

# Modelagem e Inferência de Redes Gênicas Probabilísticas (projeto de doutorado)

**Aluno:** Marcelo da Silva Reis<sup>1</sup>  
**Orientador:** Junior Barrera<sup>1</sup>

<sup>1</sup>Instituto de Matemática e Estatística, Universidade de São Paulo

Seminários de Fundamentos Metodológicos para I.A.

29 de setembro de 2008



# Agenda

## Introdução

Motivação: modelagem de Redes Gênicas Regulatórias

Soluções (modelos) existentes para o problema

Redes Gênicas Probabilísticas (PGNs)

Subredes: uma generalização do problema

## Objetivos do Projeto

## Metodologia

Modelagem da Rede Gênica Probabilística (PGN)

Inferência (estimação) de PGNs

## Trabalhos Correlatos

## Resultados e Contribuições

## Cronograma

Atividades e Publicações esperadas

## Bibliografia



# Agenda

## Introdução

Motivação: modelagem de Redes Gênicas Regulatórias

Soluções (modelos) existentes para o problema

Redes Gênicas Probabilísticas (PGNs)

Subredes: uma generalização do problema

## Objetivos do Projeto

## Metodologia

Modelagem da Rede Gênica Probabilística (PGN)

Inferência (estimação) de PGNs

## Trabalhos Correlatos

## Resultados e Contribuições

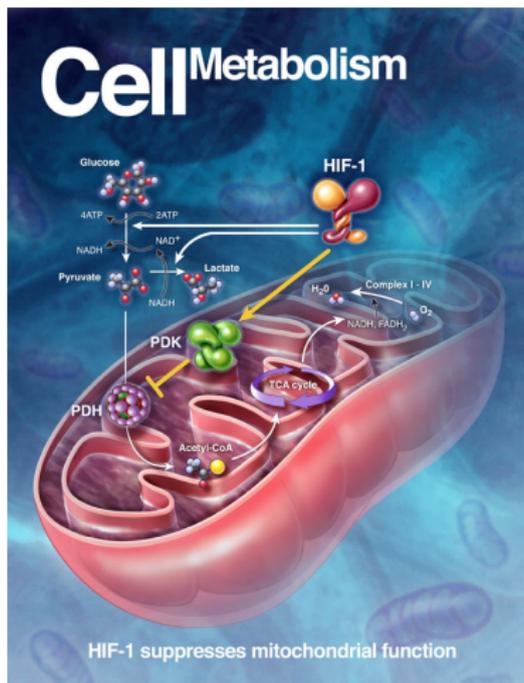
## Cronograma

Atividades e Publicações esperadas

## Bibliografia



# A célula: entendendo o seu funcionamento

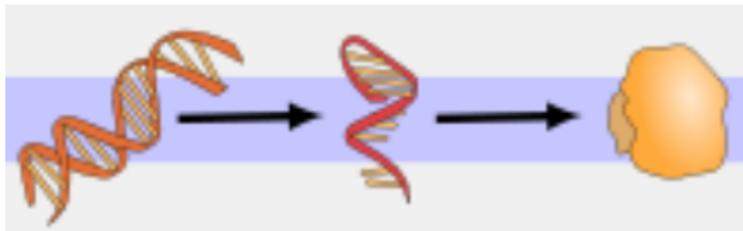


Metabolismo celular: um conjunto de complexas reações químicas.



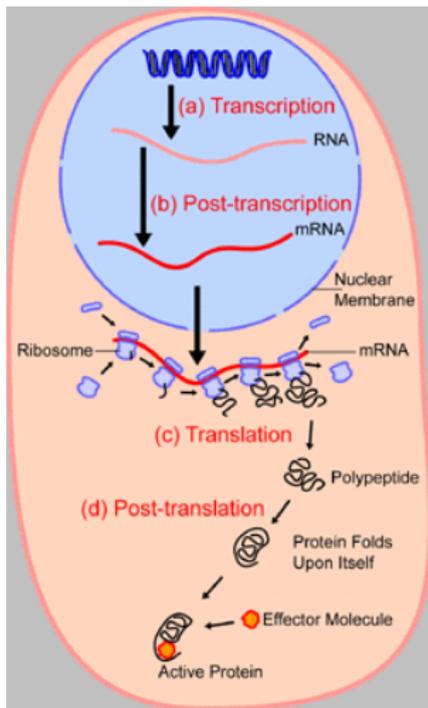
# Genes e Dogma Central

- ▶ O “controle” dessas reações é feito através dos genes.
- ▶ Genes são codificados através de DNA.
- ▶ Proteínas (e.g. enzimas) são produzidas através da expressão de seus genes correspondentes (Dogma Central).



