Reasoning about Complex Networks: A Logic Programming Approach

Gerardo I. Simari
Department of Computer Science, University of Oxford, UK

Joint work with:
Paulo Shakarian and Devon Callahan
United States Military Academy at West Point, USA

29th International Conference on Logic Programming (ICLP 2013)
Istanbul, Turkey – August 26–29, 2013
Complex Networks

- **Multi-attribute** time-series (1952–2012) author/paper network; authors colored green and papers red.
- Data extracted from *Thomson-Reuters Web of Knowledge*. 
Introduction

• Reasoning about complex networks has many applications:
  – Adoption of commercial products
  – Spread of disease
  – Diffusion of ideas, etc.

• We identified seven criteria that formalisms for this kind of reasoning “should satisfy”, and propose:
  – a formalism that satisfies all criteria;
  – algorithms for finding minimal models; and
  – study the problem of deciding group membership in social networks.
Seven Desiderata (1)

1. **Multiple** labels and weights for nodes and edges:
   - Most existing formalisms assume a single type of node that may become “active” or “mutate”; and
   - Only one type of relationship between nodes.
   - In reality, this is an oversimplification: nodes can be complex, relationships can be strong/weak, etc.

2. **Explicit representation of time:**
   - Existing works usually assume static models or simple models of temporal decay.
   - We seek a rich model of temporal relationships between conditions in the network structure, state of cascades, etc.
Seven Desiderata (1)

1. Multiple labels and weights for nodes and edges:
   - Most existing formalisms assume a single type of node that may become “active” or “mutate”; and
   - Only one type of relationship between nodes.
   - In reality, this is an oversimplification: nodes can be complex, relationships can be strong/weak, etc.

2. Explicit representation of time:
   - Existing works usually assume static models or simple models of temporal decay.
   - We seek a rich model of temporal relationships between conditions in the network structure, state of cascades, etc.
Seven Desiderata (2)

3. Non-Markovian temporal relationships:
   - “Memoryless” mode of Markov processes is insufficient to model dependencies that span multiple time units.

4. Representation of uncertainty:
   - In practice, it is not always feasible to judge all attributes of all individuals.
   - In connection to point 7, management of uncertainty should not come at a high computational cost.

5. Competing cascades:
   - Real-world situations often present competing network processes, where the success of one depends on another’s failure.
Seven Desiderata (2)

3. Non-Markovian temporal relationships
   – “Memoryless” mode of Markov processes is insufficient to model dependencies that span multiple time units.

4. Representation of uncertainty:
   – In practice, it is not always feasible to judge all attributes of all individuals.
   – In connection to point 7, management of uncertainty should not come at a high computational cost.

5. Competing cascades:
   – Real-world situations often present competing network processes, where the success of one depends on another’s failure.
Seven Desiderata (2)

3. Non-Markovian temporal relationships
   - “Memoryless” mode of Markov processes is insufficient to model dependencies that span multiple time units.

4. Representation of uncertainty:
   - In practice, it is not always feasible to judge all attributes of all individuals.
   - In connection to point 7, management of uncertainty should not come at a high computational cost.

5. Competing cascades:
   - Real-world situations often present competing network processes, where the success of one depends on another’s failure.
Seven Desiderata (3)

6. **Non-monotonic** cascades:
   - A common assumption is to only allow the number of nodes attaining a certain property to increase.
   - Competing processes are not compatible with such an assumption.

7. **Tractability**:
   - Social networks of interest today often have millions/billions of nodes.
   - Any framework for dealing with reasoning problems in such networks must be reasonably tractable.
   - In general, we take this to mean (low-degree) PTIME.
Seven Desiderata (3)

6. Non-monotonic cascades:
   - A common assumption is to only allow the number of nodes attaining a certain property to increase.
   - Competing processes are not compatible with such an assumption.

7. Tractability:
   - Social networks of interest today often have millions/billions of nodes.
   - Any framework for dealing with reasoning problems in such networks must be reasonably tractable.
   - In general, we take this to mean (low-degree) PTIME.
The MANCaLog Formalism (1)

• Preliminaries:
  – Assume a directed graph (network) \( G = (V,E) \)
  – Set of labels \( L = L_f \cup L_{nf} \) (fluent and non-fluent);
    labels are assigned to either nodes or edges in the network.
  – Network components: the set of all nodes and edges.

• Network atoms:
  Given label \( L \) and interval \( bnd \in [0,1] \), \( \langle L, bnd \rangle \) is a network atom.

