



UNIVERSIDADE ESTADUAL DO CEARÁ  
CENTRO DE CIÊNCIAS E TECNOLOGIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA  
DA COMPUTAÇÃO

**FABIANO CARNEIRO RIBEIRO**

**ESTIMAÇÃO DE MATRIZES DE TRÁFEGO ORIGEM-DESTINO  
UTILIZANDO ALGORITMO GENÉTICO**

FORTALEZA – CEARÁ  
AGOSTO/2009

FABIANO CARNEIRO RIBEIRO

ESTIMAÇÃO DE MATRIZES DE TRÁFEGO ORIGEM-DESTINO UTILIZANDO  
ALGORITMO GENÉTICO

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Estadual do Ceará - UECE como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

Área de Concentração: *Sistema de Computação*

Orientador: Prof. Dr. Jorge Luiz de Castro e Silva

FORTALEZA – CE

AGOSTO/2009

R484e

Ribeiro, Fabiano Carneiro

Estimação de Matrizes de Tráfego Origem-Destino Utilizando Algoritmo Genético. Fortaleza, 2009.

79p.;il.

Orientador: Prof. Dr. Jorge Luiz de Castro e Silva

Dissertação (Mestrado Acadêmico em Ciência da Computação) – Universidade Estadual do Ceará, Centro de Ciências e Tecnologia.

1. Estimação. 2. Matriz de tráfego. 3. Algoritmo Genético.

CDD:001.6

FABIANO CARNEIRO RIBEIRO

ESTIMAÇÃO DE MATRIZES DE TRÁFEGO ORIGEM-DESTINO UTILIZANDO  
ALGORITMO GENÉTICO

Dissertação submetida à Coordenação do curso de Mestrado Acadêmico em Ciência da Computação da Universidade Estadual do Ceará, como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação.

Aprovada em 28 de agosto de 2009

BANCA EXAMINADORA

---

Prof. Dr. Jorge Luiz de Castro e Silva (Orientador)  
Universidade Estadual do Ceará – UECE

---

Prof. Dr. Gustavo Augusto Lima de Campos  
Universidade Estadual do Ceará – UECE

---

Prof. Dr. Jerffeson Teixeira de Souza  
Universidade Estadual do Ceará – UECE

---

Prof. Dr. Raimir Holanda Filho  
Universidade de Fortaleza – UNIFOR

À Ágatha, minha filha amada e a minha esposa Luciana,  
com gratidão pelo apoio.

# Agradecimentos

Ao Senhor, pela luz de meu caminho.

A minha esposa Luciana e minha filha Ágatha, que estavam sempre a me esperar durante esses dois anos de afastamento.

Aos meus pais, irmãos, cunhados e minhas amáveis sobrinhas, sempre presentes e atuantes, com união e conforto.

Ao meu orientador Prof. Jorge Luiz, pela confiança, apoio, ensinamentos permanentes e pela paciência única.

Aos professores Gustavo, Celestino, Jerffeson, André, pelo carinho e ensinamentos de sala de aula e de todos os dias.

Aos velhos e novos amigos conquistados durante o período inicial: Alex, Carlos, Márcia, Renato, Robson, Vigno e Viviane.

Aos amigos posteriores e companheiros de tantas horas: Daladier, Gilzamir, Walisson.

Aos amigos funcionários, sempre atenciosos e prestativos: Marcos e Vagner.

Aos colegas de orientador pelo carinho e gentileza: Cecília, Silas e Arnoldo.

Aos amigos do dia-a-dia por sempre nos apoiar.

Ao Instituto Centro de Ensino Tecnológico pela liberação e investimentos feitos.

A FUNCAP pelo apoio financeiro.

*A minha tia Guiguiá (in memoria)  
pela sua presença em espírito!*

# Resumo

Matrizes de tráfego (MT) contêm informações a respeito do fluxo de tráfego entre todos os possíveis pares de nós em uma rede. É uma preocupação determinar qual a melhor maneira para transmissão de dados que atenda mais prontamente requisitos de desempenho nas aplicações, o que é visto como um grande desafio aos engenheiros de tráfego. Nas últimas décadas, tem-se observado uma considerável parcela de pesquisas em estimação de matriz de tráfego no contexto da Internet, redes de transporte e redes de telecomunicações. Mas essa medição de dados sobre fluxos em redes de comunicação pode gerar um alto custo de coleta e computacional. Em função disso, torna-se necessário o uso de técnicas de estimação baseadas em informações parciais.

Os métodos de estimação existentes não resolvem satisfatoriamente o problema da estimação de MT por se tratar de um problema complexo. Esta dissertação considera deficiências de vários desses métodos e propõe uma nova abordagem utilizando algoritmos genéticos que combinam os conceitos de adaptação seletiva e sobrevivência dos indivíduos mais capazes. Foram realizados experimentos com dados sintéticos e dados reais. Outra contribuição esperada da dissertação é a criação de um método de inicialização para geração de uma população inicial que se mostre eficiente na redução do tempo computacional e do erro médio de estimação.

**Palavras-chave:** Estimação. Matriz de tráfego. Algoritmo Genético.

# Abstract

Traffic Matrix (TM) has information about the traffic flow between all possible pairs of nodes in a network. It's a concern to determine how best to transmit data more readily meet the performance requirements in applications, which is seen as challenge to traffic engineers. In recent decades, there has been a considerable portion of research on traffic matrix estimation in the context of Internet, transmission and telecommunications networks. But this measurement data on flows in networks can generate a high cost of collecting and computing. As a result, it becomes necessary to use estimation techniques based on partial information.

The existing estimation methods don't solve satisfactorily the problem of estimating the MT because it is a complex problem. This dissertation considers shortcomings of several of these methods and proposes a new approach using genetic algorithms that combine the concepts of selective adaptation and survival of the most capable individuals. Experiments were carried out with synthetic and real data. Another contribution of this work is expected to create a boot method for generating an initial population that proves effective in reducing the computational time and average error of estimation.

**Key-words:** Estimation. Traffic Matrix. Genetic Algorithm.

## Lista de Tabelas

Tabela 6.1 – Estimação para topologia de quatro nós e sete enlaces – MT Constante .....	52
Tabela 6.2 – Estatísticos da estimação para topologia de quatro nós e sete enlaces – MT Constante .....	53
Tabela 6.3 – Estimação para topologia de quatro nós e sete enlaces – MT Poisson.....	55
Tabela 6.4 – Estatísticos da estimação para topologia de quatro nós e sete enlaces – MT Poisson.....	56
Tabela 6.5 – Estimação para topologia de quatro nós e cinco enlaces – MT Constante.....	57
Tabela 6.6 – Estatísticos da estimação para topologia de quatro nós e cinco enlaces – MT Constante .....	58
Tabela 6.7 – Estimação para topologia de quatro nós e cinco enlaces – MT Poisson .....	59
Tabela 6.8 – Estatísticos da estimação para topologia de quatro nós e cinco enlaces – MT Poisson.....	60
Tabela 6.9 – Estimação para topologia de quatro nós e nove enlaces – MT Constante.....	62
Tabela 6.10 – Estatísticos da estimação para topologia de quatro nós e nove enlaces – MT Constante .....	63
Tabela 6.11 – Estimação para topologia de quatro nós e nove enlaces – MT Poisson .....	64
Tabela 6.12 – Estatísticos da estimação para topologia de quatro nós e nove enlaces – MT Poisson.....	65
Tabela 6.13 – Estatísticos da estimação - Resumo.....	65
Tabela 6.14 – Matriz de tráfego real (exemplo) - RNP .....	67
Tabela 6.15 – Estatísticos de matriz de tráfego real exemplo - RNP .....	67
Tabela 6.16 – Matriz de tráfego real (exemplo) - GÉANT .....	70

# Lista de Figuras

Figura 1.1 – Exemplo de uma rede simples .....	17
Figura 3.1 – Estrutura de um Algoritmo Genético .....	27
Figura 3.2 – Cruzamento em um Ponto .....	32
Figura 3.3 – Operador de mutação .....	33
Figura 4.1 – Abordagem de Estimação de MT da pesquisa .....	38
Figura 5.1 – Rede com 4 (quatro) nós simples (sete enlaces) .....	45
Figura 5.2 – A rede Ipê - RNP .....	46
Figura 5.3 – Visão geral - Anel de 10 Gbps e dois anéis de 2,5 Gbps - RNP .....	47
Figura 5.4 – Mapa Geral da GÉANT – Outubro de 2004. ....	48
Figura 6.1 – Estimação de Matriz de Tráfego – Topologia de quatro nós e sete enlaces – MT Constante .....	53
Figura 6.2 – Estimação de Matriz de Tráfego – Topologia de quatro nós e sete enlaces – MT Poisson.....	55
Figura 6.3 – Rede com 4 (quatro) nós simples (cinco enlaces).....	56
Figura 6.4 – Estimação de Matriz de Tráfego – Topologia de quatro nós e cinco enlaces – MT Constante .....	58
Figura 6.5 – Estimação de Matriz de Tráfego – Topologia de quatro nós e cinco enlaces – MT Poisson.....	60
Figura 6.6 – Rede com 4 (quatro) nós simples (nove enlaces).....	61
Figura 6.7 – Estimação de Matriz de Tráfego – Topologia de quatro nós e nove enlaces – MT Constante .....	62
Figura 6.8 – Estimação de Matriz de Tráfego – Topologia de quatro nós e nove enlaces – MT Poisson.....	64
Figura 6.9 - Matriz de tráfego original x erros de estimação – tráfego constante e utilizando AG .....	66
Figura 6.10 – Estimação da MT utilizando AG e o erro de estimação.....	68
Tráfego de Poisson – RNP .....	68
Figura 6.11 – Estimação da MT utilizando AG e o erro de estimação.....	69
Tráfego de Poisson – GÉANT.....	69

# Lista de Abreviaturas

AG	Algoritmo Genético
EM	Maximização de expectativa
Géant	Rede Gigabit de Pesquisa Pan-Européia
IGP	Interior Gateway Protocols
IP	Internet Protocol
IS-IS	Intermediate System
MCMC	Markov Chain Monte Carlo
MIB	Management Information Base
MT	Matriz de tráfego
OD	Origem-Destino
OSPF	Open Shortest Path First
PL	Programação Linear
PoP	Pontos-de-Presença
QoS	Quality of service
RNP	Rede Nacional de Ensino e Pesquisa
SBC	Sociedade Brasileira de Computação
SBRC	Simpósio Brasileiro de Rede de Computadores
SNMP	Simple Network Management Protocol

# Sumário

<b>Resumo .....</b>	<b>7</b>
<b>Abstract .....</b>	<b>8</b>
<b>Lista de Tabelas .....</b>	<b>9</b>
<b>Lista de Figuras .....</b>	<b>10</b>
<b>Lista de Abreviaturas.....</b>	<b>11</b>
<b>Capítulo 1 .....</b>	<b>14</b>
Introdução.....	14
1.1 Definição do Problema .....	14
1.2 Formalização do Problema de Estimação de Matrizes de Tráfego .....	16
1.3 Objetivos.....	18
1.3.1 Objetivo Geral da Dissertação .....	18
1.3.2 Objetivos Específicos .....	18
1.4 Relevância .....	18
1.5 Organização da Dissertação.....	19
1.6 Comentários.....	20
<b>Capítulo 2 .....</b>	<b>21</b>
Trabalhos relacionados .....	21
2.1 O Estado da Arte .....	21
2.2 Comentários.....	23
<b>Capítulo 3 .....</b>	<b>24</b>
Fundamentação Teórica.....	24
3.1 Modelo de Gravidade Simples .....	24
3.2 Algoritmo Genético .....	25
3.2.1 Introdução .....	25
3.2.2 Estrutura do Algoritmo Genético .....	26
3.2.3 Codificação .....	27
3.2.3.1 Considerações Gerais .....	27
3.2.3.2 Codificação Binária .....	28
3.2.3.3 Codificação Gray .....	28
3.2.3.4 Codificação Real.....	28
3.2.4 Inicialização.....	29
3.2.5 Avaliação .....	29
3.2.6 Seleção.....	30
3.2.7 Operadores Genéticos.....	31
3.2.7.1 Cruzamento.....	31
3.2.7.2 Mutação .....	32
3.3 Programação Linear.....	33
3.4 Abordagem Bayesiana.....	34
3.5 Comentários.....	36

<b>Capítulo 4 .....</b>	<b>37</b>
Proposta .....	37
4.1 Considerações Iniciais .....	37
4.2 Modelo de Estimação .....	37
4.2.1 Algoritmo de Inicialização .....	38
4.2.2 Algoritmo Genético para Estimação de MT .....	39
4.2.2.1 Configuração dos Parâmetros .....	40
4.2.2.2 Definição do Algoritmo Genético .....	41
4.2.2.3 Pseudocódigo do Algoritmo Genético proposto .....	42
4.3 Comentários .....	42
<b>Capítulo 5 .....</b>	<b>44</b>
Ambiente Operacional e Cenário .....	44
5.1 Cenários .....	44
5.1.1 Dados Sintéticos .....	44
5.1.2 Dados Reais .....	45
5.1.2.1 Rede Nacional de Ensino e Pesquisa - RNP .....	45
5.1.2.2 Rede Gigabit de Pesquisa Pan-Européia - Géant .....	48
5.2 Comentários .....	49
<b>Capítulo 6 .....</b>	<b>50</b>
Resultados e Discussões .....	50
6.1 Avaliação comparativa .....	50
6.2 Resultados utilizando dados sintéticos .....	51
6.2.1 Topologia de quatro nós com sete enlaces .....	51
6.2.1.1 Topologia de quatro nós com sete enlaces e tráfego de Poisson .....	54
6.2.2 Topologia de quatro nós com cinco enlaces .....	56
6.2.2.1 Topologia de quatro nós com cinco enlaces e tráfego de Poisson .....	59
6.2.3 Topologia de quatro nós com nove enlaces .....	61
6.2.3.1 Topologia de quatro nós com nove enlaces e tráfego de Poisson .....	63
6.3 Resultados utilizando dados reais .....	66
6.3.1 primeiro caso: Rede Nacional de Ensino e Pesquisa - RNP .....	67
6.3.1.1 Aplicação da estimação com Inferência Bayesiana e Algoritmo Genético ..	68
6.3.2 segundo caso: Rede Gigabit de Pesquisa Pan-Européia - Géant .....	69
6.3.2.1 Aplicação da estimação com Inferência Bayesiana e Algoritmo Genético ..	69
6.4 Comentários .....	71
<b>Capítulo 7 .....</b>	<b>72</b>
Conclusões e Trabalhos Futuros .....	72
7.1 Conclusões .....	72
7.2 Trabalhos Futuros .....	74
<b>Referências Bibliográficas .....</b>	<b>75</b>

# Capítulo 1

## Introdução

O presente trabalho apresenta a dissertação, submetida como requisito para obtenção do grau de mestre no Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências e Tecnologia da Universidade Estadual do Ceará (UECE).

Este capítulo é uma introdução à dissertação. Inicialmente, a seção 1.1 apresenta o problema e a motivação do trabalho, destacando a originalidade da solução. A seção seguinte, 1.2, descreve a formalização do problema de estimação de matrizes de tráfego. A seção 1.3 apresenta o objetivo geral e os objetivos específicos da dissertação. A relevância da pesquisa é apresentada na seção 1.4. A seção 1.5 apresenta como a dissertação está organizada. Por fim, a seção 1.6 encerra o capítulo com os comentários finais.

### 1.1 Definição do Problema

Uma Matriz de Tráfego (MT) reflete o volume do tráfego transmitido entre todos os possíveis nós Origem-Destino (OD) em um domínio de rede de comunicação, representando as demandas dos usuários da rede. Esses nós podem ser enlaces, roteadores ou Pontos-de-Presença (PoPs) na Internet, durante um determinado período de tempo. Existe a necessidade de distribuir corretamente esse tráfego, visando combater o congestionamento na rede e beneficiando os usuários do sistema.