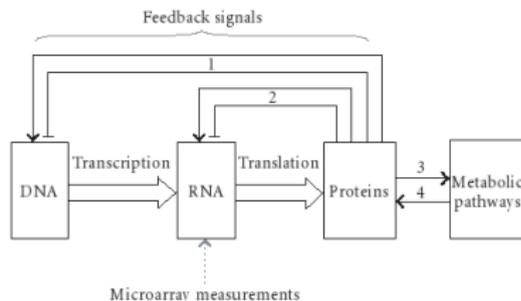
# Genes e Dogma Central (2)

Síntese de proteínas na célula:



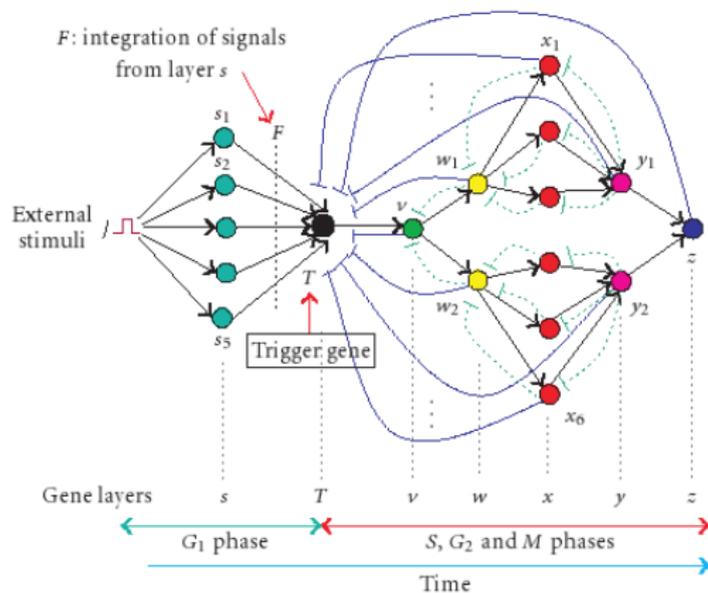
# Redes Gênicas Regulatórias

- ▶ A expressão dos genes não é constante.
- ▶ Ela pode ser ativada (ou inibida) devido à expressão de outros genes, através de sinais de *feedback*.
- ▶ Consequência direta do item acima: existe uma relação de dependência entre os genes.



## Redes Gênicas Regulatórias (2)

Um exemplo de modelo de rede gênica regulatória:



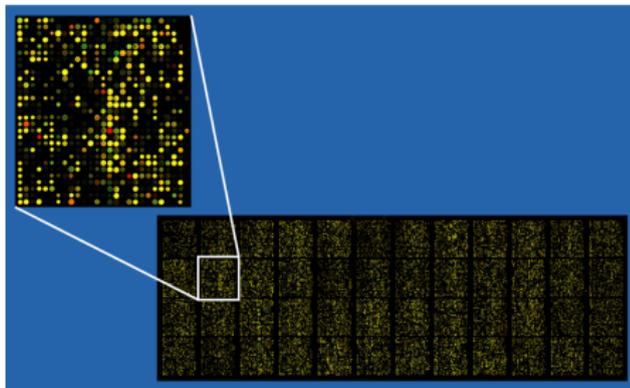
→ Forward signal

— Feedback to  $T$



# Microarranjos e Séries Temporais

- ▶ Microarranjos são experimentos biológicos onde os níveis de expressão de diversos genes são medidos simultaneamente.

