Formal definition of Bounded Linear Spatial Temporal Logic (BLSTL)

Syntax

The syntax of BLSTL will be defined by a context-free grammar using the Backus-Naur Form notation. A definition of a non-terminal symbol (element) in such grammars has the following form:

\[
\langle \text{defined-element} \rangle ::= \langle \text{element1} \rangle \\
| \langle \text{element2} \rangle \\
| \ldots
\]

where ::= introduces a new definition and | represents an alternative. In natural language this reads \(<\text{defined-element}>\) is either \(<\text{element1}>\) or \(<\text{element2}>\) or (...).

In contrast to the BLTL definition the symbol \(\phi\) was replaced by the non-terminal symbol \(<\text{logic-property}>\). Furthermore new spatial (e.g. \(<\text{numeric-spatial-measure}>\)) and arithmetic (e.g. \(<\text{mean}>\)) functions were introduced.

Definition. The syntax of BLSTL is given by the following grammar formally expressed in Backus-Naur Form:

\[
\langle \text{logic-property} \rangle ::= \langle \text{numeric-spatial-measure} \rangle \langle \text{comparator} \rangle \langle \text{numeric-measure} \rangle \\
| \langle \text{numeric-state-variable} \rangle \langle \text{comparator} \rangle \langle \text{numeric-measure} \rangle \\
| d(\langle \text{numeric-measure} \rangle) \langle \text{comparator} \rangle \langle \text{numeric-measure} \rangle \\
| \neg \langle \text{logic-property} \rangle \\
| \langle \text{logic-property} \rangle \lor \langle \text{logic-property} \rangle \\
| \langle \text{logic-property} \rangle U[\langle \text{natural-number} \rangle, \langle \text{natural-number} \rangle] \langle \text{logic-property} \rangle \\
| X \langle \text{logic-property} \rangle \\
| (\langle \text{logic-property} \rangle)
\]

\[
\langle \text{numeric-measure} \rangle ::= \langle \text{numeric-spatial-measure} \rangle \\
| \langle \text{real-number} \rangle \\
| \langle \text{numeric-state-variable} \rangle \\
| \langle \text{unary-numeric-measure} \rangle(\langle \text{numeric-measure} \rangle) \\
| \langle \text{binary-numeric-measure} \rangle(\langle \text{numeric-measure} \rangle, \langle \text{numeric-measure} \rangle)
\]

\[
\langle \text{numeric-spatial-measure} \rangle ::= \langle \text{unary-subset-measure} \rangle(\langle \text{subset} \rangle) \\
| \langle \text{binary-subset-measure} \rangle(\langle \text{subset} \rangle, \langle \text{spatial-measure} \rangle) \\
| \langle \text{ternary-subset-measure} \rangle(\langle \text{subset} \rangle, \langle \text{spatial-measure} \rangle, \langle \text{real-number} \rangle) \\
| \langle \text{quaternary-subset-measure} \rangle(\langle \text{subset} \rangle, \langle \text{spatial-measure} \rangle, \langle \text{subset} \rangle, \langle \text{spatial-measure} \rangle)
\]

\[
\langle \text{unary-subset-measure} \rangle ::= \text{count} \\
| \text{clusteredness} \\
| \text{density}
\]

\[
\langle \text{binary-subset-measure} \rangle ::= \text{avg} \\
| \text{geomean} \\
| \text{harmean} \\
| \text{kurt} \\
| \text{max}
\]
median  
min  
mode  
product  
skew  
stdev  
sum  
var

⟨ternary-subset-measure⟩ ::= percentile  
| quartile

⟨quaternary-subset-measure⟩ ::= covar

⟨unary-numeric-measure⟩ ::= abs  
| ceil  
| floor  
| round  
| sign  
| sqrt  
| trunc

⟨binary-numeric-measure⟩ ::= add  
| div  
| log  
| mod  
| multiply  
| power  
| subtract

⟨subset⟩ ::= ⟨subset-specific⟩  
| filter(⟨subset-specific⟩, ⟨constraint⟩)

⟨subset-specific⟩ ::= regions  
| clusters

⟨constraint⟩ ::= ¬ ⟨constraint⟩  
| ⟨constraint⟩ ∨ ⟨constraint⟩  
| ⟨spatial-measure⟩ ⟨comparator⟩ ⟨filter-numeric-measure⟩