• A world \( w \) is a set of network atoms s.t. for each \( L \in L \) there is exactly one atom of the form \( \langle L, bnd \rangle \) in \( w \).
The MANCaLog Formalism (2)

- Network formulas are defined over network atoms with the standard connectives ($\land, \lor, \neg$);
- Satisfaction:
  - World $w$ satisfies atom $\langle L, bnd \rangle$ iff there exists $\langle L, bnd' \rangle \in w$ such that $bnd \subseteq bnd'$;
  - satisfaction of formulas is defined inductively as usual.
- Facts:
  - Time is considered as discrete points in range $[0, t_{max}]$.
  - Facts are of the form $(A, C):[t_1, t_2]$, where $A$ is an atom, $C$ is a network component, and $[t_1, t_2] \subseteq [0, t_{max}]$.
  - If $A$ is formed with a non-fluent label, we allow only $[0, t_{max}]$ as the temporal annotation.
The MANCaLog Formalism (3)

Rules in the language are of the form:

\[ L \xrightarrow{\Delta t} f, (g_{edge}, g_{node}, h)_{ifl} \]

where:

- \( L \) is a label;
- \( \Delta t \geq 0 \);
- \( f \) is a formula over non-fluent network atoms;
- \( g_{edge}, g_{node} \) are non-fluent formulas (formed over edge and node atoms, resp.);
- \( h \) is a conjunction of (possibly fluent) network atoms;
- \( ifl \) is an influence function
The MANCaLog Formalism (3)

Rules in the language are of the form:

\[ L \overset{\Delta t}{\leftarrow} f, (g_{edge}, g_{node}, h)_{ifl} \]

Intuitively, this rule is read as follows:

“The \( L \) label of nodes meeting criteria described by \( f \) are influenced (within \( \Delta t \) time steps) by a set of neighbors that meet criteria described by \( g_{edge}, g_{node}, \) and \( h \) to a degree determined by function \( ifl \).”
Influence Functions

• **Influence functions** are of the form $ifl: \mathbb{N} \times \mathbb{N} \rightarrow [0,1] \times [0,1]$

• Intuitively, $ifl$ takes the number of neighbors and the number of *qualifying* neighbors and returns a new interval for the weight of the label in question.

• Example influence functions:
  
  **Tipping:**
  
  $tip(x,y) = \begin{cases} 
  [1,1] & \text{if } x/y \geq 0.5 \\
  [0,1] & \text{otherwise}
  \end{cases}$

  **Soft tipping:**
  
  $st(x,y) = \begin{cases} 
  [0.7,1] & \text{if } x/y \geq 0.5 \\
  [0,1] & \text{otherwise}
  \end{cases}$
Semantics

- Network interpretations map network components to worlds.
- Interpretations map time points to network interpretations.
- Models: Interpretations that satisfy all rules in a program.
- We define a partial ordering over equivalence classes of interpretations, which allows to define minimal models.
- Minimal models allow us to answer consistency and entailment queries.
- We developed a fixpoint operator for computing minimal models in PTIME.
Conclusions

• We outlined a set of **desirable criteria** for a language to model diffusion processes in complex networks.

• Essentially, we want to better understand how information and other phenomena **diffuse** in a network while leveraging information **beyond** network **topology**.

• Proposed the MANCaLog language, designed with these criteria in mind.

• Ongoing work involves experiments with real-world **police data** to address the problem of **group membership** in social networks.
A Comparison of Models

<table>
<thead>
<tr>
<th>Criterion</th>
<th>MANCaLog</th>
<th>IC/LT</th>
<th>SNOP</th>
<th>CD</th>
<th>EGT/VM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Labels</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>2. Explicit Representation of Time</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>3. Non-Markovian Time</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>4. Uncertainty</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>5. Competing Processes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>6. Non-monotonic Processes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>7. Tractablity</td>
<td>PTIME</td>
<td>#P-hard</td>
<td>PTIME</td>
<td>PTIME</td>
<td>NP-hard</td>
</tr>
</tbody>
</table>

IC/LT: Independent Cascade / Linear Threshold (Kempe et al., 2003) [2];

SNOP: Social Network Optimization Problems (Shakarian et al., 2010) [3];

CD: Competitive Diffusion (Broecheler et al., 2010) [4];

EGT/VM: Evolutionary Graph Theory / Voter Model (Liberman et al., 2005 and Sood et al., 2008) [5,6].
References