Mudanças no comportamento dos usuários, mudanças nas políticas de roteamento, falhas dos componentes da rede pode resultar em significativas (e repentinas) volubilidade de carga em uma grande rede de *backbone* IP [FELDMANN et al., 2000]. Dessa forma, o conhecimento de uma MT é bastante valioso para os operadores de rede devido a ampla variedade de controle e tarefas de gestão, incluindo provisão de serviço, equilíbrio de carga, configuração de protocolo de roteamento, análise de confiança, detecção de anomalia, implementação de políticas de QoS e gerenciamento da rede [NI, TATIKONDA, YEH, 2006]. Em casos de falha em algum enlace do domínio é possível, através da MT, prever as novas cargas nos enlaces restantes. Grande parte das abordagens anteriores usadas para estimar o tráfego, a partir de matrizes de dados disponíveis, basicamente usa medições de carga de enlaces e configurações de roteamento, a partir da MIB (*Management Information Base*) do SNMP (*Simple Network Management Protocol*). Mas é importante mencionar que, de acordo com Roughan, Thourp, e Zhang (2003), os dados de cargas nos enlaces medidos via

SNMP podem possuir restrições. Dados perdidos ou incorretos e um intervalo de amostragem longo são exemplos de limitações. Portanto, os dados medidos via SNMP têm erros e sua confiabilidade pode variar.

Embora seja importante conhecer a matriz de tráfego, a sua exata determinação não se trata de uma tarefa trivial quando no tocante das redes de comunicações modernas. As dificuldades desta estimação surgem de várias origens. Primeiro, as redes em geral apresentam baixa conectividade de forma que o número de enlaces resultante é menor que o número de fluxos OD. Isto dá origem a um problema inverso sub-determinado que permite infinitas soluções. Segundo, pode não ser fácil obter a matriz de roteamento correspondente aos fluxos medidos.

Várias técnicas já foram implementadas para a resolução do problema de estimação de matriz de tráfego. De um modo geral, os mecanismos atuais mais usados na estimação da MT são de duas classes de métodos principais: métodos baseados em inferência estatística, métodos baseados em otimização.

Nos métodos baseados em inferência estatística seleciona-se uma distribuição para os elementos da matriz de tráfego e, posteriormente, inferem-se os parâmetros desta distribuição que melhor se adaptam aos dados colhidos [VARDI, 1996; CAO et al., 2000], motivo da denominação do método. Medina et al., (2002) concluíram que os métodos de inferência estatística são muito sensíveis a erros e pequenas distorções nos dados de entrada podem comprometer fortemente a qualidade dos resultados. Esta limitação tem levado a outros modelos que empregam informações extras e pressupostos diferentes [XIANG, MAKLER, SILVA, 2008].

Já com métodos baseados em otimização, tenta-se filtrar as melhores soluções em um sistema e consiste em estabelecer uma função objetivo que de alguma forma possa ser usada para ordenar as soluções do problema de acordo com suas qualidades, e, em seguida, aplicar um algoritmo de otimização para encontrar as soluções ótimas dessa função.

Este trabalho propõe um método diferenciado para estimação de MT, desenvolvido a partir de um algoritmo genético baseado em uma população inicial de indivíduos gerada por um modelo de gravidade que busca uma maior precisão e robustez em redes reais.

A motivação da utilização de algoritmo genético para a estimação de matriz de tráfego deve-se, principalmente, ao fato de inexistir na literatura pesquisas com essa metodologia, podendo ser apresentado seus benefícios ou deficiência em relação a outros métodos de otimização e métodos estatísticos já pesquisados.

## 1.2 Formalização do Problema de Estimação de Matrizes de Tráfego

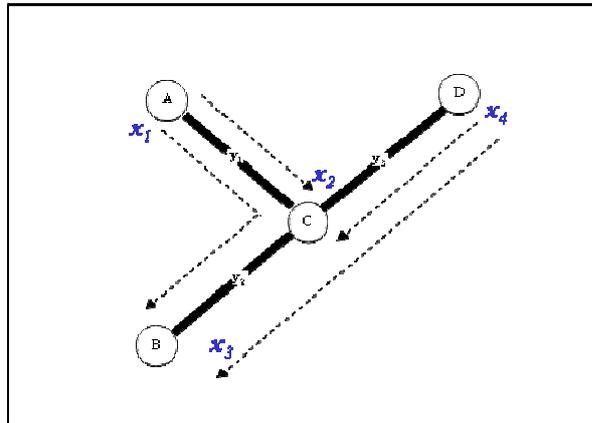
O problema de estimação de matrizes de tráfego pode ser formalizado da seguinte maneira. Um fluxo origem-destino consiste de todo tráfego que entra na rede em um ponto de ingresso comum e sai em outro ponto da rede. O tráfego observado em um enlace da rede é formado pela superposição de fluxos OD [MEDINA et al., 2002]. Seja  $c$  o número de pares origem-destino em um domínio de rede, formada por  $n$  nós de interesse em sua fronteira, então  $c = n(n-1)$ . Conceitualmente, os fluxos de tráfego podem ser representados por pares OD ordenados em um vetor  $\mathbf{X}$ , onde a quantidade de dados transmitida do nó  $i$  para o nó  $j$  é denotada pelo elemento  $x_{ij}$ . Seja  $\mathbf{Y} = [y_1, \dots, y_r]$  o vetor que representa o volume de tráfego nos enlaces individualmente, onde  $y_l$  indica o volume de tráfego para o enlace  $l$  e  $r$  denota o número de enlaces na rede. Os vetores  $\mathbf{X}$  e  $\mathbf{Y}$  estão relacionados através de uma matriz de roteamento  $\mathbf{A}$  de dimensões  $r$  por  $c$ . A matriz  $\mathbf{A}$  é composta por valores  $\{0,1\}$  com linhas representando os enlaces da rede e colunas representando os pares OD. O elemento  $a_{ij} = 1$  indica que o enlace  $i$  pertence ao caminho associado ao par OD  $j$ , enquanto  $a_{ij} = 0$  indica o oposto. Portanto, os fluxos OD estão relacionados aos volumes de tráfego nos enlace de acordo com a seguinte relação linear:

$$\mathbf{Y} = \mathbf{A} \cdot \mathbf{X} \quad (1)$$

Um administrador de rede conta com as medições de tráfego feitas em cada um dos canais de sua rede, ou seja, com um vetor  $\mathbf{Y}$ . Conta, ainda com a matriz de roteamento  $\mathbf{A}$ , e o seu objetivo é estimar  $\mathbf{X}$ , ou seja, nos ambientes de rede modernos  $\mathbf{Y}$  e  $\mathbf{A}$  estão disponíveis para o administrador. Realmente, o tráfego nos canais pode ser obtido a partir de medições feitas com o protocolo SNMP. No entanto, a matriz de roteamento pode ser computada, por exemplo, se assumirmos que o administrador possui informações sobre a topologia da rede, ou seja, sobre o custo de transmissão de dados em cada um dos canais. Afinal, tendo em mãos a informação sobre o custo de cada canal pode-se aplicar um algoritmo de roteamento para determinar o caminho de menor custo entre cada par de *hosts* na rede. Cada caminho

corresponde a uma das colunas da matriz  $A$ . O caminho ótimo depende de fatores como custo, capacidade e política administrativa.

Logo, o problema é calcular o vetor  $\mathbf{X}$ , ou seja, encontrar um conjunto de fluxos OD que possa reproduzir o volume de tráfego dos enlaces da maneira mais próxima possível. O problema associado à equação (1) é altamente subdeterminado, pois em praticamente todas as redes o número de pares OD é muito maior do que o número de enlaces na rede,  $r \ll c$ . Isto significa que há um número infinito de soluções possíveis para o vetor  $\mathbf{X}$ .



**Figura 1.1 – Exemplo de uma rede simples**

Considere o exemplo apresentado na Figura 1.1. Temos três nós  $A$ ,  $B$  e  $D$  interconectados por um quarto nó, o roteador  $C$ . Além disto, dispomos de seis canais direcionados, mas vamos nos deter naqueles três por onde passa o tráfego:  $L_1$  ( $AC$ ),  $L_2$  ( $CB$ ) e  $L_3$  ( $DC$ ) os outros três canais são  $CA$ ,  $BC$  e  $CD$ . Assumimos que as informações sobre o tráfego individual nestes canais estão disponíveis e compõem o vetor  $\mathbf{Y} = (y_1, y_2, y_3)^T$ , onde  $T$  representa um determinado instante. Os pares OD estão representados pelas linhas tracejadas. O problema chave consiste em estimar a matriz de tráfego  $\mathbf{X} = (x_1, x_2, x_3, x_4)^T$ . A relação  $\mathbf{Y} = \mathbf{A} \cdot \mathbf{X}$  é dada por:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \quad (2)$$

Em particular,  $x_1$  representa o tráfego transmitido do *host* A para o *host* B e  $y_1 = x_1 + x_2$ .

## 1.3 Objetivos

### 1.3.1 Objetivo Geral da Dissertação

O objetivo do trabalho é desenvolver uma abordagem fundamentada na computação evolucionária, mais especificamente um algoritmo genético, que busque uma estimação com maior precisão da matriz de tráfego em redes IP, e avaliar o seu desempenho comparando-o com os resultados obtidos por técnicas encontradas na literatura, e utilizando de topologias de dados sintéticos e reais.

### 1.3.2 Objetivos Específicos

A seguir enumeram-se os objetivos específicos deste trabalho:

- a) Analisar o mecanismo de roteamento de rede para detectar o problema da estimação da matriz de tráfego;
- b) Estudar alguns dos mais importantes métodos de estimação de matriz de tráfego presente na literatura;
- c) Investigar e identificar as técnicas convenientes que já foram implementadas para a resolução do problema de estimação de matriz de tráfego;
- d) Especificar e conceber um algoritmo para estimação da matriz de tráfego OD utilizando a idéia de computação evolucionária;
- e) Desenvolver o algoritmo proposto para um ambiente de redes IP;
- f) Simular a execução do algoritmo em um ambiente de redes de dados sintéticos e em um ambiente de rede de dados reais;
- g) Comparar a eficiência do algoritmo em relação a outras técnicas lineares e/ou estatísticas sob diferentes números de enlace.

## 1.4 Relevância

Atualmente as redes de comunicação, como a Internet, crescem em tamanho e diversidade e, devido a isso, técnicas de estimação tornam-se cada vez mais importantes para

o controle da rede, para as tarefas de gestão e planejamento de capacidade, além da previsão de degradações nos serviços oferecidos aos usuários finais [NI, TATIKONDA, YEH, 2006].

Muitas das decisões que um administrador de rede IP deve tomar dependem do fluxo de tráfego de sua rede. O planejamento das necessidades futuras em termos de capacidade e dimensionamento em uma rede IP é uma tarefa desafiadora e a estimação de matrizes de tráfego se torna fundamental para muitas questões da engenharia de tráfego, auxiliando na projeção do crescimento das redes e diagnósticos de problemas. A falta de informações sobre os fluxos implica na ausência de uma visão central das redes a qual é necessária para a implementação de uma engenharia eficiente [MEDINA et al., 2002].

Pesquisas anteriores tentaram resolver o problema de estimação de matriz de tráfego propondo a medição direta do volume de tráfego entre os nós origem e destino. No entanto, essas abordagens não foram totalmente exploradas em grandes redes IP, porque elas enfrentam obstáculos desafiadores de engenharia, como a falta de uma boa infra-estrutura de medição, elevados custos de comunicação e computacionais na transferência das grandes quantidades de informação [XIANG, MAKLER, SILVA, 2008]. São necessárias então abordagens distintas, que utilizem medições de tráfego nos enlaces [PAPAGIANNAKI et al., 2004].

Portanto, o problema é relevante e precisa de solução inovadora para resolvê-lo, principalmente em grandes redes IP, onde os métodos de estimação existentes não resolvem satisfatoriamente o problema da estimação de MT.

## **1.5 Organização da Dissertação**

Esta dissertação é constituída de 7 capítulos e das referências bibliográficas. O primeiro capítulo é uma introdução à dissertação. Descreve-se a contextualização do problema, a motivação, a formalização, o objetivo geral e os objetivos específicos que devem ser alcançados, além da relevância do problema, a contribuição do trabalho desenvolvido e a organização da dissertação.

A revisão bibliográfica do assunto está no Capítulo 2, que apresenta as técnicas relacionadas ao problema de estimação de MT existentes na literatura.

O Capítulo 3 apresenta as definições e a fundamentação teórica utilizada no trabalho. Descreve-se um estudo do modelo de gravidade como geração da informação prévia

de estimação da matriz, os conceitos pertinentes aos algoritmos genéticos e seus operadores e a formalização dos dois métodos de estimação usados como parâmetro de comparação de desempenho em relação ao modelo proposto.

O Capítulo 4 trata da proposta deste trabalho. É feita a descrição detalhada do modelo proposto e seus procedimentos de execução, explicitando os métodos e as técnicas a serem empregadas.

O Capítulo 5 apresenta o detalhamento do ambiente operacional ou cenário usado na pesquisa. O capítulo tratará da descrição das topologias de redes usadas nos casos utilizados para comprovar a eficiência do método proposto para estimação do tráfego de rede.

Os resultados obtidos nesta dissertação estão concentrados no capítulo 6. Nele se apresenta a análise de todos os resultados obtidos e comparados com métodos existentes na literatura.

Finalmente, o Capítulo 7 dedica-se às conclusões, dando ênfase aos resultados, encerrando a dissertação com as conclusões do trabalho e futuras pesquisas.

## **1.6 Comentários**

Este capítulo expôs inicialmente o problema que se quer resolver e seu formalismo, os objetivos que devem ser alcançados com a conclusão da dissertação, além da relevância do problema para as redes IP. Por fim, foi apresentada a estrutura geral da dissertação.

## Capítulo 2

### Trabalhos relacionados

Este capítulo trata do estudo do que já foi publicado sobre MT e do arcabouço teórico que fundamenta a realização da dissertação.

#### 2.1 O Estado da Arte

Existem, na literatura, muitos trabalhos tratando de métodos de estimação de MT. Na área de telecomunicações, os estudos de estimação de tráfego telefônico origem-destino tiveram início em 1937, onde Kruithof (1937) sugere um método para estimação de demanda de tráfego ponto-a-ponto em uma rede de telefonia baseada em uma matriz de tráfego conhecida e as medições de entrada e saída de tráfego. O método de Kruithof foi inicialmente analisado por Krupp (1979) e mostrou que a abordagem pode ser interpretada a partir de uma informação teórica do ponto de vista, que minimiza a distância de Kullback-Leibler de uma suposta matriz tráfego anterior. Além disso, Krupp (1979) mostrou que o método iterativo estendido converge para a única solução ótima. O método de Kruithof foi o primeiro método iterativo estatístico para estimação de MT.

No entanto, foi só na segunda metade da década de 90, que o problema foi abordado especificamente para redes IP. Vardi (1996) foi o primeiro a investigar a questão nessas redes e a chamou de “tomografia de rede”. Vardi (1996) estudou o problema para uma rede com uma topologia geral e usou um modelo de Poisson para a contagem de byte do tráfego OD. Usando o modelo de Poisson, a média e a covariância da amostra dos *links* é calculada para uma seqüência de medidas. As amostras são utilizadas como restrições adicionais. As demandas de tráfego são estimadas por estimativa de máxima verossimilhança. O modelo de Poisson também é utilizado por Tebaldi e West (1998), mas em vez de usar estimativa de máxima verossimilhança, utilizam uma abordagem Bayesiana. Como distribuições posteriores são difíceis de calcular, os autores usam a técnica MCMC (*Markov Chain Monte Carlo*) para obter a distribuição posterior. Cao et al. (2000) modelaram o tráfego OD como variáveis aleatórias normais independentes, usando o método EM (Maximização de expectativa) para estimar os parâmetros média e variância. Já Goldschmidt (2000), investigou as técnicas de otimização. Ele primeiro formulou o problema linear da estimação de matriz de tráfego usando um modelo de programação linear, e usou uma técnica constante correspondente para resolvê-lo. A abordagem Bayesiana foi refinada por Vaton e Gravey

(2003), que propuseram um método iterativo para melhorar a distribuição a priori dos elementos da matriz de tráfego. A estimação da matriz de tráfego a partir da medição do *link* é utilizada na próxima estimativa utilizando novas medidas do *link*. O processo é repetido até que nenhuma mudança significativa é feita na matriz de tráfego estimada.