⟨filter-numeric-measure⟩ ::= ⟨numeric-measure⟩  
| ⟨spatial-measure⟩  
| ⟨unary-numeric-measure⟩ ⟨filter-numeric-measure⟩  
| ⟨binary-numeric-measure⟩ ⟨filter-numeric-measure⟩, ⟨filter-numeric-measure⟩

⟨spatial-measure⟩ ::= clusteredness  
| density  
| area  
| perimeter  
| distanceFromOrigin  
| angle
Similarly to BLTL a set of additional operators can be derived from the definition of the BLSTL syntax using the following equivalences:

1. **Boolean**: (applies to both <logic-property> and <constraint> elements)
   - \( p \land q \equiv \neg(\neg p \lor \neg q) \)
   - \( p \Rightarrow q \equiv \neg p \lor q \)
   - \( p \Leftrightarrow q \equiv (p \Rightarrow q) \land (q \Rightarrow p) \)

2. **Temporal**:
   - \( F[a, b] \) \( p \equiv \text{True} \ U[a, b] \) \( p \)
   - \( G[a, b] \) \( p \equiv \neg F[a, b] \) \( \neg p \)
   - \( X[k] \) \( p \equiv \underbrace{X\ldots X}_{k \text{ times}} \) \( p \)

The order of precedence of the operators is given by the definition of the extended BLSTL syntax (see Appendix A). In the absence of parentheses the logic expressions are evaluated from left to right.

**Semantics**

The semantics of BLSTL is defined with respect to executions/simulations of a SS\(\mu\)DES \( M \); see [1, 2] for similar examples. Let us assume that

\[ \sigma = (s_0, t_0), (s_1, t_1), \ldots \]

is an execution of \( M \) along the sequence of states \( s_0, s_1, \ldots \) with \( t_0, t_1, \ldots \in \mathbb{R} \) time durations spent in each state. The execution trace starting at the \( k \)-th state is denoted by \( \sigma^k \), the length of the execution trace by \( |\sigma| \), the \( i \)-th state of the execution trace by \( \sigma[i] \), and the fact that the execution \( \sigma \) satisfies a property
\(\phi\) by \(\sigma \models \phi\). For an execution \(\sigma\) at state \(s\) the value of a numeric state variable \(nsv\) is given by \(NV(\sigma, s, nsv)\) and the value of a spatial state variable \(ssv\) is given by \(SpV(\sigma, s, ssv)\).

In order to have a compact and easy to follow semantics’ description the full symbol names provided in the BLSTL syntax definition were replaced with shorter abbreviations as described in Table 1.

Table 1: Translation of full BLSTL symbol names to abbreviated forms. The left column contains the full BLSTL symbol name. The right column contains the corresponding abbreviated form.

<table>
<thead>
<tr>
<th>Full BLSTL symbol name</th>
<th>Abbreviated BLSTL symbol name</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;numeric-measure&gt;</td>
<td>nm</td>
</tr>
<tr>
<td>&lt;unary-numeric-measure&gt;</td>
<td>unm</td>
</tr>
<tr>
<td>&lt;binary-numeric-measure&gt;</td>
<td>bnm</td>
</tr>
<tr>
<td>&lt;numeric-spatial-measure&gt;</td>
<td>nsm</td>
</tr>
<tr>
<td>&lt;unary-subset-measure&gt;</td>
<td>usm</td>
</tr>
<tr>
<td>&lt;binary-subset-measure&gt;</td>
<td>bsm</td>
</tr>
<tr>
<td>&lt;ternary-subset-measure&gt;</td>
<td>tsm</td>
</tr>
<tr>
<td>&lt;quaternary-subset-measure&gt;</td>
<td>qsm</td>
</tr>
<tr>
<td>&lt;subset&gt;</td>
<td>ss</td>
</tr>
<tr>
<td>&lt;spatial-measure&gt;</td>
<td>sm</td>
</tr>
<tr>
<td>&lt;filter-numeric-measure&gt;</td>
<td>fnm</td>
</tr>
<tr>
<td>&lt;comparator&gt;</td>
<td>(\sim)</td>
</tr>
<tr>
<td>&lt;numeric-state-variable&gt;</td>
<td>nsv</td>
</tr>
<tr>
<td>&lt;spatial-state-variable&gt;</td>
<td>ssv</td>
</tr>
<tr>
<td>&lt;meta-spatial-state-variable&gt;</td>
<td>mssv</td>
</tr>
<tr>
<td>&lt;logic-property&gt;</td>
<td>(\psi)</td>
</tr>
<tr>
<td>&lt;real-number&gt;</td>
<td>(r)</td>
</tr>
<tr>
<td>&lt;natural-number&gt;</td>
<td>(n)</td>
</tr>
</tbody>
</table>