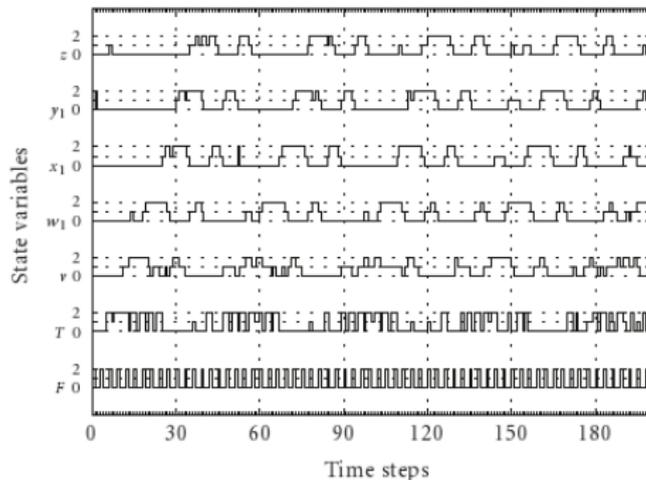


- ▶ Uma sequência sucessiva de microarranjos, com um certo intervalo de tempo entre eles, é chamada de *série temporal*.



## Microarranjos e Séries Temporais (2)

As séries temporais produzidas podem ser normalizadas e discretizadas:



# O Problema da Modelagem de Redes Gênicas

**O problema:** dada uma série temporal dos níveis de expressão de uma coleção de genes, determinar um modelo matemático que melhor explique como esses mesmos genes se relacionam entre si.



# Soluções (modelos) existentes para o problema

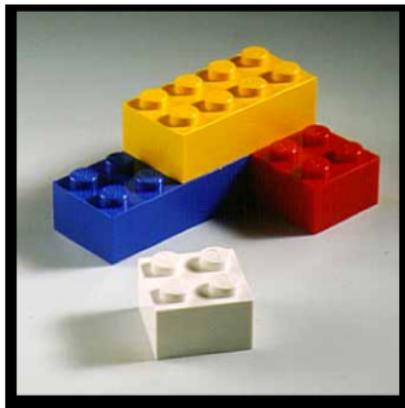
Existem na literatura diversas técnicas para abordar o problema.  
Algumas delas:

- ▶ Equações Diferenciais Ordinárias (EDOs);
- ▶ Redes Bayesianas;
- ▶ Álgebra de Processos;
- ▶ Redes Booleanas (determinísticas ou probabilísticas);
- ▶ Cadeias de Markov.
  - ▶ **Rede Gênica Probabilística (PGN, da sigla em inglês).**



# Subredes

- ▶ Uma maneira de se visualizar as relações entre os genes de um genoma seria tomá-los como uma coleção de redes gênicas regulatórias, acopladas entre si.



- ▶ *Conjectura*: redes regulatórias funcionais são “conservadas” em diferentes espécies, que herdaram a dada rede de um ancestral comum (Olson, 2006).



# O Novo Problema

**O problema, reformulado:** dada uma série temporal dos níveis de expressão de uma coleção de genes, determinar um modelo matemático que melhor explique como esses mesmos genes se relacionam entre si, tanto localmente (subredes funcionais) quanto globalmente.



# Agenda

## Introdução

Motivação: modelagem de Redes Gênicas Regulatórias

Soluções (modelos) existentes para o problema

Redes Gênicas Probabilísticas (PGNs)

Subredes: uma generalização do problema

## Objetivos do Projeto

### Metodologia

Modelagem da Rede Gênica Probabilística (PGN)

Inferência (estimação) de PGNs

### Trabalhos Correlatos

### Resultados e Contribuições

### Cronograma

Atividades e Publicações esperadas

### Bibliografia



# Objetivos do Projeto

- ▶ Desenvolver um novo modelo de Rede Gênica Probabilística, onde uma dada rede é composta de diversas PGNs que se relacionam entre si.
- ▶ Validar esse novo modelo, através de estimação de PGNs utilizando dados simulados e reais.
- ▶ Estudar a intervenção e controle desse novo modelo de PGN, desenvolvendo os algoritmos necessários. (*desejável*)
- ▶ Outras idéias... (*se der tempo!*)



# Agenda

## Introdução

Motivação: modelagem de Redes Gênicas Regulatórias

Soluções (modelos) existentes para o problema

Redes Gênicas Probabilísticas (PGNs)

Subredes: uma generalização do problema

## Objetivos do Projeto

## Metodologia

Modelagem da Rede Gênica Probabilística (PGN)

Inferência (estimação) de PGNs

## Trabalhos Correlatos

## Resultados e Contribuições

## Cronograma

Atividades e Publicações esperadas

## Bibliografia



# Modelagem da Rede Gênica Probabilística (PGN)

Redes Gênicas Probabilísticas são um tipo especial de Cadeias de Markov, com axiomas adicionais (Barrera, 2004); algumas características:

1. Cada gene tem 3 estados (inibido, não alterado, ativado);
2. Para  $n$  genes, o vetor de estados tem  $3^n$  possibilidades;
3. A matriz de transição é homogênea e quase determinística;
4. Todos os estados da rede são alcançáveis.

