.$$

Temporal quantities

Let us define TQs **0** and **1** with requirements $\mathbf{0}(t) = \maltese$ and $\mathbf{1}(t) = 1$ for all $t \in \mathcal{T}$. Again, the structure $(A_{\maltese}(\mathcal{T}), +, \cdot, \mathbf{0}, \mathbf{1})$ is a semiring.

To produce software support for computation with TQs we limit it to TQs that can be described as a sequence of disjoint time intervals with a constant value

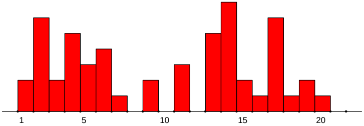
$$a = [(s_i, f_i, v_i)]_{i \in 1..k}$$

where s_i is the starting time and f_i the finishing time of the i-th time interval $[s_i, f_i)$, $s_i < f_i$ and $f_i \leq s_{i+1}$, and v_i is the value of a on this interval (over combinatorial semiring). Outside the intervals, the value of TQ a is undefined, $\maltese$.

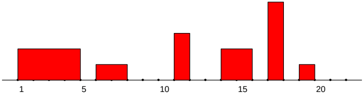
See also [6, 20]; and another approach based on semirings [17, 14]

Sum and product of temporal quantities

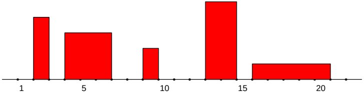
$a + b$:



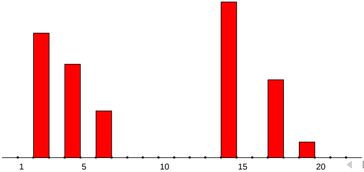
a :



b :



$a \cdot b$:



$$a = [(1,5,2),(6,8,1),(11,12,3),(14,16,2),(17,18,5),(19,20,1)]$$

$$b = [(2,3,4),(4,7,3),(9,10,2),(13,15,5),(16,21,1)]$$

Conclusions

- A framework for symbolic networks is presented based on experiences with analysis of temporal networks.
- For extending research to other types of symbolic networks the corresponding data sets are needed – creating a repository of symbolic data / networks on GitHub? Well documented!
- SDAjson – json format for SDA?
- New symbolic semirings.

The last version of these slides is available at https://github.com/bavla/SDA

Acknowledgments

This work is supported in part by the Slovenian Research Agency (research program P1-0294 and research projects J1-9187 and J5-2557), and prepared within the framework of the HSE University Basic Research Program.

References I

- Abdali SK, Saunders BD (1985) Transitive closure and related semiring properties via eliminants. Theor Com-put Sci 40:257—274
- Alves HFC (2020) Interval-Weighted Networks: Community Detection and Centrality Measures. PhD thesis Faculdade de Ciências da Universidade do Porto, Universidade de Aveiro e Universidade do Minho Matemática Aplicada.
- Baras JS, Theodorakopoulos G (2010) Path problems in networks. Morgan & Claypool, Berkeley
- Batagelj V (1994) Semirings for social networks analysis. J Math Soc 19(1):53-–68
- Batagelj V (2017) Symbolic network analysis of bike sharing data / Citi Bike. 6th Symbolic Data Analysis workshop, Ljubljana, 12-14. June 2017.

References II

- Batagelj V, Praprotnik S (2016) An algebraic approach to temporal network analysis based on temporal quantities. Social Network Analysis and Mining 6(1): 1–22
- Batagelj V (2018) Complex aggregation of large data sets. Slides from 7th Symbolic Data Analysis workshop, Viana do Castelo, 18-20. October 2018.
- Batagelj V, Maltseva D (2020) Temporal bibliographic networks. Journal of Informetrics, 14(1) 101006.
- Brandes U (2001) A faster algorithm for betweenness centrality. The Journal of Mathematical Sociology, Volume 25, Issue 2, 163–177
- Brunotte H (2013) Discrete Convolution Rings. Romanian Journal of Mathematics and Computer Science, Volume 3, Issue 2, p.155–159
- Carré B (1979) Graphs and networks. Clarendon, Oxford

References III

- Cerinšek M, Batagelj V (2017) Semirings and Matrix Analysis of Networks. In: Alhajj R., Rokne J. (eds) Encyclopedia of Social Network Analysis and Mining. Springer, New York, NY.
- Dongol B, Hayes IJ, Struth G (2016) Convolution as a Unifying Concept: Applications in Separation Logic, Interval Calculi and Concurrency ACM Transactions on Computational Logic, Article No.: 15
- Falzon L, Quintane E, Dunn J, Robins G (2018) Embedding time in positions: Temporal measures of centrality for social network analysis. Social Networks 54, 168–178
- Glazek K (2002) A Guide to the Literature on Semirings and their Applications in Mathematics and Information Sciences. Springer.
- Gondran M, Minoux M (2008) Graphs, dioids and semirings: new models and algorithms. Springer, New York

References IV

- Kontoleon N, Falzon L, Pattison P (2013) Algebraic structures for dynamic networks. Journal of Mathematical Psychology 57, 310–319
- Moore RE, Kearfott RB, Cloud MJ (2009). Introduction to Interval Analysis. Philadelphia: SIAM.
- Ostoic JAR (2021) Algebraic Analysis of Social Networks: Models, Methods and Applications Using R. Wiley
- Praprotnik S, Batagelj V (2016) Semirings for temporal network analysis. arXiv:1603.08261 [cs.SI]
- Schvaneveldt RW, Dearholt DW, Durso FT (1988) Graph theoretic foundations of Pathfinder networks. Comput Math Appl 15(4):337–345

References V

- Quirin A, Cordón O, Santamaria J, Vargas-Quesada B, Moya-Anegón F (2008) A new variant of the Pathfinder algorithm to generate large visual science maps in cubic time. Inf Process Manag 44(4): 1611–1623
- Vavpetič A, Batagelj V, Podpečan V (2009) An implementation of the Pathfinder algorithm for sparse networks and its application on text networks. SiKDD.
- Wikipedia: Regular language.