**Definition.** Let \(M = (S, T, \mu, NSV, SpSV, NV, SpV)\) be a SSpDES and \(\sigma\) an execution of \(M\). The semantics of BLSTL for \(\sigma\) is defined as follows:

- \(\sigma \models nsm \preceq nm\) if and only if \(nsm \preceq nm\), where \(nsm\) and \(nm\) \(\in\) \(\mathbb{R}\) and \(\preceq \in \{\geq, =, <, <\}, \leq\};
- \(\sigma \models nsv \succeq nm\) if and only if \(NV(\sigma, \sigma[0], nsv) \succeq nm\), where \(nsv\) \(\in\) \(NSV\), \(nm\) \(\in\) \(\mathbb{R}\) and \(\succeq \in \{\geq, =, <, <\}, \leq\};
- \(\sigma \models d(nm1) \approx nm2\) if and only if \(|\sigma| > 1\) and \((nm1^1 - nm1^0) \approx nm2\), where \(nm1^1 \in \mathbb{R}\) represents the value of the numeric measure \(nm1\) considering the execution suffix \(\sigma^1\) and \(\approx \in \{\geq, =, <, <\}, \leq\};
- \(\sigma \models \neg \psi\) if and only if \(\sigma \not\models \psi\);
- \(\sigma \models \psi_1 \lor \psi_2\) if and only if \(\sigma \models \psi_1\) or \(\sigma \models \psi_2\);
- \(\sigma \models \psi_1 U[a, b] \psi_2\) if and only if there exists \(i\), \(a \leq i \leq b\) such that \(\sigma^i \models \psi_2\), and for all \(j\), \(a \leq j < i\), \(\sigma^j \models \psi_1\);
• $\sigma \models X\psi$ if and only if $|\sigma| > 1$ and $\sigma^1 \models \psi$;

• $\sigma \models (\psi)$ if and only if $\sigma \models \psi$;

The $nm$ symbol represents the category of real-valued numeric measures. Considering a given execution $\sigma$, $nm$ is evaluated according to one of the definitions described below:

- **Numeric spatial measure**: $nm = nsm$;
- **Real number**: $nm = r \in \mathbb{R}$;
- **Numeric state variable**: $nm = NV(\sigma, \sigma[0], nsv)$ where $nsv$ is a numeric state variable;
- **Unary numeric measure**: $nm = unm(nm')$ where $nm'$ is a numeric measure;
- **Binary numeric measure**: $nm = bnm(nm', nm'')$ where $nm'$ and $nm''$ are numeric measures.

The values of the **unary (unm)** and **binary (bnm)** numeric measures are computed as described in Appendix B.

The $nsm$ symbol represents the category of numeric (real-valued) spatial measures. Considering a given execution $\sigma$, $nsm$ is evaluated according to one of the definitions described below:

- **Unary subset measure**: $nsm = usm(ss)$ where $ss$ is a subset of the considered spatial entities (clusters or regions);
- **Binary subset measure**: $nsm = bsm(ss, sm)$ where $ss$ is a subset of the considered spatial entities (clusters or regions) and $sm$ is a spatial measure;
- **Ternary subset measure**: $nsm = tsm(ss, sm, r)$ where $ss$ is a subset of the considered spatial entities (clusters or regions), $sm$ is a spatial measure and $r$ is a real value;
- **Quaternary subset measure**: $nsm = qsm(ss, sm, ss', sm')$ where $ss$ and $ss'$ are subsets of the considered spatial entities (clusters or regions), and $sm$ and $sm'$ are spatial measures;

In case the considered subset of spatial entities is empty the numeric spatial measures are evaluated to zero.

Spatial measures $sm$ are defined over the set \{clusteredness, density, area, perimeter, distanceFromOrigin, angle, triangleMeasure, rectangleMeasure, circleMeasure, centroidX, centroidY\} which is identical to the set of spatial measures recorded in an STML file for each detected region/cluster.

