Semirings in network data analysis

an overview

Vladimir Batagelj

IMFM Ljubljana, IAM UP Koper, and NRU HSE Moscow

Mathematics for Social Sciences and Arts

Algebraic Modeling

MS^{2}A^{2}M 2021

on Zoom, May 24-26, 2021

Outline

1 Semirings
2 Examples
3 Kinship relations
4 References

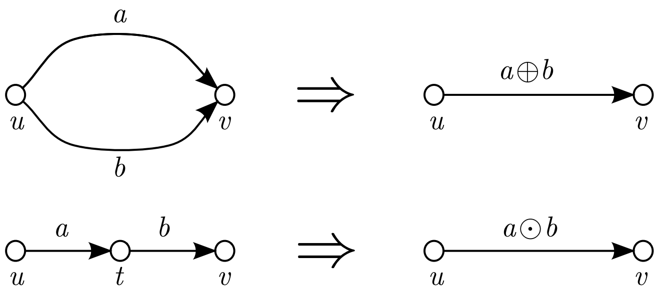
Vladimir Batagelj: vladimir.batagelj@fmf.uni-lj.si
Current version of slides (May 23, 2021 at 07 : 17): slides PDF
https://github.com/bavla/semirings

Semirings

Claude Pair (2019)

I became interested in networks and semirings already as a student [3]. My colleague Tomaž Pisanski studied for a year in Nancy, France. He provided me with a copy of lectures of Claude Pair on networks [23, 28].
I submitted (unsuccessfuly) my results to the IFIP 1971 Conference that was held in Ljubljana.

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 $\mathbb{K}$ be a set and a, b, c elements from $\mathbb{K}$. A *semiring* (Abdali and Saunders 1985 [1]; Carré 1979; Baras and Theodorakopoulos 2010; Batagelj 1994) is an algebraic structure $(\mathbb{K}, \oplus, \odot, 0, 1)$ with two binary operations (addition $\oplus$ and multiplication $\odot$) where:

$(\mathbb{K}, \oplus, 0)$ 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

$(\mathbb{K}, \odot, 1)$ 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 $\oplus$:
$a \odot (b \oplus c) = a \odot b \oplus a \odot c$ $\quad$ $(b \oplus c) \odot a = b \odot a \oplus c \odot a$

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

Semiring

A semiring $(\mathbb{K}, \oplus, \odot, 0, 1)$ 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 $a \in \mathbb{K}$. In this case the semiring over a finite set $\mathbb{K}$ is complete.

Semiring

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

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 $(\mathbb{K}, \oplus, \odot, 0, 1)$ the *absorption* law holds iff for all $a, b, c \in \mathbb{K}$:
$a \odot b \oplus a \odot c \odot b = a \odot b$.
It is sufficient for the absorption law to check the property $1 \oplus c = 1$ for all $c \in \mathbb{K}$ because of the distributivity.

Semiring

We present an overview of semirings used in network data analysis, and network matrices and vectors over a semiring (addition, multiplication, power, closure). We conclude with some research directions. Supporting materials (slides, software, data sets, and other resources) will be available at
Godran Glazek

Some examples

Some examples of 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: $(\mathbb{R}_0^+, \min, +, \infty, 0)$
4. Pathfinder [26, 25, **?**]: $(\overline{\mathbb{R}}_0^+, \min, \boxed{r}, \infty, 0)$
   where $a \boxed{r} b = \sqrt[r]{a^r + b^r}$ (Minkowski)
5. Interval [21, **?**]: $[a, A], [b, B] \subset \mathbb{R}_0^+$
   $[a, A] \oplus [b, B] = [a + b, A + B]$ and
   $[a, A] \odot [b, B] = [a \cdot b, A \cdot B]$