Em [MEDINA et al., 2002] são comparados três métodos para a estimação de matrizes de tráfego. O primeiro método aplica diretamente uma abordagem de programação linear descrita em [GOLDSCMIDT, 2000], onde a carga em determinado enlace da rede é a soma de todas as demandas que utilizam este enlace, sendo possível determinar um conjunto de equações que permita solucionar o problema. O segundo método utiliza técnicas de inferência Bayesiana [TEBALDI e WEST, 1998]. Neste método, o objetivo é computar a distribuição de probabilidade condicional para todas as demandas, dada a carga nos enlaces. Para isto, é necessária uma distribuição prévia para as demandas. O terceiro método adota a abordagem EM [CAO et al., 2000] para calcular estimativas de máxima verossimilhança (*maximum likelihood*) onde os pares OD são modelados de acordo com uma distribuição gaussiana, procurando-se por uma solução mais adequada. Uma nova abordagem baseada em modelos de escolha também é sugerida. O modelo de escolha tenta estimar a probabilidade de um nó origem enviar um pacote para um nó destino na rede.

Similar ao modelo de escolha de [MEDINA et al., 2002] é o modelo de gravidade introduzido por Zhang et al. (2003b). Na sua forma mais simples o modelo de gravidade assume uma relação de proporcionalidade entre o tráfego entrando na rede em um nó  $i$  e destinado a um nó  $j$  e a quantidade total do tráfego com origem nesse nó  $i$  e a quantidade total de tráfego da rede com destino a esse nó  $j$ . Os autores utilizaram informações adicionais sobre a estrutura e configuração da rede para melhorar o desempenho do método.

Ao obter-se uma estimação da matriz de tráfego, esta estimação pode conter erros que a afastam da matriz real. Em [ROUGHAN et al., 2004] são avaliados os efeitos de estimativas inexatas de matrizes em engenharia de tráfego.

Outras abordagens para o problema de estimação de matrizes de tráfego foram propostas recentes, tais como uma abordagem baseada em teoria da informação [ZHANG et al., 2003b]. Nela, à distância Kullback-Leibler é utilizada para minimizar a informação mútua entre origem e destino. Outra técnica adota um estimador da variância das matrizes de tráfego [SOULE et al., 2004]. O efeito das mudanças no roteamento da rede e na variação das

matrizes de tráfego é investigado em [TEIXEIRA et al., 2005]. Uma avaliação do equilíbrio entre a adoção de medições, o uso de inferência e modelagem para a estimação de matrizes de tráfego pode ser encontrada em [SOULE et al., 2005].

Diferentemente dos trabalhos anteriores onde o roteamento é considerado constante, em [NUCCI et al., 2004], é proposta a idéia de coletar medições dos enlaces sob vários cenários de roteamento para determinar a demanda de tráfego. Feldmann et al. (2000) utilizam uma abordagem um pouco diferente para o cálculo da demanda de tráfego. Em vez de estimar a partir de *links* eles colecionam medições de fluxo de rotas utilizando ferramentas Netflow da Cisco e derivam a demanda de tráfego ponto-a-multiponto usando informações de roteamento dos protocolos de roteamento inter e intra domínio.

Recentemente, Xiang, Makler, Silva (2008) propuseram uma abordagem que formula o problema de estimação de matriz de tráfego como um problema de otimização não-negativo, e então um método de projeção é utilizado para resolvê-lo. Desenvolveram também uma nova abordagem para a estimação de matrizes de tráfego quando o roteamento *optimal multi-path* é empregado. O problema foi considerado como um problema de programação binível e em seguida, um algoritmo genético é utilizado para resolvê-lo.

## 2.2 Comentários

O problema de estimação origem-destino para o tráfego telefônico é um problema bastante estudado no mundo Telecom. Várias técnicas foram usadas, transformando a função de roteamento para acomodar mais tráfego. No entanto, somente na década de 1990, se tornou uma importante área de investigação em redes IP. Este capítulo analisou as principais pesquisas na área nos últimos anos, caracterizando cada uma delas.

## Capítulo 3

### Fundamentação Teórica

Este capítulo trata do estudo dos fundamentos teóricos que darão suporte ao modelo proposto neste trabalho. A primeira parte deste capítulo é dedicada ao estudo do modelo de gravidade como geração da informação prévia para estimação da matriz. Também neste capítulo são expostos os conceitos de algoritmo genético e operadores. A quarta e a quinta parte tratarão de expor a formalização dos dois métodos de estimação usados como parâmetro de comparação de desempenho em relação ao modelo proposto.

#### 3.1 Modelo de Gravidade Simples

O modelo de gravidade, proposto por Zhang et al. (2003a), baseado no modelo de gravidade de Newton, é uma técnica que estima os valores do vetor  $X$  da equação (1). Esses modelos estão baseados na suposição de que a origem (fonte) e o destino (*sink*) de um fluxo de tráfego são independentes. Devido a isso, nesses modelos, o volume de fluxo de tráfego  $X_{ij}$  com origem em  $i$  e destino em  $j$  é proporcional a  $X_{i\bullet}$  (o volume total de tráfego gerado por  $i$ ) e a  $X_{\bullet j}$  que corresponde o volume de tráfego total com destino em  $j$ , conforme a equação (3):

$$X_{ij} \propto X_{i\bullet} \times X_{\bullet j} \quad (3)$$

onde, no caso de uma rede *backbone* IP,  $X_{i\bullet}$  e  $X_{\bullet j}$  podem ser obtidos facilmente através do SNMP.

Uma grande vantagem dos modelos de gravidade é que não são baseados em suposições de distribuição de probabilidade de tráfego (por exemplo, tráfego Poisson ou Gaussiano). Outra vantagem, é que esses modelos são capazes de produzir facilmente estimativas aproximadas de matriz de tráfego OD, podendo ser usadas como pontos de partida em métodos de estimação mais sofisticados.

Uma formulação geral do modelo de gravidade pode ser dada pela seguinte equação:

$$X_{ij} = \frac{f(R_i, A_j)}{g_{ij}} \quad (4)$$

onde  $f(.)$  é uma função não-decrescente,  $X_{ij}$  é o volume de tráfego de  $i$  para  $j$ ,  $R_i$  é um parâmetro que representa fatores *repulsivos* que são associados com “partida”  $i$ ,  $A_j$  é um parâmetro que representa fatores *atrativos* relacionado a “ida” para  $j$ , e  $g_{ij}$  representa o fator de fricção entre  $i$  e  $j$ . Esse fator codifica a informação específica de localização para diferentes pares origem-destino. Porém, como localização não é um grande fator no tráfego da internet, assume-se uma constante comum para fatores de fricção.

## 3.2 Algoritmo Genético

### 3.2.1 Introdução

Inspirados nos princípios da evolução biológica de Charles Darwin, em que indivíduos mais aptos estão mais capacitados a sobreviver e gerar descendentes pela própria seleção natural, perpetuando suas características através da hereditariedade, os Algoritmos Genéticos (AG) são uma versão simplificada do que se passa na natureza. Durante a reprodução, genes dos pais se combinam para formar um novo indivíduo. O descendente criado pode eventualmente sofrer mutação, resultado principalmente de erros na cópia dos genes dos pais, o que possibilita a incorporação de novo material genético a uma população, e o reaparecimento de alguma característica que possa ter desaparecido ao longo da evolução. A aptidão do indivíduo da população é medida pelo sucesso do organismo na sua vida, suas chances de sobrevivência. Os indivíduos mais aptos, que sobrevivem por mais tempo na população têm, portanto, maiores chances de gerar descendentes e, por conseguinte, perpetuar as suas características na população.

Os AGs foram introduzidos em 1975 por John Holland [HOLLAND, 1975] da Universidade de Michigan [GOLDBERG, 1989], com o intuito de formalizar matematicamente e explicar rigorosamente processos de adaptação em sistemas naturais e desenvolver sistemas artificiais (simulados em computador) que reproduzam os mecanismos originais encontrados em sistemas naturais. O tema central de sua pesquisa em algoritmos genéticos foi à robustez, isto é, o balanço entre a eficiência e a eficácia.

Eficientes para busca de soluções aproximadamente ótimas, os AGs não impõem muitas restrições encontradas em métodos tradicionais de busca. São capazes de convergir para soluções ótimas em níveis globais, contornando a ocorrência de ótimos locais e identificar e explorar aspectos do ambiente onde o problema está inserido e convergir

globalmente para soluções ótimas ou aproximadamente ótimas [HOLLAND, 1975; GOLDBERG, 1997].

Nas últimas décadas os AGs vêm sendo amplamente utilizados na solução de problemas nas mais diversas áreas do conhecimento que visam a encontrar um indivíduo ou um conjunto de indivíduos que melhor se adapte a uma série de condições ambientais previamente estabelecidas.

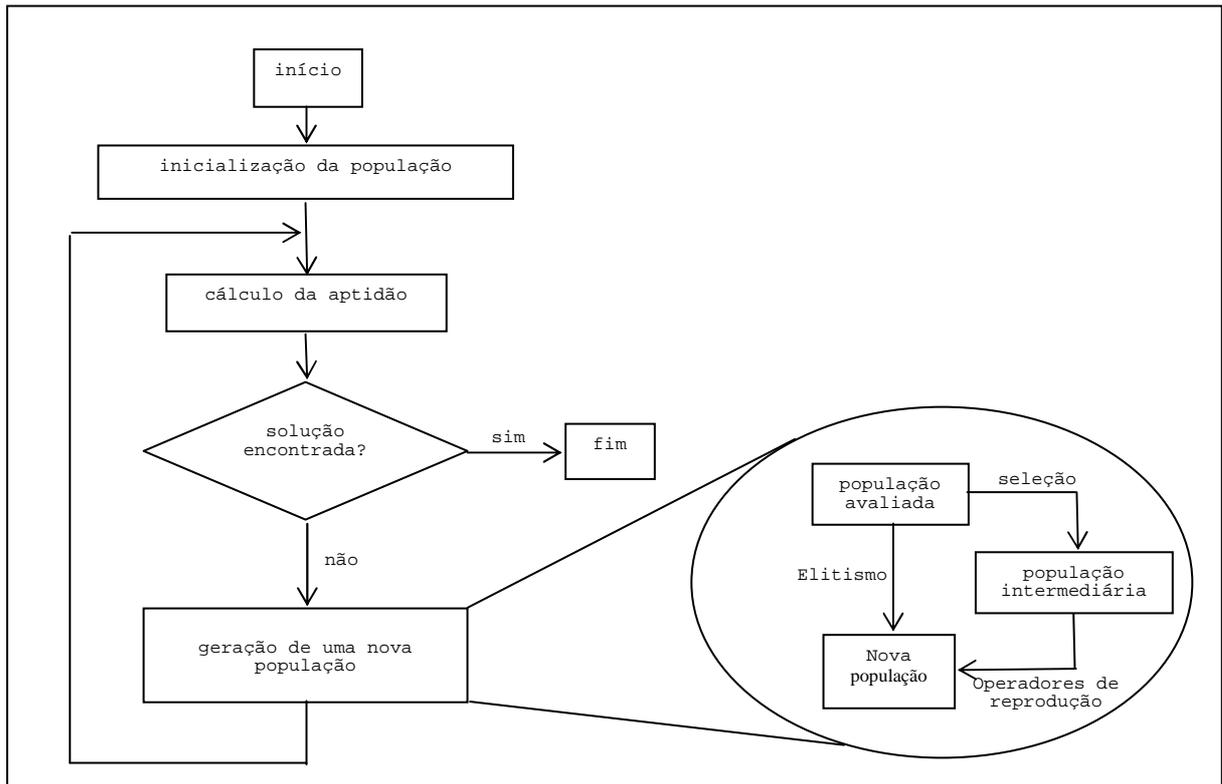
### 3.2.2 Estrutura do Algoritmo Genético

A implementação de um AG tem início com a definição de uma função objetivo que se deseja minimizar ou maximizar e uma população de cromossomos, os quais representam possíveis soluções do problema a ser resolvido, aplicando um cruzamento (*crossover*) e mutação como mecanismo de busca. Durante o processo evolutivo essas estruturas são avaliadas segundo uma função de aptidão e a elas são associadas uma probabilidade de reprodução, de tal forma que as maiores probabilidades são atribuídas aos cromossomos que representam uma melhor solução para o problema de otimização. Trata-se de uma técnica de busca randômica global no espaço de solução do problema e, usualmente, permite a captura em uma otimização local [GEN, 2000].

Por ser uma analogia da seleção natural de Darwin, o algoritmo genético adotou um vocabulário provindo da genética. Assim, um conjunto de soluções candidatas é denominado de população; cada solução candidata é denominada indivíduo ou cromossomo, que por sua vez é formada por genes. Cada iteração do algoritmo genético é chamada de geração; a combinação de dois (ou mais) indivíduos para se criar novos indivíduos é dito de recombinação ou cruzamento; e a modificação aleatória de um indivíduo é chamada de mutação.

Os Algoritmos Genéticos diferenciam-se dos algoritmos de otimização tradicionais em dois aspectos principalmente: Um AG gera uma população de pontos a cada iteração e essa população aproxima a uma solução ótima, enquanto que um algoritmo padrão gera um único ponto a cada iteração e a seqüência de pontos aproxima a uma solução ótima. O AG também seleciona a próxima população por cálculo que envolve escolha randômica, já os algoritmos padrões selecionam o próximo ponto na seqüência por cálculo determinístico.

A estrutura típica de um AG pode ser observada na Figura 3.1.



**Figura 3.1 – Estrutura de um Algoritmo Genético**

Embora possam parecer simplistas do ponto de vista biológico, esses algoritmos são suficientemente complexos para fornecer mecanismos poderosos e robustos de busca adaptativa [GOLDBERG, 1997; COELHO, 2003].

### 3.2.3 Codificação

#### 3.2.3.1 Considerações Gerais

A codificação é o primeiro passo para aplicar o AG num problema real, e, se for inadequada, o problema pode ficar difícil ou até impossível [LOPES, 1999]. Por outro lado, a codificação torna o AG independente dos parâmetros diretos do problema e, portanto, livre de eventuais restrições do espaço de solução inicial. Na codificação, as variáveis do problema são representadas, dentro do espaço de busca, por uma seqüência de símbolos (genes), que advêm de um alfabeto finito de opções (alelos). Após a codificação, as variáveis são reunidas num conjunto, formando um cromossomo e um ou mais cromossomos associados formam um indivíduo. Para facilitar a compreensão e simplificar o tratamento computacional, é comum usar indivíduos com um só cromossomo de tamanho fixo e o conjunto desses indivíduos gerar uma população também de tamanho constante.

### 3.2.3.2 Codificação Binária

A estrutura simples e analogia direta com a genética natural fazem da codificação binária uma das mais usadas. No caso elementar, o alfabeto de alelos pode ser binário, 0 ou 1, e a dimensão da seqüência que compõe o cromossomo depende da precisão requerida.