The value of **unary (usm)**, **binary (bsm)**, **ternary (tsm)** and **quaternary (qsm)** subset measures are computed as described in Appendix C. Some of the binary and all ternary and quaternary subset measures are statistical functions which can be employed for reasoning about the distribution of the regions/-clusters measures at a particular time point. In contrast to traditional logic formalisms BLSTL allows specifying properties of both single spatial properties and/or distributions of spatial properties.
Subsets of the collections of regions/clusters are represented by the \( ss \) symbol. Considering a given execution \( \sigma \), \( ss \) is evaluated according to one of the definitions described below:

- **Specific subset**: \( ss = \text{specificSubset} \) where \( \text{specificSubset} \) represents either the collection of all clusters (see Definition 4) or the collection of all regions (see Definition 3) corresponding to \( \sigma[0] \);

- **Filtered specific subset**: \( ss = \text{filter}(\text{specificSubset}, \text{constraints}) \) where \( \text{specificSubset} \) has the semantics defined above, and \( \text{constraints} \) is a set of logic properties restricting the considered spatial entities to a subset of \( \text{specificSubset} \);

Given an execution \( \sigma \) the value of the \( \text{specificSubset} \) symbol is computed using one of the definitions described below:

- **Regions**: \( \text{specificSubset} = \bigcup_{ssv \in \text{SpSV}} \{ \text{region} \mid \text{region} \in \text{regionDetectionMechanism}(ssv) \} \), \( \forall ssv \in \text{SpSV} \) considering the state \( \sigma[0] \);

- **Clusters**: \( \text{specificSubset} = \text{clustersDetectionMechanism}(se) \), where \( se = \bigcup_{ssv \in \text{SpSV}} \{ \text{region} \mid \text{region} \in \text{regionDetectionMechanism}(ssv) \} \), \( \forall ssv \in \text{SpSV} \) considering the state \( \sigma[0] \);

Subsets of the collection returned by \( \text{specificSubset} \) can be computed using the \( \text{filter} \) predicate. Considering an execution \( \sigma \) \( \text{filter} \) is evaluated using the definition described below:

\[
\text{filter} = \{ e \in \text{specificSubset} \mid e \models c, \forall c \in \text{constraints} \}.
\]

The semantics of the constraint satisfaction problem considering a region/cluster \( e \) and a constraint \( c \) is defined below:

- \( e \models sm \bowtie fnm \text{ if and only if } sm(e) \bowtie fnm \), where \( sm(e) \) evaluates the spatial measure \( sm \) for the given spatial entity \( e \), \( sm \in \{ \text{clusteredness, density, area, perimeter, distanceFromOrigin, angle, triangleMeasure, rectangleMeasure, circleMeasure, centroidX, centroidY} \} \), \( fnm \) is a filter numeric measure \( \in \mathbb{R} \) and \( \bowtie \in \{ >, >=, <, <=, = \} \);

- \( e \models \neg c \text{ if and only if } e \not\models c \);

- \( e \models c_1 \lor c_2 \text{ if and only if } e \models c_1 \text{ or } e \models c_2 \);

The \( fnm \) symbol represents the (real-valued) numeric measure computed for the \( \text{filter} \) predicate. Given an execution \( \sigma \) and a region/cluster \( e \), the value of \( fnm \) is computed using one of the definitions given below:

- **Numeric measure**: \( fnm = nm \) where \( nm \) is a numeric measure;

- **Spatial measure**: \( fnm = sm(e) \) where \( sm \in \{ \text{clusteredness, density, area, perimeter, distanceFromOrigin, angle, triangleMeasure, rectangleMeasure, circleMeasure, centroidX, centroidY} \} \);

- **Unary filter numeric measure**: \( fnm = unm(fnm') \) where \( fnm' \) is a filter numeric measure;
- **Binary filter numeric measure**: \( fnm = bnm(fnm', fnm'') \) where \( fnm' \) and \( fnm'' \) are filter numeric measures.

