Analysis of Large-Scale Biological Networks with Constraint-Based Approaches over Static Models

Carito Guziolowski

Hamamatsu TIGA Center - BioQuant - Institute of Medical Biometry and Informatics, University Hospital Heidelberg, Germany

October 14, 2011
Systems Biology Paradigm

- **Question**
  - Cancer treatment
  - Fattening
  - Illness resistance
  - Skin aging
  - Wound healing

- **Model**
  - Regulatory information
  - Expression information
  - Network reconstruction
  - Data analysis
  - Modeling

- **Model**: network & experimental data
- **Multidisciplinary**: biologists, physicians, computer scientists, mathematicians, physicists
Modeling influence graphs (causal networks)

**Dynamic**
- Trajectories
- System evolution
- Formalisms:
  - Boolean networks (Kauffman 1969)
  - Quantitative and qualitative differential equations (QSim, BioSim, GNA)
  - Rule-based formalisms (Kappa, BIOCHAM)
  - Small-scale networks
  - Unknown parameters

**Static**
- Structural, functional analyses
- Local correlation check (Herrgard et al., 2003)
  - Large-scale transcriptional networks
  - Genome scale data: RNA
- Flux balance analysis & Boolean regulatory constraints (Covert et al., 2004)
  - Large-scale networks: transcriptional and metabolic networks
  - Genome scale data: RNA, metabolic constants
- Boolean and Fuzzy Logic (Klamt et al., 2006, Morris et al., 2011)
  - Middle-scale signaling networks
  - Perturbation experiments: (phospho)protein measures
**Sign Consistency Model**

**Large-scale analysis ↔ Causal graphs + logical rules + consistency**

(Siegel et al. 2006, Radulescu et al. 2006, Veber et al. 2005)

- Quantitative ⇒ Qualitative data
- Trajectories ⇒ Global variations between steady states
- Simulations ⇒ Efficient constraints resolution
- Boolean on/off states ⇒ Up/down shifts
- Boolean algebra ⇒ Consistency rule
- Computation: *Automatic reasoning, predictions, experimental design*
Reasoning over qualitative variations (idea)

- Inconsistency: nothing explains \( rpmC \) variation

- Deduction: no other choice for \( arcA \) and \( fnr \) variations

Intuitive for small cases
Reasoning over qualitative variations (idea)

*Inconsistency:* nothing explains \( rpmC \) variation

*Deduction:* no other choice for \( arcA \) and \( fnr \) variations

Intuitive for small cases
Reasoning over qualitative variations (idea)

**Inconsistency:** nothing explains $rpmC$ variation

**Deduction:** no other choice for $arcA$ and $fnr$ variations

Intuitive for small cases
Reasoning over qualitative variations (idea)

- **Inconsistency**: nothing explains \( rpmC \) variation

- **Deduction**: no other choice for \( arcA \) and \( fnr \) variations

Intuitive for small cases
Reasoning over qualitative variations (idea)

- **Inconsistency**: nothing explains \( rpmC \) variation

- **Deduction**: no other choice for \( arcA \) and \( fnr \) variations

Intuitive for small cases
Reasoning over qualitative variations (idea)

**Inconsistency:** nothing explains $rpmC$ variation

**Deduction:** no other choice for $arcA$ and $fnr$ variations

Intuitive for small cases
Reasoning over qualitative variations (formalization)

1. **Influence graph**
   - The source influences the production of the target.
   - $\{+, -, ?\}$

2. **Experimental Data**
   - $\{+, -\}$ global variations

3. **Reasoning rule** to relate network regulations with expression variations

![Influence graph diagram]

```
Influence graph

⇒ The source influences the production of the target.
⇒ \{+, −, ?\}

Experimental Data
⇒ \{+, −\} global variations

Reasoning rule to relate network regulations with expression variations
```
Consistency rule

“The variation of one molecule in the network must be explained by an influence received from at least one of its predecessors in the network” (Siegel et al., Biosystems 2006)

⇒ Generic rule = allowed and forbidden color patterns
Global Consistency

Consistency coloring model:

- Assign a consistent \{+,−\} “coloring”
- Predictions = Intersection of all consistent coloring models
- How is it approached?
  - Cross Graph coloring ⇒ delicate algorithmic
  - Tick Solving qualitative constraints

Justification:

- Imposing constraints allows extending the consistency rule
- Diagnosis visualization
Qualitative constraints

Sign algebra

<table>
<thead>
<tr>
<th>⊕</th>
<th>+</th>
<th>−</th>
<th>?</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>−</td>
<td>?</td>
<td>−</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

(close to OR)

<table>
<thead>
<tr>
<th>⊗</th>
<th>+</th>
<th>−</th>
<th>?</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>−</td>
<td>?</td>
</tr>
<tr>
<td>−</td>
<td>−</td>
<td>+</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>≈</th>
<th>+</th>
<th>−</th>
<th>?</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>T</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>−</td>
<td>F</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>?</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
</tbody>
</table>