O uso da codificação binária dá origem a alguns problemas, como vetores muito extensos para representar indivíduos com alta precisão e o fenômeno denominado *Hamming cliffs* [TANOMARU, 1995]. Esse fenômeno consiste em grandes diferenças nas cadeias de “bits” que codificam dois números inteiros adjacentes, o que é enfatizado em ambientes dinâmicos e evidenciado em perturbações no “bit” mais significativo do parâmetro codificado, podendo causar um grande deslocamento no espaço de busca. Apesar de ampliar a busca em alguns casos, a presença de *Hamming cliffs* pode atrapalhar a otimização, pois, a menos de saltos aleatórios no espaço de busca codificado, haverá dificuldades de alcançar o ponto ótimo, se ele tem muitos “bits” diferentes de seus pontos vizinhos.

### 3.2.3.3 Codificação Gray

Uma alternativa à codificação binária é o código Gray, pois minimiza o efeito das *Hamming cliffs*, já que o alfabeto de alelos também pode ser binário, 0 ou 1, mas os números inteiros consecutivos são representados por seqüências que diferem num único “bit”. No entanto, a recíproca não é verdadeira e duas seqüências com apenas um “bit” de diferença podem não ser dois inteiros consecutivos (por exemplo, 3 e 12). Com base nisso conclui-se que a codificação binária amplia a região de exploração e localiza o ótimo global, mas tem dificuldade em refinar a solução, enquanto a codificação Gray favorece a convergência com maior precisão, mas pode produzir um ótimo local. A conversão entre os códigos binários e Gray é simples e facilmente é implantada num AG.

### 3.2.3.4 Codificação Real

Os problemas físicos em geral têm melhor codificação se ela for semelhante ao parâmetro modelado [WEILE e MICHELSEN, 1997; HOUCK, JOINES e KAY, 1995], ou seja, na otimização de variáveis contínuas por natureza, é muito útil usar números reais, empregando a codificação real ou contínua. Se comparada à binária, a codificação real pode reduzir a complexidade computacional, aumentar a precisão sem restringir a codificação e facilitar a hibridização, pois não há transformações decimal-binário-decimal para formar cromossomos e calcular a aptidão, a precisão não depende de número de bits dos parâmetros e é mais compatível com outros algoritmos de otimização. A codificação real também minimiza

o fenômeno *Hamming cliffs* e, segundo [HAUPT e HAUPT, 2004] e [HOUCK, JOINES e KAY, 1995], é mais rápida, mais precisa e tem resultados mais consistentes ao longo de repetições da execução do algoritmo.

Contudo, a codificação real possui um espaço de busca infinito (contínuo), exigindo certas adaptações na codificação e tornando os processos de recombinação e mutação mais complexos. Muitas vezes, essa complexidade requer soluções específicas, o que restringe a característica de flexibilidade proposta nos AGs, onde problemas distintos são tratados usando poucas ou nenhuma alteração. Isso posto, também merece ser registrado que alguns projetos com grande variedade de parâmetros e alto tempo de processamento, podem utilizar indivíduos com cromossomos de codificação híbrida, isto é, contendo partes binárias e partes reais, de acordo com as variáveis contidas no problema [WEILE e MICHIELSSEN, 1997].

### **3.2.4 Inicialização**

Em geral, a população é inicializada com pontos do espaço amostral (combinações de variáveis) escolhidos aleatoriamente, mas, dependendo da aplicação, podem existir formas heurísticas de selecionar uma população inicial mais favorável. De qualquer maneira, é importante que a população inicial esteja bem distribuída pelo espaço de busca, amostrando toda sua área. Primeiro, para que todas as partes do espaço amostral estejam representadas, e segundo por que o AG necessita de uma boa diversidade inicial de combinações para funcionar bem, da mesma forma que populações de organismos vivos precisam de uma boa diversidade genética para poderem se adaptar eficientemente ao seu meio ambiente.

### **3.2.5 Avaliação**

Cada indivíduo da população é avaliado no início de cada geração através de uma função aptidão, também chamada de *fitness*. Mede, tipicamente, algum custo a ser minimizado ou alguma recompensa a ser maximizada. Comparando com a natureza, indivíduos com maior aptidão têm mais chances de sobreviver no ambiente e reproduzir-se, passando seu material genético para as gerações posteriores. Assim, cada indivíduo deve possuir um valor de aptidão não-negativo, obtido de uma função de avaliação, denominada função objetivo, que, em casos simples, pode ser a própria função a otimizar. Onde a função objetivo fornece valores negativos, inadequados para alguns métodos de seleção, ou fornece valores muito próximos entre os indivíduos da população ou muito elevados para um pequeno

grupo destes indivíduos, se faz necessário mapear a função objetivo em valores de aptidão utilizando uma função de aptidão distinta da função objetivo. Pode-se considerar um único objetivo, ou pode agregar várias metas e restrições.

A função de avaliação deve também ser relativamente rápida. Como um algoritmo genético trabalha com uma população de soluções potenciais, este incorre o custo de avaliar-se esta população. Além disto, a população é substituída (em parte ou totalmente) a cada geração. Os membros da população reproduzem e sua cria deve ser então avaliada, e todos estes processos podem levar a pouca eficiência na solução se não se tem uma adequada função avaliação.

### 3.2.6 Seleção

Para emular os processos de seleção natural e reprodução, os AGs realizam uma seleção dos indivíduos da população, privilegiando as melhores soluções e gerando descendentes mais qualificados, implementando o mecanismo da sobrevivência do mais apto. Se não houvesse seleção, os AGs perderiam a essência de processo evolutivo e seriam ineficientes, semelhantes a buscas aleatórias. Todavia, a seleção não pode escolher apenas a melhor solução da população atual, já que essa pode não estar próxima da solução ótima global. Portanto, deve-se manter alguma chance de soluções piores participarem do processo de reprodução. Isso posto, as técnicas de seleção podem ser probabilísticas ou mesmo determinísticas, mas todas utilizam o valor da aptidão, isto é, quanto maior esse valor, maior a probabilidade de seleção. Dentre os métodos de seleção mais comuns, pode-se citar:

**Roleta (*Roulette Wheel*)** - Método de seleção proposto por Holland (1975), o qual usa uma seleção proporcional à função avaliação. O número esperado de vezes que um indivíduo será selecionado para reprodução é proporcional à sua avaliação dividida pela avaliação média da população. Para cada indivíduo é dada uma fatia de uma roleta (*roulette wheel*), proporcional à sua avaliação. Essa roleta é girada  $N$  vezes, onde  $N$  é o número de indivíduos da população. A cada giro, o indivíduo marcado na roleta é selecionado para a reprodução. Este método estocástico resulta no número esperado de filhos para cada indivíduo, mas de forma estatística. Quando se usa uma população com poucos indivíduos, o número alocado de filhos para cada indivíduo pode ficar longe de seu valor esperado [PAVANI, 2003].

**Amostragem Universal Estocástica (*Stochastic Universal Sampling*)** - Melhora o método Roleta, minimizando a diferença entre o número alocado de filhos para cada indivíduo e o seu valor esperado [BAKER, 1987].

**Seleção Rank (*Rank Selection*)** - Método de seleção proposto para evitar uma convergência prematura do algoritmo genético. Os indivíduos de uma população são ordenados segundo sua avaliação e o número esperado de vezes que um indivíduo será selecionado para reprodução depende da sua posição em relação aos demais indivíduos e não da sua avaliação. Após essa ordenação, o procedimento de amostragem poderia ser a Amostragem Universal Estocástica.

**Torneio (*Tournament Selection*)** - Método de seleção proposto para ser computacionalmente mais eficiente que a Seleção *Rank*, pois não necessita ordenar toda a população de acordo com a avaliação de cada indivíduo. Dois (ou mais) indivíduos são selecionados aleatoriamente da população e o “melhor” indivíduo nesse conjunto é selecionado para reprodução.

### 3.2.7 Operadores Genéticos

#### 3.2.7.1 Cruzamento

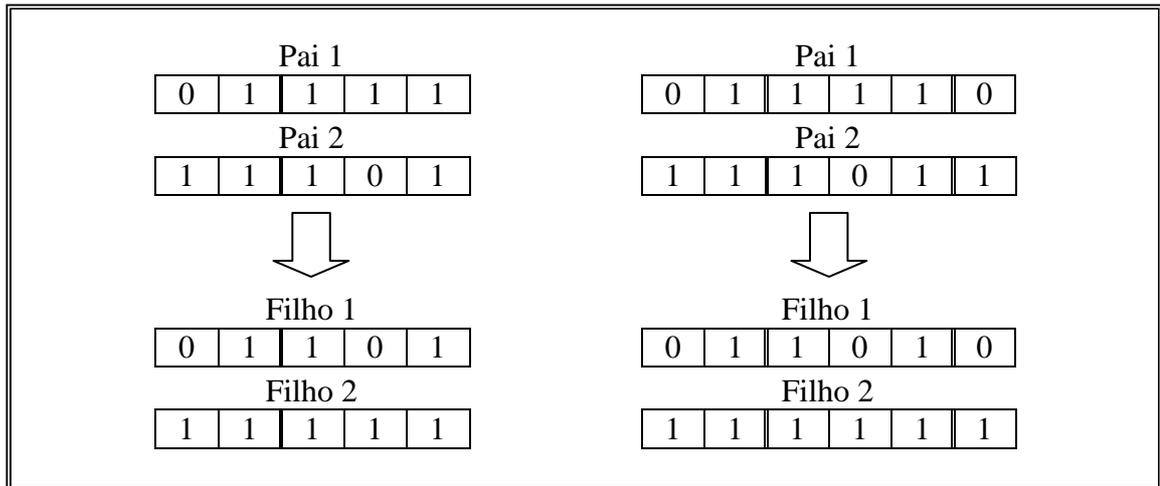
Também chamado de reprodução ou *crossover*. O princípio básico dos operadores de cruzamento é transformar a população através de sucessivas gerações, estendendo a busca até chegar a um resultado satisfatório. Os operadores de cruzamento são necessários para que a população se diversifique e mantenha características de adaptação adquiridas pelas gerações anteriores.

Durante a fase de reprodução de um algoritmo genético, selecionam-se indivíduos da população que serão recombinados para formar descendentes, que, por sua vez, constituirão a geração seguinte. Os pares são selecionados aleatoriamente, usando-se um método que favoreça os indivíduos melhores adaptados. Logo que forem escolhidos os pares, seus cromossomos se mesclam e se combinam, usando os operadores de cruzamento (*crossover*) e mutação. Eles são utilizados para assegurar que a nova geração seja totalmente nova, porém, mantendo características de adaptação adquiridas pelas populações anteriores.

Quanto maior for essa taxa, mais rapidamente novas estruturas serão introduzidas na população. Mas se for muito alta, estruturas com boas aptidões poderão ser retiradas mais

rapidamente que a capacidade da seleção em criar melhores estruturas. Se a taxa for muito baixa, a busca pode estagnar.

O operador de cruzamento cruza parte de dois indivíduos, trocando parte de suas informações genéticas e gerando dois novos indivíduos, como podemos observar na Figura 3.2.

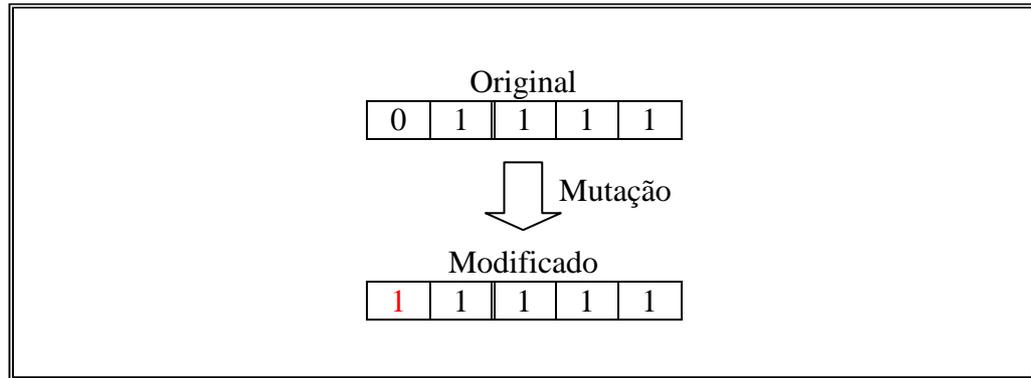


**Figura 3.2 – Cruzamento em um Ponto**

### 3.2.7.2 Mutação

A hereditariedade possibilita a estabilidade em sistemas biológicos. Porém, nenhum mecanismo composto de moléculas e sujeito ao impacto do mundo físico pode ser perfeito. Podem ocorrer erros na cópia, que produzem seqüências alteradas de DNA, ou seja, mutações que são perpetuadas.

O operador de mutação é necessário para a introdução e manutenção da diversidade genética da população, alterando arbitrariamente um ou mais componentes de uma estrutura escolhida, conforme ilustrado na Figura 3.3, o que fornece meios para introdução de novos elementos na população.



**Figura 3.3 – Operador de mutaçãõ**

Desta maneira, a mutaçãõ assegura que a probabilidade de chegar a qualquer ponto do espaço de busca nunca serã zero, além de contornar o problema de m´ınimos locais, pois este mecanismo altera levemente a direçãõ da busca.

O operador de mutaçãõ ´e aplicado aos indiv´duos com uma probabilidade dada pela taxa de mutaçãõ que comumente varia de 0,1 a 10%. Uma baixa taxa de mutaçãõ previne que a busca fique estagnada em sub-regiões do espaço de busca. Al´m disso, possibilita que qualquer ponto do espaço de busca seja atingido. Com uma taxa muito alta a busca se torna essencialmente aleat´ria.

### 3.3 Programaçãõ Linear

O m´etodo de estimaçãõ da matriz de tráfego usando Programaçãõ Linear (PL) foi proposto por Goldschmidt (2000) e, como visto na seçãõ 2.1, ´e o primeiro m´etodo usado em [MEDINA et al., 2002].

O problema de estimaçãõ do vetor  $X$  (fluxos OD) estã fortemente ligado aos relacionamentos lineares descritos pelo sistema da equaçãõ (1). Este problema pode ser facilmente formulado como um modelo de Programaçãõ Linear. Sabendo que o tráfego em cada enlace  $y_i$  ´e formado pela soma dos fluxos OD que utilizam esse enlace, uma funçãõ objetivo para este problema de otimizaçãõ da estimaçãõ do vetor  $X$  ´e dada por:

$$\max \sum_{j=1}^c w_j x_j, \quad (5)$$

onde  $w_j$  ´e o peso associado ao par OD $_j$ .

Além da função objetivo, o modelo básico de PL referenciado na literatura para este problema é composto por três conjuntos de restrições:

- restrições de enlace;
- restrições de conservação de fluxo; e
- restrições de positividade.

As restrições de enlace estabelecem que a soma dos fluxos OD que passam pelo enlace  $y_i$  não pode ser maior que o valor do tráfego medido nesse enlace. Estas restrições são dadas por:

$$\sum_{j=1}^c A_{lj} x_j \leq y_l \quad l = 1, \dots, r \quad (6)$$

As restrições de conservação do fluxo [Goldscmidt 2000] estabelecem que o tráfego total obtido na saída de um nó é igual à quantidade de tráfego que chega mais a quantidade de tráfego gerada por esse nó, menos a quantidade de tráfego que permanece no nó. Estas restrições são dadas por:

$$\sum_{l=(i,j)} y_l a_{lk} - \sum_{l=(j,i)} y_l a_{lk} = \begin{cases} x_k & \text{se } j = \text{origem de } k \\ -x_k & \text{se } i = \text{destino de } k \\ 0 & \text{caso contrário} \end{cases} \quad (7)$$

onde  $a_{lk}$  é a fração do tráfego  $x_k$  passa no canal  $l$ .

As restrições de positividade são dadas por  $x_j \geq 0$ , para todo  $j$ , significando que nem um fluxo pode ser negativo.