The extended semantics of BLSTL for an execution \( \sigma \) (corresponding to the extended syntax of BLSTL) additionally includes the following rules:

1. **Boolean**:
   - \( \sigma \models \psi_1 \land \psi_2 \) if and only if \( \sigma \models \psi_1 \) and \( \sigma \models \psi_2 \);
   - \( \sigma \models \psi_1 \Rightarrow \psi_2 \) if and only if \( \sigma \models \neg \psi_1 \) or \( \sigma \models \psi_2 \);
   - \( \sigma \models \psi_1 \Leftrightarrow \psi_2 \) if and only if \( \sigma \models \psi_1 \Rightarrow \psi_2 \) and \( \sigma \models \psi_2 \Rightarrow \psi_1 \);

2. **Temporal**:
   - \( \sigma \models F[a,b] \psi \) if and only if there exists \( i, a \leq i \leq b \) such that \( \sigma^i \models \psi \);
   - \( \sigma \models G[a,b] \psi \) if and only if for all \( i, a \leq i \leq b, \sigma^i \models \psi \);
   - \( \sigma \models X[k] \psi \) if and only if \( |\sigma| > k \) and \( \sigma^k \models \psi \);

Conversely the extended semantics of BLSTL for a spatial entity \( e \) and constraint \( c \) (corresponding to the extended syntax of BLSTL) additionally includes the following Boolean rules:

- \( e \models c_1 \land c_2 \) if and only if \( e \models c_1 \) and \( e \models c_2 \);
- \( e \models c_1 \Rightarrow c_2 \) if and only if \( e \models \neg c_1 \) or \( e \models c_2 \);
- \( e \models c_1 \Leftrightarrow c_2 \) if and only if \( e \models c_1 \Rightarrow c_2 \) and \( e \models c_2 \Rightarrow c_1 \);

**References**


Appendices

Appendix A  Extended BLSTL syntax

Definition. Let $M = \langle S, T, \mu, NSV, SpSV, NV, SpV \rangle$ be a SSpDES. The extended syntax of BLSTL is given by the following grammar expressed in Backus-Naur Form (BNF):

$$
\langle \text{logic-property} \rangle ::= \langle \text{numeric-spatial-measure} \rangle \langle \text{comparator} \rangle \langle \text{numeric-measure} \rangle \\
| \langle \text{numeric-state-variable} \rangle \langle \text{comparator} \rangle \langle \text{numeric-measure} \rangle \\
| d(\langle \text{numeric-measure} \rangle) \langle \text{comparator} \rangle \langle \text{numeric-measure} \rangle \\
| \neg \langle \text{logic-property} \rangle \\
| \langle \text{logic-property} \rangle \land \langle \text{logic-property} \rangle \\
| \langle \text{logic-property} \rangle \lor \langle \text{logic-property} \rangle \\
| \langle \text{logic-property} \rangle \Rightarrow \langle \text{logic-property} \rangle \\
| \langle \text{logic-property} \rangle \Leftrightarrow \langle \text{logic-property} \rangle \\
| \langle \text{logic-property} \rangle \ U[\langle \text{natural-number} \rangle, \langle \text{natural-number} \rangle] \langle \text{logic-property} \rangle \\
| F[\langle \text{natural-number} \rangle, \langle \text{natural-number} \rangle] \langle \text{logic-property} \rangle \\
| G[\langle \text{natural-number} \rangle, \langle \text{natural-number} \rangle] \langle \text{logic-property} \rangle \\
| X \langle \text{logic-property} \rangle \\
| X[\langle \text{natural-number} \rangle] \langle \text{logic-property} \rangle \\
| (\langle \text{logic-property} \rangle)
$$

$$
\langle \text{numeric-measure} \rangle ::= \langle \text{numeric-spatial-measure} \rangle \\
| \langle \text{real-number} \rangle \\
| \langle \text{numeric-state-variable} \rangle \\
| \langle \text{unary-numeric-measure} \rangle (\langle \text{numeric-measure} \rangle) \\
| \langle \text{binary-numeric-measure} \rangle (\langle \text{numeric-measure} \rangle, \langle \text{numeric-measure} \rangle)
$$

$$
\langle \text{numeric-spatial-measure} \rangle ::= \langle \text{unary-subset-measure} \rangle (\langle \text{subset} \rangle) \\
| \langle \text{binary-subset-measure} \rangle (\langle \text{subset} \rangle, \langle \text{spatial-measure} \rangle) \\
| \langle \text{ternary-subset-measure} \rangle (\langle \text{subset} \rangle, \langle \text{spatial-measure} \rangle, \langle \text{real-number} \rangle) \\
| \langle \text{quaternary-subset-measure} \rangle (\langle \text{subset} \rangle, \langle \text{spatial-measure} \rangle, \langle \text{subset} \rangle, \langle \text{spatial-measure} \rangle)
$$