Harary
Ostoic

World trade 1999 network

Semirings in network data analysis

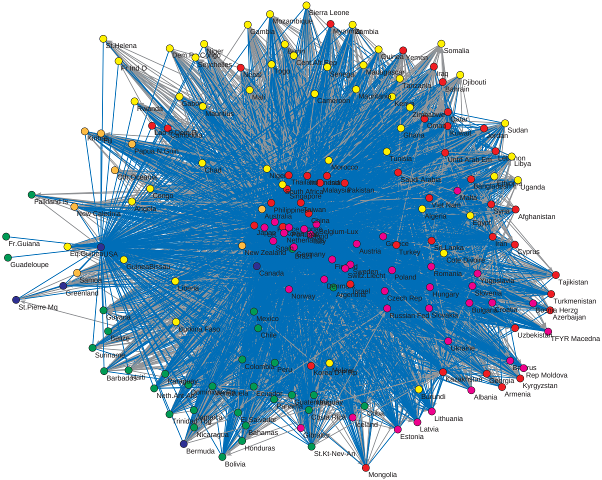
V. Batagelj

Semirings

Examples

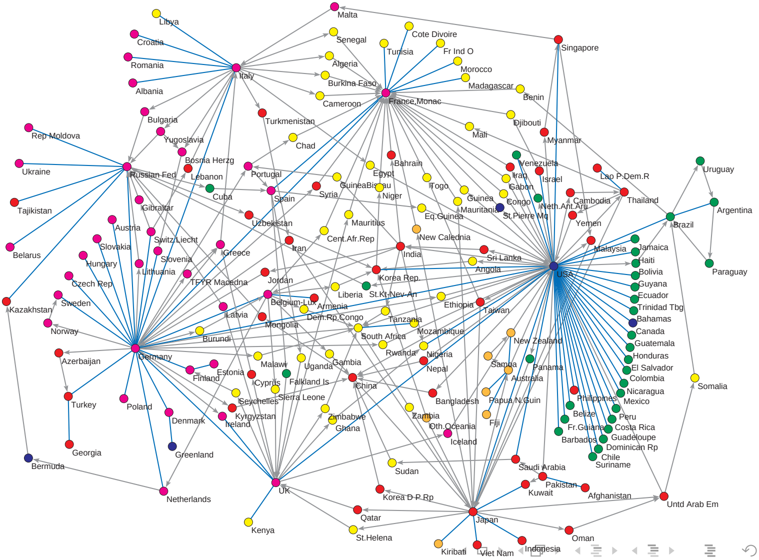
Kinship relations

References



understanding

World trade 1999 Pathfinder skeleton



Geodetic semiring

6 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). It was used in an algorithm for computing the betweenness in a network. In 2001 Brandes proposed a faster algorithm [11].

Balance semiring for signed networks

7 To construct a semiring corresponding to the balance problem we take the set A with four elements [4]

0	no walk;
n	all walks are negative;
p	all walks are positive;
a	at least one positive and at least one negative walk.

$\oplus$	0	n	p	a
0	0	n	p	a
n	n	n	a	a
p	p	a	p	a
a	a	a	a	a

$\odot$	0	n	p	a
0	0	0	0	0
n	0	p	n	a
p	0	n	p	a
a	0	a	a	a

x	$x^\star$
0	p
n	a
p	p
a	a

Partition semiring for signed networks

8 The set A with five elements [4]

0	no walk;
n	at least one walk with exactly one negative arc; no walk with only positive arcs;
p	at least one walk with only positive arcs; no walk with exactly one negative arc;
a	at least one walk with only positive arcs; at least one walk with exactly one negative arc;
q	each walk has at least two negative arcs.

$\oplus$	0	n	p	a	q
0	0	n	p	a	q
n	n	n	a	a	n
p	p	a	p	a	p
a	a	a	a	a	a
q	q	n	p	a	q

$\odot$	0	n	p	a	q
0	0	0	0	0	0
n	0	q	n	n	q
p	0	n	p	a	q
a	0	n	a	a	q
q	0	q	q	q	q

x	$x^\star$
0	p
n	a
p	p
a	a
q	p

Histograms

9 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 [15, 10]

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

Temporal quantities

10 A temporal quantity (TQ) a is a function $a : \mathcal{T} \to A \cup \{\maltese\}$ where $\maltese$ denotes the value *undefined*. 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 $\mathbf{0}$ and $\mathbf{1}$ 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.

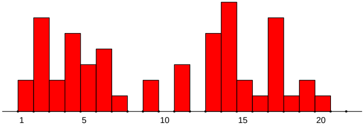
To produce a software support for computation with TQs we limit it to TQs that can be described as a sequence of disjoint time intervals with a constant value

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

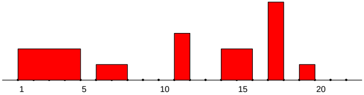
where s_i is the starting time and f_i the finishing time of the i-th time interval $[s_i, f_i)$, $s_i < f_i$ and $f_i \leq s_{i+1}$, and v_i is the value of a on this interval (over combinatorial semiring). Outside the intervals the value of TQ a is undedined, $\maltese$.