Cada subrede seria modelada como uma PGN. O acoplamento entre essas subredes seria implementado através de técnicas existentes na Teoria de Sistemas.



# Inferência (estimação) de PGNs

Existem diversos métodos para estimação de uma PGN. Os mais utilizados são:

- ▶ Informação Mútua Média (Barrera, 2004).
- ▶ Coeficiente de Determinação – CoD (Dougherty, 2000).

Atualmente é desenvolvido um trabalho de comparação entre esses dois métodos.

A estimação é feita utilizando-se tanto dados simulados quanto séries temporais reais (e.g. dados do patógeno causador da Malária).



# Ferramentas e Implementações

- ▶ A dinâmica de PGNs pode ser avaliada utilizando o simulador SGEN (Armelin e colegas, 2001).
- ▶ O SGEN também é utilizado para a produção de séries temporais simuladas.
- ▶ Métodos de estimação e outros algoritmos normalmente são implementados em C/C++ e Perl.



# Agenda

## Introdução

Motivação: modelagem de Redes Gênicas Regulatórias

Soluções (modelos) existentes para o problema

Redes Gênicas Probabilísticas (PGNs)

Subredes: uma generalização do problema

## Objetivos do Projeto

## Metodologia

Modelagem da Rede Gênica Probabilística (PGN)

Inferência (estimação) de PGNs

## Trabalhos Correlatos

## Resultados e Contribuições

## Cronograma

Atividades e Publicações esperadas

## Bibliografia



## Do ponto de vista matemático / computacional:

Dougherty, E. R. e Schmulevich, I. **Mappings Between Probabilistic Boolean Networks**. *Signal Processing* (2003) 83(4):799–809.

- ▶ Nesse trabalho (e também em posteriores) são exploradas sub-redes, porém através de projeções e de reduções da rede original.
- ▶ Maior preocupação com questões computacionais.



## Trabalhos Correlatos (2)

### Do ponto de vista biológico:

Olson, E.N. **Gene Regulatory Networks in the Evolution and Development of the Heart.** Science (2006) 313:1922–1927.

- ▶ Nesse trabalho o acoplamento entre sub-redes é estudado, porém do ponto de vista biológico.
- ▶ É feita a análise dos fatores de transcrição envolvidos no processo.



# Agenda

## Introdução

Motivação: modelagem de Redes Gênicas Regulatórias

Soluções (modelos) existentes para o problema

Redes Gênicas Probabilísticas (PGNs)

Subredes: uma generalização do problema

## Objetivos do Projeto

## Metodologia

Modelagem da Rede Gênica Probabilística (PGN)

Inferência (estimação) de PGNs

## Trabalhos Correlatos

## Resultados e Contribuições

## Cronograma

Atividades e Publicações esperadas

## Bibliografia



# Resultados e Contribuições

## Contribuições Científicas:

- ▶ A caracterização de uma generalização de uma família de Cadeias de Markov que melhor explique a integração de subredes dentro de um sistema estocástico complexo.
- ▶ Um método eficiente de inferência de sistemas desse tipo, a partir de uma amostra.

## “Efeito colateral benigno”:

- ▶ Um modelo estocástico que pode ser utilizado para o estudo de redes gênicas regulatórias, a partir de séries temporais de diferentes espécies.