$$
\langle \text{unary-subset-measure} \rangle ::= \text{count} \\
| \text{clusteredness} \\
| \text{density}
$$

$$
\langle \text{binary-subset-measure} \rangle ::= \text{avg} \\
| \text{geomean} \\
| \text{harmean} \\
| \text{kurt} \\
| \text{max} \\
| \text{median} \\
| \text{min} \\
| \text{mode} \\
| \text{product} \\
| \text{skew} \\
| \text{stdev}
$$
ternary-subset-measure ::= percentile
  | quartile

quaternary-subset-measure ::= covar

unary-numeric-measure ::= abs
  | ceil
  | floor
  | round
  | sign
  | sqrt
  | trunc

binary-numeric-measure ::= add
  | div
  | log
  | mod
  | multiply
  | power
  | subtract

subset ::= subset-specific
  | filter((subset-specific), (constraint))

subset-specific ::= regions
  | clusters

constraint ::= ¬ (constraint)
  | (constraint) ∧ (constraint)
  | (constraint) ∨ (constraint)
  | (constraint) ⇒ (constraint)
  | (constraint) ⇔ (constraint)
  | (spatial-measure) (comparator) (filter-numeric-measure)
  | ((constraint))

filter-numeric-measure ::= (numeric-measure)
  | (spatial-measure)
  | (unary-numeric-measure)((filter-numeric-measure))
  | (binary-numeric-measure)((filter-numeric-measure), (filter-numeric-measure))

spatial-measure ::= clusteredness
  | density
  | area
  | perimeter
  | distanceFromOrigin
  | angle
  | triangleMeasure
  | rectangleMeasure
\[
\begin{align*}
\text{circleMeasure} & \quad \text{centroidX} \\
\text{centroidY} & \\
\langle \text{real-number} \rangle & ::= \langle \text{unsigned-real-number} \rangle \\
& \quad | \langle \text{sign} \rangle \langle \text{unsigned-real-number} \rangle \\
\langle \text{unsigned-real-number} \rangle & ::= \langle \text{fractional-part} \rangle \\
& \quad | \langle \text{fractional-part} \rangle \langle \text{exponent-part} \rangle \\
\langle \text{fractional-part} \rangle & ::= \langle \text{digit-sequence} \rangle . \langle \text{digit-sequence} \rangle \\
& \quad | . \langle \text{digit-sequence} \rangle \\
& \quad | \langle \text{digit-sequence} \rangle . \\
\langle \text{digit-sequence} \rangle & ::= \langle \text{digit} \rangle \\
& \quad | \langle \text{digit} \rangle \langle \text{digit-sequence} \rangle \\
\langle \text{digit} \rangle & ::= 0 \\
& \quad | 1 \\
& \quad | 2 \\
& \quad | 3 \\
& \quad | 4 \\
& \quad | 5 \\
& \quad | 6 \\
& \quad | 7 \\
& \quad | 8 \\
& \quad | 9 \\
\langle \text{natural-number} \rangle & ::= \langle \text{digit-sequence} \rangle \\
& \quad | + \langle \text{digit-sequence} \rangle \\
\langle \text{exponent-part} \rangle & ::= e \langle \text{digit-sequence} \rangle \\
& \quad | E \langle \text{digit-sequence} \rangle \\
& \quad | e \langle \text{sign} \rangle \langle \text{digit-sequence} \rangle \\
& \quad | E \langle \text{sign} \rangle \langle \text{digit-sequence} \rangle \\
\langle \text{sign} \rangle & ::= + \\
& \quad | - \\
\langle \text{comparator} \rangle & ::= > \\
& \quad | >= \\
& \quad | < \\
& \quad | <= \\
\langle \text{numeric-state-variable} \rangle & ::= \langle \text{state-variable} \rangle \\
\langle \text{state-variable} \rangle & ::= \{ \langle \text{string} \rangle \} \\
\langle \text{string} \rangle & ::= \langle \text{character} \rangle | \langle \text{character} \rangle \langle \text{string} \rangle
\end{align*}
\]
(character) ::= based on the Unicode character set except “{” and “}”