### 3.4. Abordagem Bayesiana

Outro método de estimação da matriz de tráfego conhecido utiliza técnicas de inferência Bayesiana e foi proposto originalmente por Tebaldi e West, (1998). [MEDINA et al., 2002] fez uso da técnica.

A abordagem Bayesiana para a estimação do vetor  $X$  calcula a distribuição de probabilidade condicional,  $p(X | Y)$ , de todos os fluxos OD (representados pelo vetor  $X$ )

dado o vetor de enlaces  $Y$ . Para alcançar esse objetivo é necessário ter a distribuição *a priori* de  $X$ ,  $p(X)$ . Tebaldi e West (1998) assumem que  $p(X_j)$  segue uma distribuição de Poisson com média  $\lambda_j$ , isto é,  $X_j \sim \text{Poisson}(\lambda_j)$ . Seja  $\Lambda = \{\lambda_1, \dots, \lambda_c\}$  o vetor que representa as taxas médias das distribuições dos elementos do vetor  $X$ . Como  $\Lambda$  é desconhecido e precisa ser estimado, torna-se necessário então definir uma distribuição para  $\Lambda$ . Isto nos conduz a um modelo de distribuição conjunta dado por  $P(X, \Lambda)$ . A idéia é então observar os valores dos enlaces e verificar se eles condicionam o valor de  $X$  para obter uma distribuição condicional conjunta dada por  $P(X, \Lambda | Y)$ .

A obtenção de distribuições *a posteriori*, como  $P(X, \Lambda | Y)$ , é computacionalmente muito difícil. A abordagem adotada por [Tebaldi e West 1998] para facilitar esse cálculo é aplicar métodos de simulação iterativa como Cadeia de Markov Monte Carlo (MCMC). A simulação da distribuição significa obter um grande número de amostras para representar um histograma completo da distribuição desejada. O objetivo final dessa simulação é calcular  $P(X | Y)$ .

O mecanismo de simulação iterativo mencionado relaciona a distribuição de probabilidade de  $X$  dado  $Y$  que se quer obter com a distribuição conjunta que envolve tanto  $X$  e  $\Lambda$  pela seguinte equação:

$$p(X | Y) = \sum_{\Lambda} P(X | \Lambda, Y) \quad (8)$$

Assim, o problema consiste em calcular as probabilidades a posterior de  $P(\Lambda | X, Y)$  e  $p(X | \Lambda, Y)$ . O procedimento de simulação começa com uma matriz  $X^0$  que pode ter valores quaisquer e as iterações seguintes são executadas da seguinte forma:

---

**Algoritmo 1** Algoritmo Bayesiano

---

1. Inicializar  $\Lambda^0$
2. Calcular  $X^0$  a partir de  $\Lambda^0$
3. Obter valores de  $\Lambda^i$  a partir de  $P(\Lambda | X^i, Y)$
4. Usando este  $\Lambda^i$ , obter valores para o vetor  $X^{i+1}$  a partir de  $p(X | \Lambda^i, Y)$
5. Repetir os passos 3 e 4 até uma solução viável ser encontrada.

### **3.5 Comentários**

Este capítulo apresentou o estudo que dará suporte ao objetivo do trabalho. Descreveu-se o modelo de gravidade como geração da informação prévia de estimação e o algoritmo genético e seus operadores. Foram apresentados, ainda, os dois métodos (PL e abordagem Bayesiana) usados para comparar a eficiência do algoritmo proposto.

## Capítulo 4

### Proposta

Este capítulo traz uma descrição detalhada da abordagem proposta onde são expostos os procedimentos, através dos métodos e das técnicas a serem empregadas.

#### 4.1 Considerações Iniciais

Os valores nas conexões dos nós representam o volume de tráfego nos enlaces individualmente. Cada nó na topologia representa um PoP e cada conexão representa a conectividade entre os pares adjacentes aos PoP. O objetivo é estimar a quantidade de tráfego em cada fluxo OD, dado a quantidade de tráfego total coletado na entrada e saída de cada nó, respectivamente, oriunda e direcionada para os nós. Os resultados obtidos a partir da estimação proposta são comparados com os valores reais.

Assim como em [MEDINA et al., 2002], serão usadas matrizes de tráfego com dados sintéticos e matrizes de dados reais. Nessa última, a matriz de tráfego é representada por informações fornecidas para a comunidade acadêmica com o propósito de desenvolvimento da pesquisa.

Em todos os casos, o vetor que caracteriza o tráfego da rede passando pelos canais pode ser obtido através do cálculo do produto da matriz de tráfego e da matriz de roteamento **A**.

A abordagem proposta considera o roteamento pelo caminho mais curto como mecanismo de roteamento das redes citadas, ou seja, os usuários seguem o seu caminho mais curto para chegar ao seu destino. Cada enlace está associado com um peso e o comprimento do caminho é definido como a soma dos pesos de todos esses enlaces do caminho. O caminho mais curto tem comprimento mínimo. O tráfego é roteado ao longo do caminho mais curto para o destino.

#### 4.2 Modelo de Estimação

O modelo proposto de estimação da matriz de tráfego é ilustrado na Figura 4.1. O objetivo final é obter uma matriz de tráfego, a partir da MT *a priori* **X** e apresentar uma taxa de erro de estimação o mais aceitável possível.

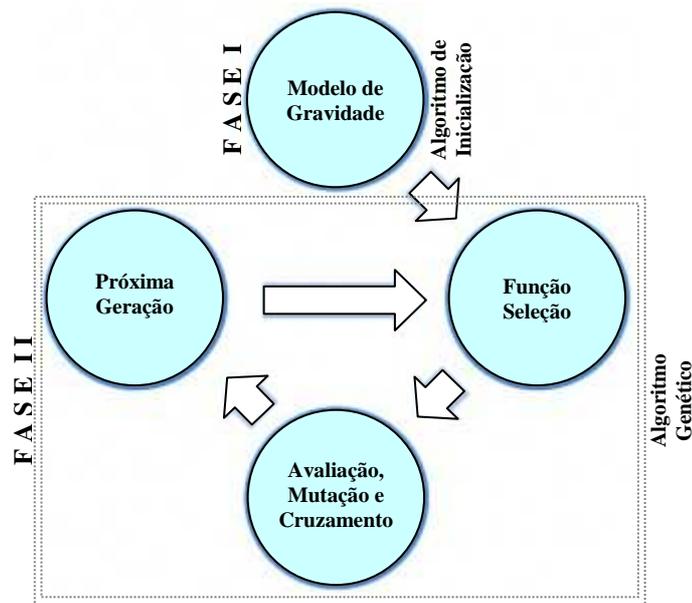


Figura 4.1 – Abordagem de Estimação de MT da pesquisa

O esquema proposto se divide em duas fases:

- **Fase I (algoritmo de inicialização):** Baseia-se no modelo de gravidade para gerar a população inicial do algoritmo genético proposto na Fase II.
- **Fase II (algoritmo genético):** Depois de ter a população inicial gerada a partir das idéias do modelo de gravidade da Fase I, tende a buscar a cada nova geração uma solução ótima, em um espaço de busca global.

#### 4.2.1 Algoritmo de Inicialização

Antes de analisar a primeira fase do proposto, é importante mencionar que, em relação à codificação (primeiro passo para aplicar o AG num problema real), são usados valores reais (codificação real), visando reduzir a complexidade computacional.

O algoritmo de inicialização proposto é dividido em duas etapas. Inicialmente, uma matriz de tráfego gravitacional, baseada no modelo gravitacional, é gerada. Supõe-se que a origem e o destino de um fluxo de tráfego OD são independentes e se leva em consideração que a troca de tráfego entre pontos na rede deve ser proporcional ao volume de tráfego que entra e sai em qualquer par OD. A MT gravitacional pode ser representada pela equação (9):

$$X_{ij}^g = \frac{\sum_{j=1}^c x_{ij} \sum_{i=1}^c x_{ij}}{\sum_{i=1}^c \sum_{j=1}^c x_{ij} - (c * \sum_{i=1}^c x_{ij})} \quad (9)$$

onde  $X_{ij}$  é o volume de tráfego que sai do nó  $i$  para o destino  $j$  e  $c$  representa o número de pares origem destino.

Nessa primeira etapa do algoritmo de inicialização, sabe-se que  $\sum_{j=1}^c x_{ij}$  representa o volume total de tráfego gerado por  $i$  ( $X_{i\bullet}$ ) e  $\sum_{i=1}^c x_{ij}$  corresponde o volume de tráfego total com destino em  $j$  ( $X_{\bullet j}$ ). Já fora mencionado na seção 3.2 que em uma rede *backbone* IP,  $X_{i\bullet}$  e  $X_{\bullet j}$  podem ser obtidos facilmente através do SNMP.

A segunda etapa consiste exatamente na criação da população inicial, baseada na matriz de tráfego gravitacional, gerada anteriormente.

$$\{x_1^{(k=0)}, \dots, x_N^{(k=0)}\} = \frac{(2 * rand(c)) * \left( X_{ij}^g * \sum_{i=1}^c \sum_{j=1}^c x_{ij} \right)}{\sum_{i=1}^c \sum_{j=1}^c X_{ij}^g} \quad (10)$$

onde  $k = 0$  indica se tratar da primeira geração de população,  $N$  representa o tamanho da população e a função *rand* é usada para gerar valores pseudo-aleatórios baseados no número de pares OD.

Ao final da Fase I, tem-se uma população inicial bem distribuída pelo espaço de busca, amostrando toda sua área. Nesse caso, todas as partes do espaço amostral estão representadas dando uma boa diversidade inicial de combinações para o AG funcionar bem.

#### 4.2.2 Algoritmo Genético para Estimção de MT

A partir da população inicial gerada anteriormente, inicia-se a Fase II do modelo concebido. O algoritmo genético começa com a população  $\{x_1^{(0)}, \dots, x_N^{(0)}\}$ , onde  $x_i^{(k)}$  ( $k = 0, 1, 2, \dots$ ),  $k = 0$  para primeira geração e  $i$  o número de representação da população na geração  $k$ . Calcula-se  $y_i^{(k)}$  correspondente a cada indivíduo da população. Em cada par

$(x_i^{(k)}, y_i^{(k)})$  faz-se uma avaliação de acordo com a função aptidão, gerando assim uma seleção da população  $x_i^{(k+1)}$ .

#### 4.2.2.1 Configuração dos Parâmetros

É preciso tomar a decisão de como atribuir os valores para os vários parâmetros no AG, tais como o tamanho da população, a probabilidade de cruzamento, a probabilidade de mutação, etc. Esses parâmetros tipicamente interagem um com outro de maneira não-linear, de forma que não se pode otimizar um independentemente de outro [LÓPEZ-PASTOR, 2007]. Normalmente, tem-se empregado em AGs valores nas seguintes faixas: N: 10 a 200,  $P_{cru}$ : 0,5 a 1,0 e  $P_{mut}$ : 0,001 a 0,05. Na literatura, podem-se encontrar diversos valores para esses parâmetros. Por exemplo, em [DE JONG, 1975; MITCHELL, 1996] é indicado que a população deve ter de 50 a 100 indivíduos, a probabilidade de cruzamento deve ser de 0,6 e a probabilidade de mutação deve ser de 0,001. Já GOLDBERG (1989) sugere uma taxa de cruzamento superior a 60%, chegando muitas vezes a 100% e igual para todos os indivíduos da população. Segundo ainda o autor, o operador de mutação é aplicado aos indivíduos com uma probabilidade dada pela taxa de mutação que em geral varia numa faixa pequena entre 0,1% e 3%.

Diante desse problema, existe ainda muita pesquisa voltada para estratégias em que os parâmetros do algoritmo genético se adaptam durante sua execução [BACK, 2000].

Os valores dos parâmetros para a pesquisa foram escolhidos inicialmente conforme sugestão de [DE JONG, 1975; MITCHELL, 1996; GOLDBERG, 1989]. Foram testadas diversas possibilidades e investigou-se qual apresentava os melhores resultados e algumas modificações foram executadas:

- Tamanho da população: 100
- Método de seleção: Roleta
- Probabilidade de cruzamento: 0.6
- Probabilidade de mutação: 0.033

Além desses valores, optou-se por 1.000 (um mil) interações nesse experimento.

#### 4.2.2.2 Definição do Algoritmo Genético

A proposta do algoritmo genético é dada a seguir:

Cruzamentos: O operador de cruzamento tem a função de recombinar as características dos pais durante a reprodução, permitindo que as próximas gerações herdem estas características. São selecionados indivíduos da população aleatoriamente a partir da taxa de cruzamento estipulada. Estes serão recombinados e formarão descendentes. A determinação de um ponto de corte marca o ponto de troca de material genético, o gerando dois descendentes, cada um com informação genética de cada um dos pais.

Mutação: Visando melhorar a diversidade das características dos indivíduos da população, a mutação consiste na alteração do valor dos genes de um indivíduo, resultando em um novo indivíduo a partir da taxa de mutação sugerida.

Mecanismo de Seleção: O propósito da seleção é fazer com que os indivíduos mais aptos na população tenham algum tipo de prioridade na escolha para reprodução, aumentando a probabilidade de transmitir seu código genético às próximas gerações. A seleção deve ser balanceada com os operadores de cruzamento e mutação. Foi escolhido o método Roleta. É calculada a soma da aptidão de todos os indivíduos da população  $F_s$ , gerando uma aptidão total. Posteriormente, é gerado um número aleatório entre 0 e 1 e é multiplicado por  $F_s$ . É tomado, então, o indivíduo da população que terá sua aptidão adicionada à soma das aptidões de todos os membros da população anterior a este. Se o resultado é igual ou superior ao produto do número aleatório gerado entre 0 e 1 com  $F_s$ , haverá a seleção do indivíduo atual para reprodução.

Função de avaliação: Durante cada iteração, os princípios de seleção e reprodução são aplicados a população de candidatos. A função de avaliação calcula o percentual de erro de estimação de  $y_i^{(k)}$ . O indivíduo mais apto da população, ou seja, que apresente menor taxa de erro de estimação é escolhido para fazer parte da solução mínima e esta solução será usada nas demais gerações para criação de uma nova população. Vale salientar que o menor percentual de erro de estimação de  $y_i^{(k)}$  não garante uma melhor solução para  $x_i^{(k)}$  uma vez que o problema permite infinitas soluções para determinação de  $x_i^{(k)}$ .

### 4.2.2.3 Pseudocódigo do Algoritmo Genético proposto

A proposta do algoritmo genético é dada a seguir:

---

#### Algoritmo 2 Algoritmo Genético para estimação de Matriz de Tráfego

---

**Passo 0.** Segundo as recomendações de [DE JONG, 1975; MITCHELL, 1996; GOLDBERG, 1989], selecionar os valores dos parâmetros: tamanho de população método de seleção, probabilidade de cruzamento, probabilidade de mutação e número de geração.

**Passo 1.** A partir da população inicial gerada na Fase I do modelo de estimação proposto, calcular a função de avaliação para cada indivíduo com os parâmetros de  $N$  de  $x_i^{(k)}$ , onde  $i \in N$ . Definir  $k = 0$  (população inicial).

**Passo 2.** Calcular o  $y_i^{(k)}$  correspondente a cada indivíduo  $x_i^{(k)}$  da população.

**Passo 3.** Em cada par  $(x_i^{(k)}, y_i^{(k)})$  fazer uma avaliação de acordo com a função aptidão.

**Passo 4.** A seleção por Roleta (*Roulette Wheel*) da população é utilizada para selecionar os indivíduos adequados  $x_i^{(k)}$ . Executar a operação de cruzamento com probabilidade  $P_c$ . Executar a operação de mutação com probabilidade  $P_m$ . Calcular a função de avaliação dos descendentes mutados. Inserir esses descendentes mutados na nova geração. Isso é feito para evitar perder a melhor solução encontrada até agora. Depois gerar uma nova população  $x_i^{(k+1)}$ .