Influence graph ⇒ Qualitative system of constraints

Experimental data ⇒ Fix variables of the system

Solution = {+, −} valuation of variables

Intersection of solutions = Predictions
Qualitative constraints

Sign algebra

<table>
<thead>
<tr>
<th>⊕</th>
<th>+</th>
<th>-</th>
<th>?</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>-</td>
<td>?</td>
<td>-</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

(close to OR)

<table>
<thead>
<tr>
<th>⊗</th>
<th>+</th>
<th>-</th>
<th>?</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>-</td>
<td>?</td>
</tr>
<tr>
<td>-</td>
<td>?</td>
<td>-</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>≈</th>
<th>+</th>
<th>-</th>
<th>?</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>T</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>-</td>
<td>F</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>?</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
</tbody>
</table>

(+ \not\approx -)

- Influence graph \Rightarrow Qualitative system of constraints
- Experimental data \Rightarrow Fix variables of the system

Qualitative system of constraints

Solution = \{+, -\} valuation of variables

Intersection of solutions = Predictions
Qualitative constraints

Sign algebra

<table>
<thead>
<tr>
<th>⊕</th>
<th>+</th>
<th>−</th>
<th>?</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>−</td>
<td>?</td>
<td>−</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

(close to OR)

<table>
<thead>
<tr>
<th>⊗</th>
<th>+</th>
<th>−</th>
<th>?</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>−</td>
<td>?</td>
</tr>
<tr>
<td>−</td>
<td>−</td>
<td>+</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>≈</th>
<th>+</th>
<th>−</th>
<th>?</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>T</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>−</td>
<td>F</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>?</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
</tbody>
</table>

(+ ≈ −)

- Influence graph ⇒ Qualitative system of constraints
- Experimental data ⇒ Fix variables of the system

Solution = {+, −} valuation of variables
- Intersection of solutions = Predictions
Qualitative constraints

Sign algebra

<table>
<thead>
<tr>
<th>⊕</th>
<th>+</th>
<th>−</th>
<th>?</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>−</td>
<td>?</td>
<td>−</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

(close to OR)

<table>
<thead>
<tr>
<th>⊗</th>
<th>+</th>
<th>−</th>
<th>?</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>−</td>
<td>?</td>
</tr>
<tr>
<td>−</td>
<td>−</td>
<td>+</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>≈</th>
<th>+</th>
<th>−</th>
<th>?</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>T</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>−</td>
<td>F</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>?</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
</tbody>
</table>

(+ $\not\approx$ −)

- **Influence graph** $\Rightarrow$ Qualitative system of constraints
- **Experimental data** $\Rightarrow$ Fix variables of the system

Solution $= \{+, −\}$ valuation of variables
Intersection of solutions $= \text{Predictions}$
Computational implementations

Bioquali Python Library
- Efficiently represents all solutions in memory (Le Borgne, 2009)
- **Strategy**: exploit the few variables shared among constraints
- Dependency (bipartite) graph
- Covering tree: optimal order to solve the constraints

BioASP
- Answer Set Programming (ASP)
- Shape of programming languages but using SAT algorithms
- Influence graph, experiments, constraints coded into a program of rules
- **Strategy**: find the *answer sets* of the program (*clasp*; Gebser et al. LPNMR'07)

Availability:
- **BioASP**: [http://www.cs.uni-potsdam.de/bioasp/](http://www.cs.uni-potsdam.de/bioasp/)
Bioinformatic tools

- Cytoscape Plugin

- Bioquali Web service
  [http://mobyle.genouest.org/cgi-bin/Mobyle/portal.py?form=bioquali](http://mobyle.genouest.org/cgi-bin/Mobyle/portal.py?form=bioquali)

Guziolowski et al., BMC Genomics, 2009
Proven Biological Applications

1. Studying the consistency of the *E.coli* transcriptional network wrt small dataset of observations (Bioquali) (Guziolowski *et al.*, JBPC, 2006; Guziolowski *et al.*, CIBB’08)

2. Detecting the origin of a phenotype in the EWS-FLI1 human oncogene signaling network (Bioquali) (Guziolowski *et al.*, IEEE-TCCB, 2010)

3. Prediction under inconsistency, minimal corrections on network and/or experiments to make them consistent (ASP) (Gebser *et al.*, KR’10)

4. Inferring the TF roles in the *S. cerevisiae* transcriptional network (Bioquali + ASP) (Veber *et al.*, BMC Bioinformatics, 2008)

Impact of real biological data over informatic implementations
MEDSYS Project
Systems Biology for Chronical Wounds
A close look to skin wound healing

System characteristics:

- Different cell types: keratinocytes, fibroblasts
- Many growth factors (GFs)
- Strong cell communication: sending and receiving GFs signals
- Multi-scale system: tissue, intercellular, intracellular
- 3 main cell states: migration, proliferation, differentiation

Whal R&D Systems 2002
1. RNA measurements (many microarray experiments)
   - Time series data from key GFs stimulation (e.g. HGF)

2. 2D keratinocyte cultures
   - Inhibition, over-expression, knockout experiments
   - Methods to quantify keratinocytes migration, proliferation

3. 3D tissue model
   - Staining of (some) relevant proteins
   - Imaging $\rightarrow$ protein localization and quantification in a tissue context

How to use this data in a Systems Biology context?
Building intracellular influence graphs?