## Appendix B  Numeric measures

Table 2: Name, description and semantics of unary numeric measures

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td>abs</td>
<td>Returns the absolute value of a number</td>
<td>$abs(n) =</td>
</tr>
<tr>
<td>ceil</td>
<td>Rounds the number upward, returning the smallest integral value that is not less than the number</td>
<td>$ceil(n) = \lceil n \rceil$</td>
</tr>
<tr>
<td>floor</td>
<td>Rounds the number downward, returning the largest integral value that is not greater than the number</td>
<td>$floor(n) = \lfloor n \rfloor$</td>
</tr>
<tr>
<td>round</td>
<td>Returns the integral value that is nearest to x, with halfway cases rounded away from zero</td>
<td>$round(n) = \begin{cases} \lfloor n + 0.5 \rfloor, &amp; \text{if } n \geq 0 \ \lceil n - 0.5 \rceil, &amp; \text{otherwise} \end{cases}$</td>
</tr>
<tr>
<td>sign</td>
<td>Returns the sign of a number</td>
<td>$sign(n) = \begin{cases} 1, &amp; \text{if } n &gt; 0 \ 0, &amp; \text{if } n = 0 \ -1, &amp; \text{otherwise} \end{cases}$</td>
</tr>
<tr>
<td>sqrt</td>
<td>Returns the square root of a number</td>
<td>$sqrt(n) = \sqrt{n}$</td>
</tr>
<tr>
<td>trunc</td>
<td>Rounds number toward zero, returning the nearest integral value that is not larger in magnitude than the number</td>
<td>$trunc(n) = sign(n) \lfloor</td>
</tr>
</tbody>
</table>
Table 3: Name, description and semantics of binary numeric measures

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td>add</td>
<td>Returns the sum of two numbers</td>
<td>add(n_1, n_2) = n_1 + n_2</td>
</tr>
<tr>
<td>div</td>
<td>Returns the integer part of a division</td>
<td>(\text{div}(n_1, n_2) = \lfloor n_1/n_2 \rfloor)</td>
</tr>
<tr>
<td>log</td>
<td>Returns the logarithm of a number in the given base</td>
<td>(\log_b(</td>
</tr>
<tr>
<td>mod</td>
<td>Returns the remainder of a division</td>
<td>(\text{mod}(n_1, n_2) = n_1 - (n_2 \times \text{div}(n_1, n_2)))</td>
</tr>
<tr>
<td>multiply</td>
<td>Return the multiplication of two numbers</td>
<td>(\text{multiply}(n_1, n_2) = n_1 \times n_2)</td>
</tr>
<tr>
<td>power</td>
<td>Returns the base raised at the power exponent</td>
<td>(\text{pow}(b, c) = b^c)</td>
</tr>
<tr>
<td>subtract</td>
<td>Returns the difference between two numbers</td>
<td>(\text{subtract}(n_1, n_2) = n_1 - n_2)</td>
</tr>
</tbody>
</table>

Appendix C  Subset measures

Table 4: Name, description and semantics of unary subset measures

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td>count</td>
<td>Returns the number of clusters/regions</td>
<td>(\text{count}(\text{subset}) =</td>
</tr>
<tr>
<td>clusteredness</td>
<td>Returns the clusteredness of a set of clusters/regions</td>
<td>As described in the main manuscript</td>
</tr>
<tr>
<td>density</td>
<td>Returns the density of a set of clusters/regions</td>
<td>As described in the main manuscript</td>
</tr>
</tbody>
</table>

Table 5: Name, description and semantics of binary subset measures

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td>avg</td>
<td>Returns the arithmetic mean considering the given subset and spatial measure (sm)</td>
<td>(\text{avg}(\text{subset}, \text{sm}) = \frac{1}{n} \sum_{i=1}^{n} \text{sm}(\text{subset}_i)), where (n =</td>
</tr>
</tbody>
</table>
geomean Returns the geometric mean considering the given subset and spatial measure \( sm \)

\[
\text{geomean}(\text{subset}, sm) = \left( \prod_{i=1}^{n} sm(\text{subset}_i) \right)^{\frac{1}{n}}, \quad \text{where} \quad n = |\text{subset}|
\]

harmean Returns the harmonic mean considering the given subset and spatial measure \( sm \)

\[
\text{harmean}(\text{subset}, sm) = \frac{n}{\sum_{i=1}^{n} \frac{1}{sm(\text{subset}_i)}}, \quad \text{where} \quad n = |\text{subset}|
\]

kurt Returns the kurtosis considering the given subset and spatial measure \( sm \)