# Agenda

## Introdução

Motivação: modelagem de Redes Gênicas Regulatórias

Soluções (modelos) existentes para o problema

Redes Gênicas Probabilísticas (PGNs)

Subredes: uma generalização do problema

## Objetivos do Projeto

## Metodologia

Modelagem da Rede Gênica Probabilística (PGN)

Inferência (estimação) de PGNs

## Trabalhos Correlatos

## Resultados e Contribuições

## Cronograma

Atividades e Publicações esperadas

## Bibliografia



# Cronograma

Período	Atividade	Resultados
Mar-08 – Jul-08	Estudos da bibliografia e da teoria de inferência de sistemas dinâmicos.	Know-how.
Jul-08 – Fev-09	Estudo comparativo entre métodos de estimação.	Resultados publicáveis.
Fev-09 – Mar-09	Realização dos exames obrigatórios (AA e Sistemas).	
Mar-09 – Jun-09	Estudos da bibliografia e da teoria de sistemas dinâmicos (“temporada na Poli”).	Know-how, talvez resultados publicáveis



## Cronograma (2)

Período	Atividade	Resultados
Mai-09 – Dez-09	Modelagem do novo paradigma de PGN.	Resultados publicáveis.
Ago-09 – Jan-10	Inferência de redes seguindo o novo modelo.	Resultados publicáveis.
Jan-10 ou Fev-10	Exame de Qualificação.	
Mar-10 – Dez-10	Melhoramentos no modelo e no método de estimação do mesmo.	Possivelmente resultados publicáveis.
Mar-10 – Dez-10	Estudos e eventuais implementações de métodos de intervenção e controle.	Resultados publicáveis.
Jan-11 ou Fev-11	Defesa do Doutorado.	



# Periódicos e Congressos

## Congressos:

- ▶ Congresso da ISMB (internacional).
- ▶ Recomb (internacional).
- ▶ X-Meeting (nacional).

## Periódicos:

- ▶ Nature? Science? (resultados biológicos de  **muito** impacto :-)
- ▶ PNAS, Nucleic Acids Research. (resultados biológicos significativos).
- ▶ Bioinformatics, Public Library of Science (PLoS).
- ▶ GMB, Anais do X-Meeting (nacionais).



# Agenda

## Introdução

Motivação: modelagem de Redes Gênicas Regulatórias

Soluções (modelos) existentes para o problema

Redes Gênicas Probabilísticas (PGNs)

Subredes: uma generalização do problema

## Objetivos do Projeto

## Metodologia

Modelagem da Rede Gênica Probabilística (PGN)

Inferência (estimação) de PGNs

## Trabalhos Correlatos

## Resultados e Contribuições

## Cronograma

Atividades e Publicações esperadas

## Bibliografia



# Bibliografia

- ▶ Trepode, N.W. and Armelin, H.A. and Bittner, M. and Barrera, J. and Gubitoso, M.D. and Hashimoto, R.F.  
**A Robust Structural PGN Model for Control of Cell-Cycle Progression Stabilized by Negative Feedbacks**  
EURASIP (2007) 1–11.
- ▶ Olson, E.N.  
**Gene Regulatory Networks in the Evolution and Development of the Heart**  
Science (2006) 313:1922–1927.
- ▶ Barrera, J. and Cesar Jr, R.M. and Merino, E.F. and Pereira, C.A.B. and Vencio, R.Z. and Leonardi, F.G. and Yamamoto, M.M. and Portillo, H.A.  
**A new annotation tool for malaria based on inference of probabilistic genetic networks**  
CAMDA (2004) 36–40.
- ▶ Datta, A. and Choudhary A. and Bittner M.L. and Dougherty, E.R.  
**External Control in Markovian Genetic Regulatory Networks**  
Machine Learning (2003) 52:169–191.



## Bibliografia (2)

- ▶ Armelin, H.A. and Barrera, J. and Dougherty, E.R. and Gubitoso, M.D. and Ferreira, J.E. and Hirata, N.S.T. and Neves, E.J.  
**A Simulator for Gene Expression Networks**  
SPIE Microarrays (2001) 4266:248–259.
- ▶ Dougherty, E.R. and Kim, S. and Chen, Y.  
**Coefficient of determination in nonlinear signal processing**  
Signal Processing (2000) 80:2219–2235.
- ▶ Shmulevich, I. and Dougherty, E.R. and Kim, S. and Zhang, W.  
**Probabilistic Boolean Networks: A rulebased uncertainty model for gene regulatory networks**  
Bioinformatics (2000) 18:261–274.
- ▶ Shmulevich, I. and Dougherty, E.R. and Zhang, W.  
**Gene perturbation and intervention in probabilistic boolean networks**  
Bioinformatics (2000) 18:1319–1331.



# Perguntas?



Obrigado!

THANK  
YOU