**Passo 5.** Se  $k =$  número máximo de gerações, o algoritmo termina. Senão,  $k = k + 1$ .

Retornar para o Passo 2.

---

## 4.3 Comentários

Este capítulo expôs no seu início algumas considerações a cerca da metodologia escolhida. O modelo de estimação proposto na pesquisa, representado graficamente seu esquema proposto das duas fases sugeridas na estimação foi descrito na seção seguinte. Ainda nessa seção uma abordagem detalhada do algoritmo de inicialização foi apresentada

especificando a idéia central do uso do modelo de gravidade na geração da população inicial, bem como uma abordagem detalhada também do algoritmo genético para estimação, foco do trabalho.

## Capítulo 5

### Ambiente Operacional e Cenário

Este capítulo trata da descrição das topologias de rede utilizadas para comprovar a eficiência computacional do método de estimação de matriz de tráfego e servirão para validar a dissertação. São dois tipos de cenários: com dados sintéticos e dados reais.

#### 5.1 Cenários

São realizados experimentos com dados sintéticos e dados reais, estes obtidos da Rede Nacional de Ensino e Pesquisa e Rede Gigabit de Pesquisa Pan-Européia.

##### 5.1.1 Dados Sintéticos

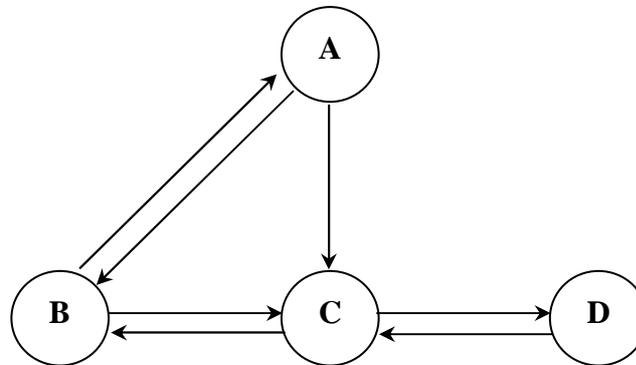
Para o uso de dados sintéticos, será considerada uma topologia pequena de 4 (quatro) nós, como apresentado na Figura 5.1. Possui 7 (sete) enlaces diretos, além de 12 (doze) pares origem-destino. Esta topologia é baseada no cenário proposto por Tebaldi e West (1998), usado da mesma forma por MEDINA et al., (2002). Todos os pares OD e seus tráfegos podem ser enumerados. Com isso, ilustra-se mais, claramente, o comportamento dos métodos.

Dados sintéticos são importantes na avaliação do desempenho das técnicas para a estimação da matriz de tráfego, ou seja, são úteis para mostrar como os métodos se comportam. Com dados sintéticos, pode-se estudar o comportamento das técnicas avaliadas com relação às matrizes de tráfego completas, ao invés das matrizes de tráfego parciais, além de permitir avaliar melhor os erros gerados pelas diferentes técnicas. [XIANG, MAKLER, SILVA, 2008]

Os dados sintéticos também permitem avaliar melhor os erros gerados pelas diferentes técnicas.

Pode-se observar na literatura, que em geral, nos estudos e nas avaliações comparativas, as matrizes de tráfego geradas sinteticamente são baseadas em hipóteses fortes que respeitam a distribuição das demandas de tráfego dos pares origem-destino [VARDI, 1996; CAO et al., 2000]. Por exemplo, MEDINA et al. (2002), em um caso, assumem que os pares OD são distribuídos de acordo com uma distribuição Poisson independente e em outro

caso, cada par OD segue uma distribuição Gaussiana. Estas distribuições apresentam determinadas propriedades que expõem os pontos fortes e fracos das técnicas avaliadas.



**Figura 5.1 – Rede com 4 (quatro) nós simples (sete enlaces)**

Nesses experimentos, serão consideradas uma matriz de tráfego constante e outra gerada com distribuição de Poisson. A MT constante está baseada nos valores usados por Tebaldi e West (1998). Em um segundo caso, a matriz de tráfego é obtida através de geração uniforme dos parâmetros de Poisson  $\lambda_i$  para cada par OD  $i$  em um intervalo [100, 500]. Então, as demandas de tráfego de Poisson são geradas de acordo com  $x_i \sim Poisson(\lambda_i)$  [MEDINA et al., 2002].

É feito também um estudo comparativo que levará em conta variações no número de enlaces presente no cenário original. Acontecerão 2 (duas) variações principais no cenário da Figura 5.1, modificando o número de conectividade: em uma situação com baixa conectividade (uma nova rede de quatro nós e apenas cinco enlaces) e outra situação com alta conectividade (uma nova rede de quatro nós e nove enlaces).

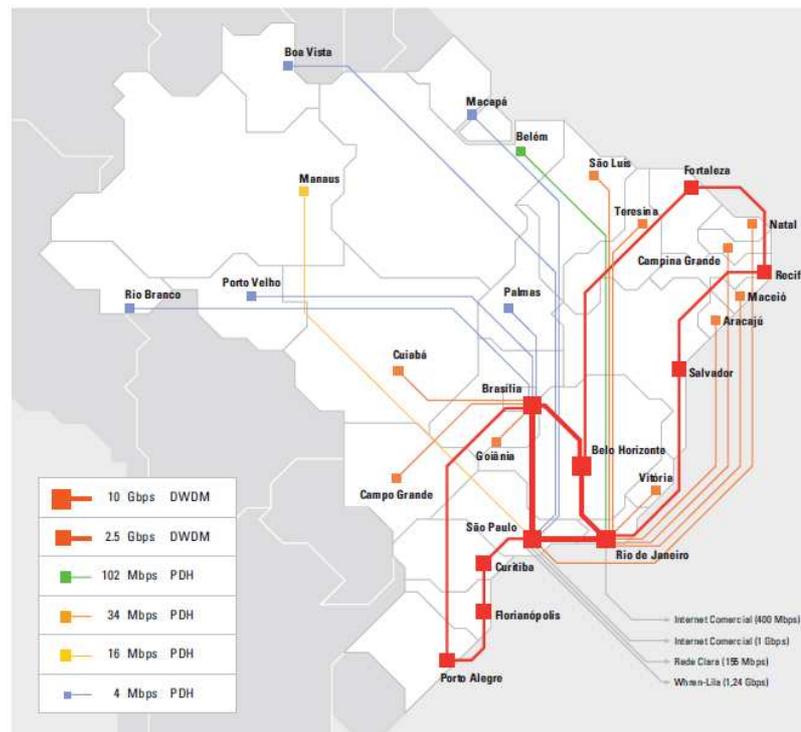
## 5.1.2 Dados Reais

### 5.1.2.1 Rede Nacional de Ensino e Pesquisa - RNP

A segunda topologia está baseada no *backbone* da Rede Nacional de Ensino e Pesquisa (RNP) (Figura 5.2), que opera uma infra-estrutura de alto desempenho com qualidade fim-a-fim. São 27 (vinte e sete) PoPs instalados em todas as capitais do Brasil, interligando cerca de 600 (seiscentas) unidades de instituições de ensino e pesquisa e algumas

iniciativas de redes regionais, principalmente redes estaduais e redes metropolitanas de ensino e pesquisa.

A RNP tem levado serviços e aplicações colaborativas de uso geral (telefonia IP, videoconferência, conferência web etc.) às instituições usuárias da rede acadêmica nacional. Também financia o desenvolvimento de aplicações inovadoras a partir do Programa Grupos de Trabalho RNP.



**Figura 5.2 – A rede Ipê - RNP**

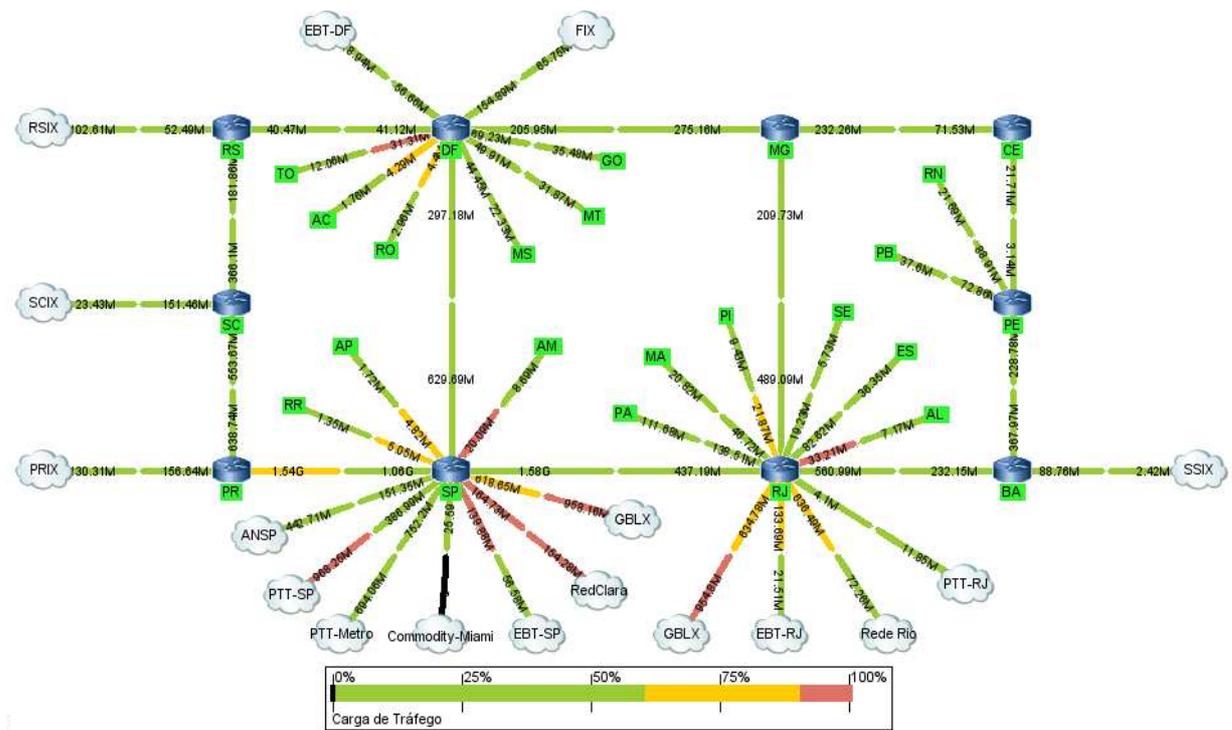
Os dados para a pesquisa da dissertação foram cedidos conforme “Acordo para a Provisão de Dados de Rede para Fins de Pesquisa”. Esclarece que os dados podem ser utilizados com o propósito de se obter uma compreensão melhor do comportamento da rede, sendo funcionais em testes ou novos desenvolvimentos de algoritmos e de instrumentos para operações de rede, monitoramento de desempenho, planejamento e segurança. Outro propósito é o desenvolvimento de diretrizes e sugestões para o uso e implementação de algoritmos novos ou existentes, e de instrumentos para os mesmos fins mencionados.

Os dados se referem a matrizes de tráfego de 119 (cento e dezenove) dias, referentes aos quatro primeiros meses do ano de 2005. Foram coletados em intervalos de 15 minutos, contados a partir de 00:30h do dia 01 de janeiro as 16:45h de 29 de abril de 2005. A

matriz de roteamento foi computada baseada em informações de topologia da rede, aplicando assim um algoritmo de roteamento para determinar o caminho mais curto entre cada par de *host* dessa rede. Já o tráfego nos canais foi obtido através do cálculo do produto da matriz de tráfego e da matriz de roteamento **A**.

Embora se trate de 27 pontos de presença instalados, optou-se por utilizar as séries temporais de matrizes de tráfego referentes apenas aos 10 (dez) PoPs. A escolha engloba um anel de 10 Gbps que liga os PoPs de São Paulo, Rio de Janeiro, Minas Gerais e Brasília e dois anéis de 2,5 Gbps que ligam os estados do Rio Grande do Sul, Santa Catarina, Paraná e Bahia, Pernambuco e Ceará.

Uma visão geral dessa topologia da rede com 10 PoPs, considerada na pesquisa, pode ser analisada na Figura 5.3.



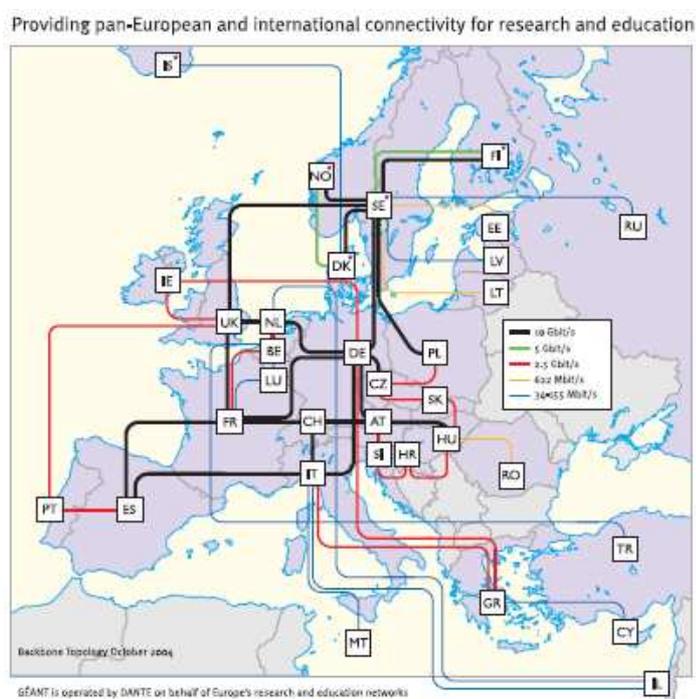
**Figura 5.3 – Visão geral - Anel de 10 Gbps e dois anéis de 2,5 Gbps - RNP**

A matriz de tráfego foi obtida adotando o procedimento do somatório das matrizes de tráfego fornecidas (por intervalos de 15 minutos), gerando uma representação de todo o tráfego no período. Com isso, é garantida a presença de tráfego entre todos os PoPs.

### 5.1.2.2 Rede Gigabit de Pesquisa Pan-Européia - Géant

A terceira topologia utilizada está baseada na Rede Gigabit de Pesquisa Pan-Européia (Géant) [UHLIG et al., 2006], uma rede de alta capacidade que engloba mais de três mil instituições de ensino e pesquisa em 30 (trinta) países, através de 26 (vinte e seis) redes nacionais e regionais de ensino e pesquisa (Figura 5.4). Além disso, a rede Géant tem 53 (cinquenta e três) ligações com outros domínios.

Atuando desde dezembro de 2001, seu objetivo principal é dar continuidade e melhorar a versão anterior da rede de pesquisa pan-européia, TEN-155.



**Figura 5.4 – Mapa Geral da GÉANT – Outubro de 2004.**

Todos os roteadores da rede Géant são roteadores de borda [UHLIG et al., 2006]. É composta por 23 (vinte e três) roteadores interconectados com 38 (trinta e oito) ligações. A rede Géant usa o protocolo intra-domínio IS-IS (*Intermediate System to Intermediate System Routing Exchange Protocol*) para calcular suas rotas. Os pesos do IGP (*Interior Gateway Protocols*) na rede GÉANT são baseados principalmente no inverso da capacidade dos enlaces com alguns aperfeiçoamentos manuais.

O *backbone* atual permite uma velocidade máxima de transmissão de 10 Gbps. Além dos nove circuitos que operam com esse enlace, há outros 11 (onze) que operam com

2,5 Gbps. Devido a esta característica e ao seu alto poder de alcance, a rede Géant é considerada a rede mais avançada do mundo no seu gênero e a que oferece maior cobertura.

As matrizes de tráfego coletadas na rede Géant referem-se a dezenas de medições em dias divididos com espaços regulares. Para a obtenção da matriz de roteamento foi considerado um algoritmo de roteamento que determinou o caminho mais curto entre cada par de *host* da rede, baseada também em informações de topologia da própria rede.