\[
\text{kurt}(\text{subset}, sm) = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^{n} \left( \frac{sm(\text{subset}_i) - \text{avg}(\text{subset}, sm)}{\text{stdev}(\text{subset}, sm)} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}, \quad \text{where} \quad n = |\text{subset}| \quad \text{and} \quad n \geq 4
\]

max Returns the maximum considering the given subset and spatial measure \( sm \)

\[
\text{max}(\text{subset}, sm) = \max_{i=1}^{n} sm(\text{subset}_i), \quad \text{where} \quad n = |\text{subset}|
\]

median Returns the median considering the given subset and spatial measure \( sm \)

\[
\text{median}(\text{subset}, sm) = \text{middle value in the ordered list of spatial measures}
\]

min Returns the minimum considering the given subset and spatial measure \( sm \)

\[
\text{min}(\text{subset}, sm) = \min_{i=1}^{n} sm(\text{subset}_i), \quad \text{where} \quad n = |\text{subset}|
\]

mode Returns the mode considering the given subset and spatial measure \( sm \)

\[
\text{mode}(\text{subset}, sm) = \text{value which appears most often in the list of spatial measures}
\]

product Returns the product considering the given subset and spatial measure \( sm \)

\[
\text{product}(\text{subset}, sm) = \prod_{i=1}^{n} sm(\text{subset}_i), \quad \text{where} \quad n = |\text{subset}|
\]

skew Returns the skewness considering the given subset and spatial measure \( sm \)

\[
\text{skew}(\text{subset}, sm) = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left( \frac{sm(\text{subset}_i) - \text{avg}(\text{subset}, sm)}{\text{stdev}(\text{subset}, sm)} \right)^3, \quad \text{where} \quad n = |\text{subset}|
\]

stdev Returns the standard deviation considering the given subset and spatial measure \( sm \)

\[
\text{stdev}(\text{subset}, sm) = \sqrt{\frac{\sum_{i=1}^{n} (sm(\text{subset}_i) - \text{avg}(\text{subset}, sm))^2}{(n-1)}}, \quad \text{where} \quad n = |\text{subset}|
\]

sum Returns the sum considering the given subset and spatial measure \( sm \)

\[
\text{sum}(\text{subset}, sm) = \sum_{i=1}^{n} sm(\text{subset}_i), \quad \text{where} \quad n = |\text{subset}|
\]
\[
\text{var} \quad \text{Returns the variance considering the given } \text{subset} \text{ and spatial measure } sm \\
\quad \text{var}(\text{subset}, sm) = \frac{\sum_{i=1}^{n} (sm(\text{subset}_i) - \text{avg}(\text{subset}, sm))^2}{n - 1}, \text{ where } n = |\text{subset}|
\]
### Table 6: Name, description and semantics of ternary subset measures

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td>percentile</td>
<td>Returns the $pv$-th ($0 \leq pv \leq 100$) percentile considering the given subset and spatial measure $sm$</td>
<td>$percentile(subset, sm, pv) = i$-th spatial measure considering an ordered list of $N$ spatial measures, $N$ is the number of regions/clusters and $i = \left\lfloor \frac{pv}{100} \times N + \frac{1}{2} \right\rfloor$</td>
</tr>
<tr>
<td>quartile</td>
<td>Returns the $i$-th ($i \in {25, 50, 75}$) quartile considering the given subset and spatial measure $sm$</td>
<td>Let $v$ be the ordered list of spatial measures, $m = median(subset, sm)$, $L$ the sublist of values in $v$ smaller than $m$ and $U$ the sublist of values in $v$ greater than $m$. $quartile(subset, sm, i) = median$ of $L$ if $i = 25$, $m$ if $i = 50$, and $median$ of $U$ if $i = 75$.</td>
</tr>
</tbody>
</table>

### Table 7: Name, description and semantics of quaternary subset measures

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td>covar</td>
<td>Returns the covariance considering $subset_1$ and $subset_2$, and the spatial measures $sm_1$ and $sm_2$</td>
<td>$covar(subset_1, sm_1, subset_2, sm_2) = \frac{1}{N-1} \sum_{i=1}^{N} ((sm_1(subset_1[i]) - mean(subset_1, sm_1)) \ast (sm_2(subset_2[i]) - mean(subset_2, sm_2)))$, where $N = min(</td>
</tr>
</tbody>
</table>