Requirements

- Use genome scale data issued from GF stimulation experiments
- Link GF stimulation with cellular process
- Not a nice cartoon → Computable object (influence graph)

Querying database information

- Explain® Systems
- Ingenuity® Systems
- Pathways Commons (Cerami et al. 2011) - BioPAX/Cytoscape

⇒ Not influence (logical) graphs

- BiNoM (Zynovyev et al. 2007) - BioPAX
⇒ Missing transcriptional/signaling levels, cellular processes connections

Inference from transcriptome

- Correlation clusters + TFs (Faith et al. 2007)
- Mathematical models: Bayesian networks (Missal et al. 2006), ODEs (Bansal et al. 2006)
- Machine learning: neural networks (Busch et al. 2008)

⇒ Mix direct with indirect links
⇒ Only transcriptional links
⇒ Require numerous data points
Intracellular view: automatic generation of RSTC networks

4 molecular scales
1. Receptors
2. Signaling events
3. Transcription
4. Cellular processes

Are computable objects
- Unique node/edge identifiers
- Preserve the (qualitative) logic of the system: activations, inhibitions
- Can be queried/analyzed by:
  - Graph algorithms
  - Bioquali, BioASP, Boolean Networks
Automatic generation of RSTC networks (pipeline)

- PID
- Master network
- Filter:
  - Small molecules
  - Proteins which RNA is absent
  - Complexes with at least one member with RNA absent
- Active network
- Subgraph extraction by linking: GF-receptor + 2-fold genes + cell procs
- 2-fold up/down regulations

RSTC network
Automatic generation of RSTC networks (pipeline)

PID
- PID.xml
- Java program

Master network

Filter:
- Small molecules
- Proteins which RNA is absent
- Complexes with at least one member with RNA absent

Active network

Barcode

Cytoscape format
- Nodes: proteins, complexes
- Edges: activations, inhibitions, state-transitions

RSTC network
- Perl, Python programs

Subgraph extraction by linking: GF-receptor + 2-fold genes + cell procs

2-fold up/down regulations

Guziolowski et al, submitted
HGF RSTC network
reinterpret state-transitions
HGF RSTC network
reinterpret state-transitions

RhoA, Arf6, Rac1, E-cad, PI3K

Up-regulated \Rightarrow CYR61, EGR1, MKP1, ET1, FOS, uPAR, IL8, LAMA3, COX2

migration, proliferation
Experimental validation of network predictions

Network predictions (extract)

- The network topology obtained was consistent with the 2-fold gene expression upon 1 hour of HGF stimulation (Bioquali).
- Validate some predictions experimentally (ongoing work).
- Example: HGF $\rightarrow$ RhoA $\rightarrow$ Erk12

Induction with 10ng/ml HGF
Conclusion

- **Sign Consistency Modeling**: Causal modeling + constraint resolutions + consistency + boolean rules.
- Main interest: reason over large-scale/complex biological networks.
- Biological applications: network curation, experimental design, in-silico analysis of the system combinatorics.
- Systems Biology for **wound healing**:
  - Propose an RSTC network for HGF signaling: C-met receptor → signaling cascades → transcription → cell-migration, -proliferation.
  - RSTC networks are a computable objects
  - Bioquali inference: active proteins that explain the RNA data.
Current projects

1. **Automatic extraction** of causal logical networks
   - RSTC networks ⋆, □
   - Rules to add logic to reaction networks □
   - Cytoscape plugin □

2. **Experimental validation** of the network structure and Bioquali outputs ⋆, ø
   - Keratinocytes migration: HGF signaling network
   - Keratinocytes differentiation: EGF + E-cadherin signals

3. Proposing a **transcriptional network** for Acidithiobacillus ferrooxidans ¹ ⋆

4. **Optimizing** the structure of signaling networks to fit protein data (DREAM) ² ⋆

---

Legend: ⋆: modeling, □: (bio)informatic, ø: biology

---

¹ Andres Aravena, Alejandro Maass @ CMM, Chile - Anne Siegel @ Rennes
² Sven Thiele, Torsten Schaub @ Potsdam University - Julio Saez-Rodriguez @ EBI
Acknowledgments

- **TIGA Center @ Bioquant Heidelberg** Niels Grabe
- **DKFZ, Heidelberg**: Christian Schuster
- **University Hospital Heidelberg**: Kai Breuhahn
- **Symbiose Project, IRISA-INRIA Rennes, France**: Anne Siegel, Pierre Blavy
- **University of Potsdam, Germany**: Torsten Schaub, Sven Thiele
- **University of Chile, Santiago, Chile**: Alejandro Maass, Andres Aravena
Thank you!
The resolution of a qualitative system of constraints is a NP-complete problem.