Sum and product of temporal quantities

$a \oplus b$:

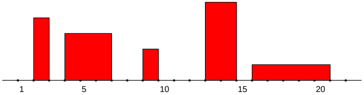


a :



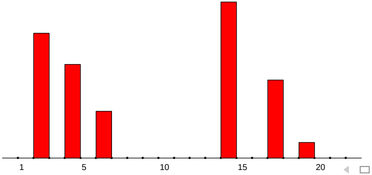
$$a = [(1,5,2),(6,8,1),(11,12,3),(14,16,2),(17,18,5),(19,20,1)]$$

b :



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

$a \odot b$:



Application

Bibliographic networks [9]

Kinship relations

Anthropologists typically use a basic vocabulary of kin types to represent genealogical relationships. One common version of the vocabulary for basic relationships:

```
Kin Type     English Type
 P           Parent
   F         Father
   M         Mother
 C           Child
   D         Daughter
   S         Son
 G           Sibling
   Z         Sister
   B         Brother
 E           Spouse
   H         Husband
   W         Wife
```

The genealogies are usually described in GEDCOM format. Examples family, Bouchards. Paper

Calculating kinship relations

Pajek generates three relations when reading genealogy as Ore graph:
F: _ **is a father of** _
M: _ **is a mother of** _
E: _ **is a spouse of** _
Additionally we must generate two binary diagonal matrices, to distinguish between male and female:
L: _ **is a male** _ / 1-male, 0-female
J: _ **is a female** _ / 1-female, 0-male

$$\mathbf{F} \cap \mathbf{M} = \emptyset, \qquad \mathbf{L} \cup \mathbf{J} \subseteq \mathbf{I}, \qquad \mathbf{L} \cap \mathbf{J} = \emptyset$$

Derived kinship relations

Other basic relations can be obtained using macros based on identities:

_ **is a parent of** _	P	$=$	$F \cup M$
_ **is a child of** _	C	$=$	P^T
_ **is a son of** _	S	$=$	$L * C$
_ **is a daughter of** _	D	$=$	$J * C$
_ **is a husband of** _	H	$=$	$L * E$
_ **is a wife of** _	W	$=$	$J * E$
_ **is a sibling of** _	G	$=$	$((F^T * F) \cap (M^T * M)) \setminus I$
_ **is a brother of** _	B	$=$	$L * G$
_ **is a sister of** _	Z	$=$	$J * G$
_ **is an uncle of** _	U	$=$	$B * P$
_ **is an aunt of** _	A	$=$	$Z * P$
_ **is a semi-sibling of** _	G_e	$=$	$(P^T * P) \setminus I$

and using them other relations can be determined

_ **is a grand mother of** _	M_2	$=$	$M * P$
_ **is a niece of** _	Ni	$=$	$D * G$

Conclusions

Python Hamilton
Kinship relations
Lattices, Boolean algebras, Regular languages

References I

- Abdali SK, Saunders BD (1985) Transitive closure and related semiring properties via eliminants. Theor Com-put Sci 40:257–274
- Baras JS, Theodorakopoulos G (2010) Path problems in networks. Morgan & Claypool, Berkeley
- Batagelj V (1970) Nekaj malega o kvazi-kolobarjih, v-grafih in vrednostnih matrikah. Ljubljana.
- Batagelj V (1994) Semirings for social networks analysis. J Math Soc 19(1):53–68
- Batagelj V, Cerinšek M (2013) On bibliographic networks. Scientometrics 96(3):845–864
- Batagelj V, Mrvar A (2008) Analysis of kinship relations with Pajek. Soc Sci Comput Rev 26(2): 224–246
- Batagelj, Doreian, Ferligoj, Kejžar
- 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

References II

Batagelj V, Maltseva D(2020): Temporal bibliographic networks. Journal of Informetrics, 14 (1), 101006.

Brunotte H (2013) Discrete Convolution Rings. Romanian Journal of Mathematics and Computer Science, Volume 3, Issue 2, p.155-159

Brandes U (2001) Between

Burkard RE, Cuninghame-Greene RA, Zimmermann U (eds) (1984) Algebraic and combinatorial methods in operations research. Annals of discrete mathematics, vol 19. North Holland, Amsterdam/New York

Carré B (1979) Graphs and networks. Clarendon, Oxford

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

References III

- Fletcher JG (1980) A more general algorithm for computing closed semiring costs between vertices of a directed graph. Commun ACM 23(6): 350–351
- Glazek K (2002) A Guide to the Literature on Semirings and their Applications in Mathematics and Information Sciences. Springer.
- Golan JS (1999) Semirings and their Applications. Springer.
- Gondran M, Minoux M (2008) Graphs, dioids and semirings: new models and algorithms. Springer, New York
- Kepner J, Gilbert J (2011) Graph algorithms in the language of linear algebra. SIAM, Philadelphia
- 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
- Pair C (1969) Notions sur la théorie des graphes. Université de Nancy, Nancy 1969/70.

References IV

Praprotnik S, Batagelj 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

Schvaneveldt RW, Dearholt DW, Durso FT (1988) Graph theoretic foundations of Pathfinder networks. Comput Math Appl 15(4):337–345

Github bavla/semirings https://github.com/bavla/semirings ; Refs

La découverte scientifique . . . qui mérite le mérite ? Le Monde / Binaire, July 6 2020. https://www.lemonde.fr/blog/binaire/tag/claude-pair/