## **5.2 Comentários**

Este capítulo tratou da descrição das topologias de redes sugeridas para comprovar a eficiência computacional do método de estimação de matriz de tráfego. Levou-se em conta tanto um cenário com dados sintéticos, quanto dois cenários reais de rede.

# Capítulo 6

## Resultados e Discussões

Este capítulo apresenta os resultados obtidos nas várias simulações efetuadas para validação do modelo proposto e suas análises. A seção 6.1 aborda a metodologia feita na avaliação comparativa, base para os resultados da pesquisa. O capítulo procede com os resultados e análise das matrizes de tráfegos geradas pelo método AG, método Bayesiano e PL, usando redes de dados sintéticos com diferentes números de enlaces e diferentes distribuições. A terceira parte do capítulo apresenta uma análise dos resultados obtidos a partir da simulação com AG em redes reais, bem como uma comparação com a Inferência Bayesiana. Consideram-se duas redes de dados reais: a RNP e Géant. Os resultados são apresentados a seguir.

### 6.1 Avaliação comparativa

Uma das principais contribuições da pesquisa é o resultado da avaliação comparativa da abordagem proposta e de técnicas de estimação de matriz de tráfego existentes. A avaliação desses métodos se dá nos erros de estimativas resultantes.

A primeira técnica de estimação existente, usada na avaliação comparativa, baseia-se na Programação Linear (PL). Foi proposta por [GOLDSCMIDT, 2000]. A formulação do problema como um modelo de Programação Linear foi descrito na seção 3.4. A segunda técnica utilizada é um método baseado em inferência estatística, utilizando-se de inferência Bayesiana. O método foi proposto inicialmente em [Tebaldi e West, 1998]. A objetividade do método foi descrita na seção 3.5.

Essas técnicas simularam uma estimação de MT sob iguais condições. Foram utilizados dados sintéticos e dados reais, obtidos das redes RNP e Géant. Como já mencionado, todas as simulações consideram que suas redes executam um algoritmo de roteamento pelo caminho mais curto entre cada par de *host* da rede.

Conforme discutido na seção 5.1.1, serão usadas uma matriz de tráfego constante e outra gerada com distribuição de Poisson. A MT constante é baseada em Tebaldi e West (1998). Foram feitas, também, variações na quantidade de enlaces da rede.

Cada simulação gera resultados referentes aos erros de estimação. Esses dados são comparados e avaliados.

## 6.2 Resultados utilizando dados sintéticos

Os resultados a seguir mostram as variações apresentadas no cenário com dados sintéticos. Segundo [MEDINA et al., 2002], os métodos de PL e inferência Bayesiana terão mais sucessos na execução se os pares OD seguirem uma distribuição de Poisson. É possível analisar sua sensibilidade a essa suposição, avaliando o seu desempenho com outras distribuições. Foi sugerido o caso uniforme porque é isso que é frequentemente utilizado por pesquisadores que precisam de uma matriz de tráfego na avaliação de algoritmos de roteamento.

### 6.2.1 Topologia de quatro nós com sete enlaces

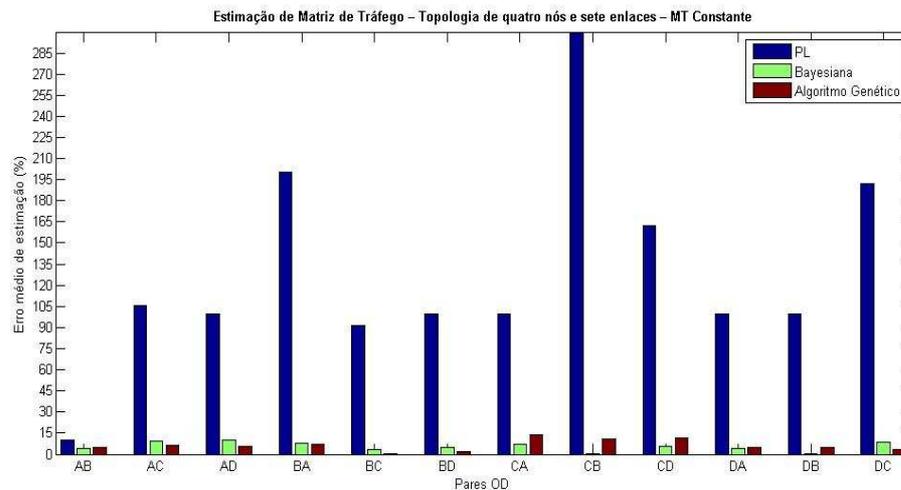
Iniciou-se a demonstração dos resultados com as simulações na rede de quatro nós e sete enlaces, utilizando uma matriz de tráfego original com dados constantes (Figura 5.1) sugerida em Tebaldi e West (1998). Os resultados apresentados a seguir mostram a eficiência do mecanismo proposto em estimação de tráfego em redes com uma topologia de poucos nós. Após várias simulações realizadas, pode-se visualizar na Tabela 6.1, a configuração adotada para todos os demais experimentos desta seção. A matriz de tráfego a priori representa a matriz estimada inicial do método proposto nessa pesquisa, gerada através de um modelo de gravidade e posteriormente usada para geração da população inicial, ou seja, primeiro passo para utilização do algoritmo genético utilizado na pesquisa. O erro médio da matriz de tráfego a priori foi de aproximadamente 56%, o que não corresponde a uma boa estimação de matriz a priori, se for considerado o tamanho da rede. Certamente que, se a própria matriz a priori fosse muito perto da solução real de estimação, o método por si só possivelmente seria adequado. Mas o que deve ser ressaltado é a separação dos valores em um espaço de busca global o que prova que a exploração do AG não fica restrita apenas a um espaço de busca local. Foram adotados os mesmos parâmetros para todas as simulações desta seção, tendo em vista os fins comparativos dos vários tipos de distribuição usados e da variação feita no número de enlaces da rede.

**Tabela 6.1 – Estimação para topologia de quatro nós e sete enlaces – MT Constante**

<i>(Constante)</i>	Original MT		<i>PL</i>		<i>Bayesiana</i>		<i>AG</i>	
	<i>MT Priori</i>	Erro (%)	MT Estimada	Erro (%)	MT Estimada	Erro (%)	MT Estimada	Erro (%)
AB: 318	109,285	65,634	287,000	9,748	305,000	4,088	302,5129	4,870
AC: 289	272,348	5,762	595,000	105,88	317,000	9,69	271,0907	6,20
AD: 312	8,308	97,337	0,000	100,00	281,000	9,94	294,1747	5,71
BA: 294	319,383	8,634	882,000	200,00	318,000	8,16	272,4831	7,32
BC: 292	396,568	35,811	559,000	91,44	282,000	3,42	292,7725	0,26
BD: 267	102,038	61,783	0,000	100,00	280,000	4,87	262,9957	1,50
CA: 305	117,202	61,573	0,000	100,00	284,000	6,89	262,5540	13,92
CB: 289	79,036	72,652	1154,000	299,31	287,000	0,69	258,6298	10,51
CD: 324	105,922	67,308	851,000	162,65	305,000	5,86	286,5139	11,57
DA: 283	69,412	75,473	0,000	100,00	295,000	4,24	269,4892	4,77
DB: 277	14,859	94,636	0,000	100,00	278,000	0,36	263,4134	4,90
DC: 291	221,249	23,969	851,000	192,44	266,000	8,59	281,7691	3,17
	<b>Erro Médio:</b>	<b>55,88</b>	<b>Erro Médio:</b>	<b>130,12</b>	<b>Erro Médio:</b>	<b>5,57</b>	<b>Erro Médio:</b>	<b>6,23</b>

Os resultados para os três métodos são apresentados na Tabela 6.1. A Tabela mostra a matriz de tráfego original, o valor estimado para cada par OD e o erro relativo. O erro médio foi de 130,12% para o método de Programação Linear, 5,57% para o método de inferência Bayesiana, e 6,23% para o método proposto na dissertação utilizando AG. O método PL é claramente pior do que os outros dois. Observe que este método atribui zero a alguns pares OD. Parece tentar combinar os pares OD ao enlace. Por exemplo, são atribuídos 595 ao par OD **AC**. Mas se sabe que o enlace **AC** também realizou o tráfego do par OD **AD**. Da mesma forma, é visto que o método PL atribuiu um valor de banda de 851 ao par OD **CD**, que coincide com **DC**. No entanto, a ligação **DC** transporta o tráfego de pares OD **DA**, **DB** e **DC**. Note que a **DA** e **DB** são atribuídos zero a fim de permitir **DC** coincidir com o enlace. Atribuir zero para alguns pares de nós implica em erro de 100% para eles, ou seja, estabelecer que alguns pares de nós combinem com o enlace, pode gerar altos erros relativos, a taxa média de erro será muito grande.

Com a abordagem proposta, o valor estimado foi muito bem aceitável. Os erros variaram entre 0,26% e 13,92% conforme apresenta a Tabela 6.1. Para esta topologia pequena utilizando tráfego constante, o método desenvolvido para a dissertação foi bem próximo ao método Bayesiano e muito melhor que o método utilizando PL, em termos de erro médio. No que diz respeito ao erro por fluxo, apresentou em alguns casos erros individuais maiores que o método Bayesiano.



**Figura 6.1 – Estimación de Matriz de Tráfego – Topologia de quatro nós e sete enlaces – MT Constante**

A Tabela 6.2 apresenta os estatísticos das taxas de erro relativo obtidos por cada estimação simulada. Os dados desta tabela mostram que: (i) A média de erro da estimação de MT através de Bayesiana e estimação de MT através de AG esta muito próximo uma da outra; (ii) em relação ao desvios-padrão, uma vez que se encontra alguns valores destoando da média, o método que usa Algoritmo Genético apresenta um leve aumento em relação ao método Bayesiano; (iii) a variância baixa do método proposto indica que a estimação não se encontra tão longe em geral dos valores esperados; (iv) para esse primeiro exemplo de rede o coeficiente de variação é sutilmente menor na estimação Bayesiana.

**Tabela 6.2 – Estatísticos da estimação para topologia de quatro nós e sete enlaces – MT Constante**

ESTATÍSTICO	<i>PL</i>	<i>Bayesiana</i>	<i>AG</i>
	MT Estimada	MT Estimada	MT Estimada
Média	1,301	0,056	0,062
Variância	0,541	0,001	0,002
Desvio-Padrão	0,736	0,032	0,040
Coef. de Variação	0,565	0,577	0,650
Mínimo	0,097	0,004	0,003
Máximo	2,993	0,099	0,139
Amplitude	2,896	0,096	0,137

Como analisado, o método de PL parece não funcionar bem para essa rede usando uma distribuição constante. A função objetiva escolhida, como sugerida em [GOLDSCHMIDT, 2000], define um peso igual ao número de saltos dos pares OD correspondente. Esses pesos podem evitar o problema de atribuição-zero se uma maximização objetiva for selecionada. Embora essa abordagem possa ter trabalhado em uma topologia de três nós considerada em [GOLDSCHMIDT, 2000], não parece ser eficaz na topologia de quatro nós usada como primeiro caso de estudo. Este mesmo método será avaliado posteriormente, usando Poisson onde MEDINA et al.(2002) relatam que este deve ser melhor executado se os pares OD seguirem a essa distribuição.

#### **6.2.1.1 Topologia de quatro nós com sete enlaces e tráfego de Poisson**

É proposto agora um novo cenário de rede onde a matriz de tráfego foi gerada a partir de uma distribuição de Poisson.

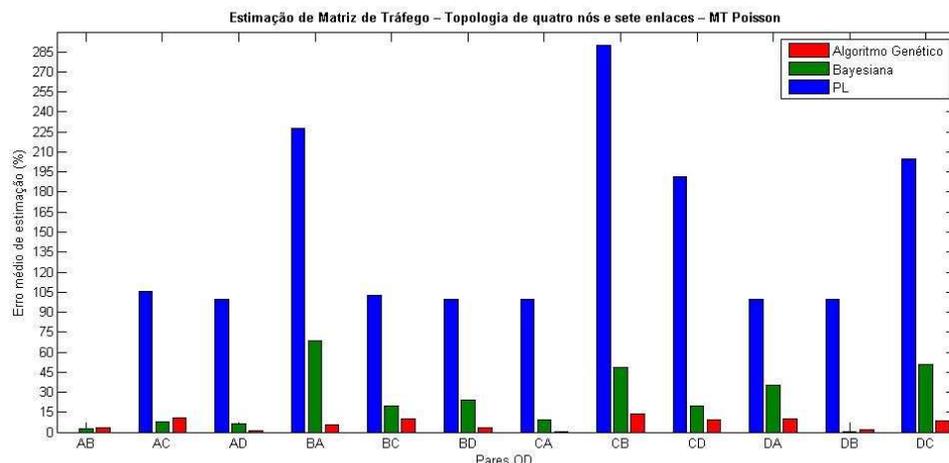
Os resultados apresentados mostram que, embora o tráfego tenha sido gerado por outra distribuição (distribuição de Poisson), o método proposto continuou produzindo baixas taxas de erro de estimação. Mais uma vez, foram realizadas várias simulações e a média das estimações é apresentada na Tabela 6.3. A matriz de tráfego a priori apresentou um erro de estimação de 76%. Sua variação no espaço de busca global continua sendo favorável a criação da população inicial usada na metodologia do trabalho.

A Programação Linear apresentou um erro médio de 135%. Os erros de estimação para o método de inferência Bayesiana e AG foram, respectivamente de 24,30% e 6,47%. Mais uma vez o método de PL é claramente pior do que os outros dois. Se comparado com a rede sintética de quatro nós, sete enlace e matriz constante, a taxa de erro final foi equivalente. A PL novamente atribui valores zero a alguns pares OD, possivelmente tentando combinar os pares OD ao enlace.

**Tabela 6.3 – Estimação para topologia de quatro nós e sete enlaces – MT Poisson**

<i>(Poisson)</i>	Original MT		PL		Bayesiana		AG	
	MT Priori	Erro (%)	MT Estimada	Erro (%)	MT Estimada	Erro (%)	MT Estimada	Erro (%)
AB: 297	615,351	107,189	297,000	0,000	289,355	2,574	287,2386	3,287
AC: 302	514,344	70,313	620,000	105,30	325,168	7,67	269,9303	10,62
AD: 318	609,639	91,710	0,000	100,00	338,652	6,49	322,1015	1,29
BA: 280	507,732	81,333	917,000	227,50	88,997	68,22	264,2449	5,63
BC: 285	410,986	44,206	578,000	102,81	340,800	19,58	257,4749	9,66
BD: 293	583,401	99,113	0,000	100,00	223,088	23,86	282,4170	3,61
CA: 288	588,306	104,273	0,000	100,00	315,365	9,50	286,5245	0,51
CB: 307	564,922	84,014	1198,000	290,23	158,504	48,37	265,0377	13,67
CD: 319	499,001	56,427	930,000	191,54	381,637	19,64	289,8878	9,13
DA: 358	474,327	32,494	0,000	100,00	482,930	34,90	320,9007	10,36
DB: 339	548,631	61,838	0,000	100,00	339,111	0,03	344,8416	1,72
DC: 305	562,216	84,333	930,000	204,92	149,987	50,82	280,0263	8,19
	<b>Erro Médio:</b>	<b>76,44</b>	<b>Erro Médio:</b>	<b>135,19</b>	<b>Erro Médio:</b>	<b>24,30</b>	<b>Erro Médio:</b>	<b>6,47</b>

Nesse novo cenário de sete enlaces, usando Poisson, o AG foi mais otimizado quanto à variação da estimação do tráfego. Para o método Bayesiano, houve um aumento considerável na taxa de erro de estimação embora presente ainda uma baixa taxa média de erro. Os erros obtidos com AG variaram entre 0,51% e 13,67%, verificados na Figura 6.2.

**Figura 6.2 – Estimação de Matriz de Tráfego – Topologia de quatro nós e sete enlaces – MT Poisson**

A Tabela 6.4 apresenta os estatísticos das taxas de erro relativo, obtidos por cada estimação para essa simulação. Os dados desta tabela mostram que as médias, variância e desvios-padrão pouco variaram se comparadas as médias da simulação anterior. Aqui a variância apresentou valor de 0,002, ou seja pode ser considerado a zero.

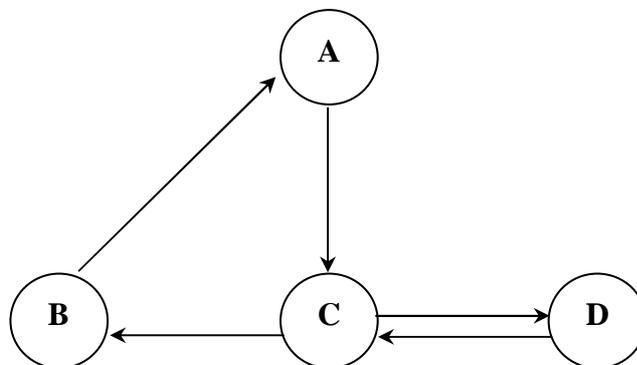
**Tabela 6.4 – Estatísticos da estimação para topologia de quatro nós e sete enlaces – MT Poisson**

ESTATÍSTICO	<i>PL</i>	<i>Bayesiana</i>	<i>AG</i>
	MT Estimada	MT Estimada	MT Estimada
Média	1,352	0,243	0,065
Variância	0,609	0,048	0,002
Desvio-Padrão	0,780	0,218	0,044
Coef. de Variação	0,577	0,898	0,672
Mínimo	0,000	0,000	0,005
Máximo	2,902	0,682	0,137
Amplitude	2,902	0,682	0,132

Embora MEDINA et al.(2002) tenham relatado que o método de PL teria melhores resultados usando Poisson, especificamente para essa simulação isso não foi percebido. A seguir mais uma variação na simulação, dessa vez modificando o número de enlaces utilizados.

### 6.2.2 Topologia de quatro nós com cinco enlaces

Para a terceira simulação com topologia de dados sintéticos foi feito uma variação da rede anterior quanto a sua topologia. O novo modelo apresenta os mesmos números de nós, no entanto apresenta apenas cinco enlaces. O objetivo é comparar o comportamento dos mesmos métodos de estimação, com as mesmas condições anteriores e constatar até que ponto o número de enlace pode influenciar em suas estimações. Essa nova rede está ilustrada na Figura 6.3. Para atender as modificações, foi gerada uma nova matriz de roteamento de acordo com o novo cenário de roteamento.



**Figura 6.3 – Rede com 4 (quatro) nós simples (cinco enlaces)**

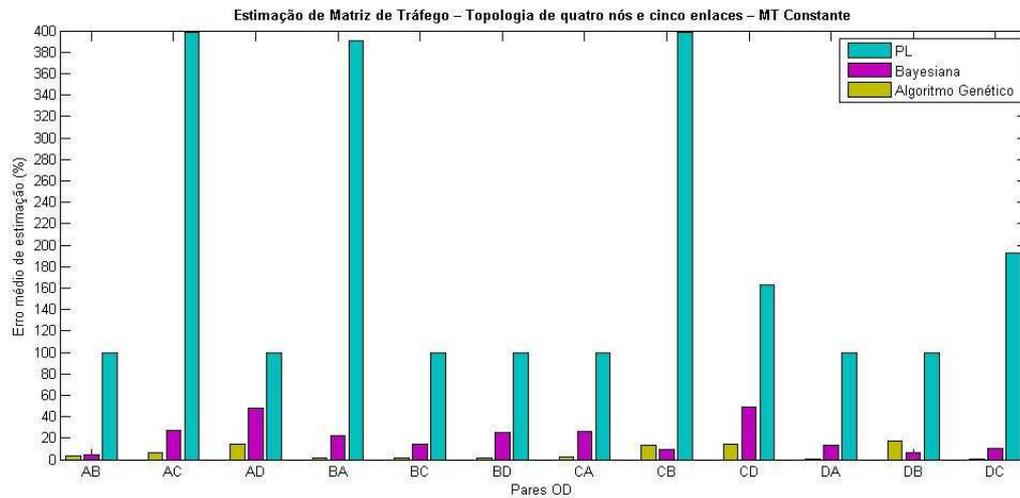
Os resultados da nova simulação são apresentados na Tabela 6.5. O método de PL, mais uma vez, apresentou resultados totalmente insatisfatórios. Acompanhando uma boa tendência, o algoritmo proposto apresenta erro médio de 6,32%. Igualmente, foram usadas várias simulações e retirado uma matriz de estimação média.

**Tabela 6.5 – Estimação para topologia de quatro nós e cinco enlaces – MT Constante**

	Original MT		PL		Bayesiana		AG	
	(Constante)	MT Priori	Erro (%)	MT Estimada	Erro (%)	MT Estimada	Erro (%)	MT Estimada
AB: 318	438,817	37,993	0,000	100,000	332,908	4,688	307,9143	3,172
AC: 289	189,285	34,50	1441,000	398,62	210,258	27,25	270,3186	6,46
AD: 312	291,022	6,72	0,000	100,00	161,119	48,36	267,9915	14,11
BA: 294	73,952	74,85	1441,000	390,14	359,928	22,42	289,7933	1,43
BC: 292	38,262	86,90	0,000	100,00	250,831	14,10	288,4672	1,21
BD: 267	43,587	83,68	0,000	100,00	334,327	25,22	261,9023	1,91
CA: 305	108,559	64,41	0,000	100,00	385,787	26,49	298,7848	2,04
CB: 289	165,711	42,66	1441,000	398,62	315,328	9,11	250,4049	13,35
CD: 324	206,080	36,39	851,000	162,65	165,806	48,83	278,0737	14,17
DA: 283	82,204	70,95	0,000	100,00	320,405	13,22	281,7413	0,44
DB: 277	81,074	70,73	0,000	100,00	294,816	6,43	229,7725	17,05
DC: 291	57,289	80,31	851,000	192,44	260,225	10,58	289,5623	0,49
	<b>Erro Médio:</b>	<b>57,51</b>	<b>Erro Médio:</b>	<b>186,87</b>	<b>Erro Médio:</b>	<b>21,39</b>	<b>Erro Médio:</b>	<b>6,32</b>

O erro médio da matriz de tráfego estimada com Programação Linear aumentou ainda mais quando se diminuiu o número de enlace. Seu alto valor de variância indica quão longe em geral os seus valores se encontram do valor esperado. O método de inferência Bayesiana apresentou um erro médio de estimação de aproximadamente 21%. O erro médio do algoritmo genético proposto foi de 6,32% próximo aos seus outros resultados. Mais uma vez o método proposto teve uma performance satisfatória. O maior erro individual foi 14%. Apresentou, ainda, mais de cinquenta por cento das estimações com valores abaixo de 4% chegando.

É interessante notar que os erros de estimação em algum dos piores casos para Bayesiano e AG correspondem aos mesmos pares OD. O comportamento dos erros de estimação é ilustrado na Figura 6.4. O método Bayesiano faz seus dois maiores erros para os pares OD **AD** e **CD**. O método AG faz tem seus maiores erros para os pares OD **AD** e **CD**, **DB**. O enlace **CB** realiza um dos maiores números de fluxos de pares OD, ou seja, cinco, entre todos os elos da pequena rede.



**Figura 6.4 – Estimação de Matriz de Tráfego – Topologia de quatro nós e cinco enlaces – MT Constante**

A Tabela 6.6 apresenta os estatísticos para esse novo caso. Aqui foi apresentada a média entre os pontos estimados igual a 0,063. A variância mais uma vez comprova a medida da dispersão estatística, indicando quão perto em geral os seus valores se encontram dos valores esperados. A variabilidade dos coeficientes é pequena, como pode ser observado através do valor da amplitude, comparada com a da estimativa Bayesiana.

**Tabela 6.6 – Estatísticos da estimação para topologia de quatro nós e cinco enlaces – MT Constante**

ESTATÍSTICO	<i>PL</i>	<i>Bayesiana</i>	<i>AG</i>
	MT Estimada	MT Estimada	MT Estimada
Média	1,869	0,214	0,063
Variância	1,677	0,022	0,004
Desvio-Padrão	1,295	0,149	0,064
Coef. de Variação	0,693	0,697	1,015
Mínimo	1,000	0,047	0,004
Máximo	3,986	0,488	0,170
Amplitude	2,986	0,441	0,166

Uma rede com poucos enlaces obriga mais fluxos passarem por um mesmo caminho (enlace) o que serve de prova na eficiência de um método, uma vez que em teoria, quanto mais pares OD no mesmo enlace, mas difícil a estimação.

### 6.2.2.1 Topologia de quatro nós com cinco enlaces e tráfego de Poisson

A mesma rede de dados sintéticos tem a sua matriz de tráfego gerada a partir de uma distribuição de Poisson. Pode-se perceber que as matrizes de tráfego a priori apresentam uma variação percentual em seu erro de estimação entre 55% e 66%, mas o mais importante é o material genético usado para a formação da população.

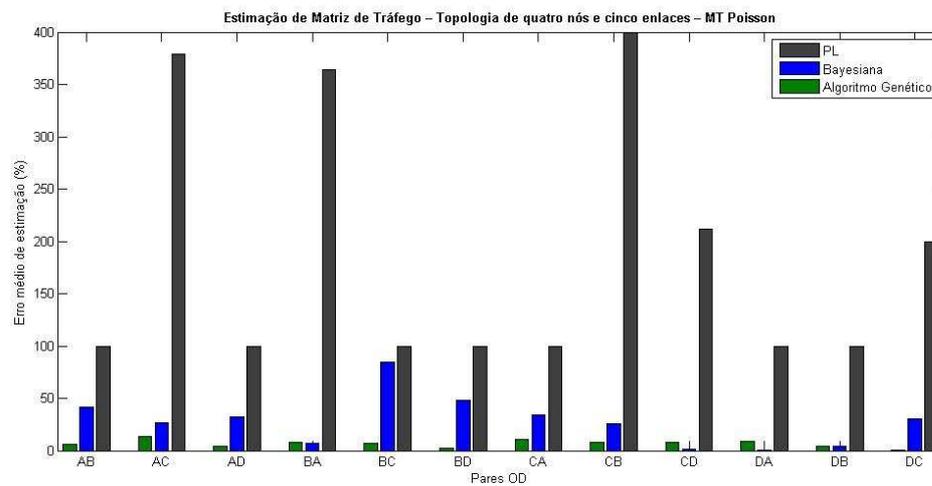
Os resultados para topologia de quatro nós com cinco enlaces e tráfego de Poisson estão apresentados na Tabela 6.7. As taxas dos três métodos usados permaneceram sem muita variação. O erro médio do método de Programação Linear foi de 188,99%, para o método de inferência Bayesiana foi de 27 % e 6 % para o método utilizando AG.

**Tabela 6.7 – Estimação para topologia de quatro nós e cinco enlaces – MT Poisson**

<i>(Poisson)</i>	Original MT		<i>PL</i>		<i>Bayesiana</i>		<i>AG</i>	
	<i>MT Priori</i>	Erro (%)	MT Estimada	Erro (%)	MT Estimada	Erro (%)	MT Estimada	Erro (%)
AB: 211	33,753	84,003	0,000	100,000	123,319	41,555	199,0564	5,660
AC: 210	20,794	90,10	1006,000	379,05	265,503	26,43	182,4408	13,12
AD: 214	173,338	19,00	0,000	100,00	283,395	32,43	205,8937	3,79
BA: 217	358,722	65,31	1006,000	363,59	231,331	6,60	200,3046	7,69
BC: 202	109,369	45,86	0,000	100,00	372,718	84,51	216,0603	6,96
BD: 223	18,835	91,55	0,000	100,00	116,159	47,91	227,6153	2,07
CA: 183	99,746	45,49	0,000	100,00	246,053	34,45	162,6270	11,13
CB: 196	95,730	51,16	1006,000	413,27	246,064	25,54	181,3404	7,48
CD: 200	137,422	31,29	624,000	212,00	202,831	1,42	184,2980	7,85
DA: 197	76,320	61,26	0,000	100,00	197,208	0,11	179,4782	8,89
DB: 219	45,926	79,03	0,000	100,00	228,004	4,11	228,8482	4,50
DC: 208	24,208	88,36	624,000	200,00	270,609	30,10	207,3601	0,31
	<b>Erro Médio:</b>	<b>62,70</b>	<b>Erro Médio:</b>	<b>188,99</b>	<b>Erro Médio:</b>	<b>27,93</b>	<b>Erro Médio:</b>	<b>6,62</b>

A estimação não foi afetada pelo fato de se ter poucos enlaces nesse exemplo. O fluxo OD **CB**, embora presente o maior número de enlaces a utilizá-lo, não é afetado, onde em toda a estimação média apresentou taxa de erro de 13 % para a pior média e 0,31 % para o caso mais otimizado.

Conforme ilustra a Figura 6.5., a variação de estimação do método proposto por par OD é mais estabilizada que as demais estimações comparativas.



**Figura 6.5 – Estimación de Matriz de Tráfego – Topologia de quatro nós e cinco enlaces – MT Poisson**

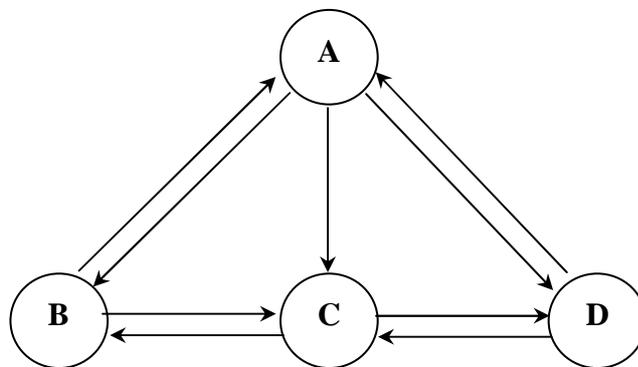
Os estatísticos das taxas de erro relativo, obtidos por cada estimação para essa simulação são da mesma forma apresentados na Tabela 6.8. Os dados desta Tabela mostram que: (i) as médias aritméticas das simulações de estimação média estão muito próximas umas das outras; (ii) os desvios padrões da rede usando tráfego do Poisson é menor que a correspondente medida no tráfego constante, indicando uma menor variabilidade do valor médio de estimação.

**Tabela 6.8 – Estatísticos da estimação para topologia de quatro nós e cinco enlaces – MT Poisson**

ESTATÍSTICO	<i>PL</i>	<i>Bayesiana</i>	<i>AG</i>
	MT Estimada	MT Estimada	MT Estimada
Média	1,890	0,279	0,066
Variância	1,573	0,058	0,001
Desvio-Padrão	1,254	0,240	0,036
Coef. de Variação	0,664	0,861	0,549
Mínimo	1,000	0,001	0,003
Máximo	4,133	0,845	0,131
Amplitude	3,133	0,844	0,128

### 6.2.3 Topologia de quatro nós com nove enlaces

Como última fase do estudo em experimentos de dados sintéticos para essa dissertação é apresentada uma nova topologia de rede com quatro nós e nove enlaces. Esta rede tem papel importante nesse método de comparação por apresentar uma maior conectividade. A Figura 6.6 ilustra a nova rede. Com uma maior quantidade de enlaces, o roteamento pode ser executado de forma a evitar vários saltos e, por consequência, aliviar em parte o problema de infinitas soluções, já que o número de canais (linha da matriz A) não se torna tão menor que o número de pares OD (coluna da matriz A).



**Figura 6.6 – Rede com 4 (quatro) nós simples (nove enlaces)**

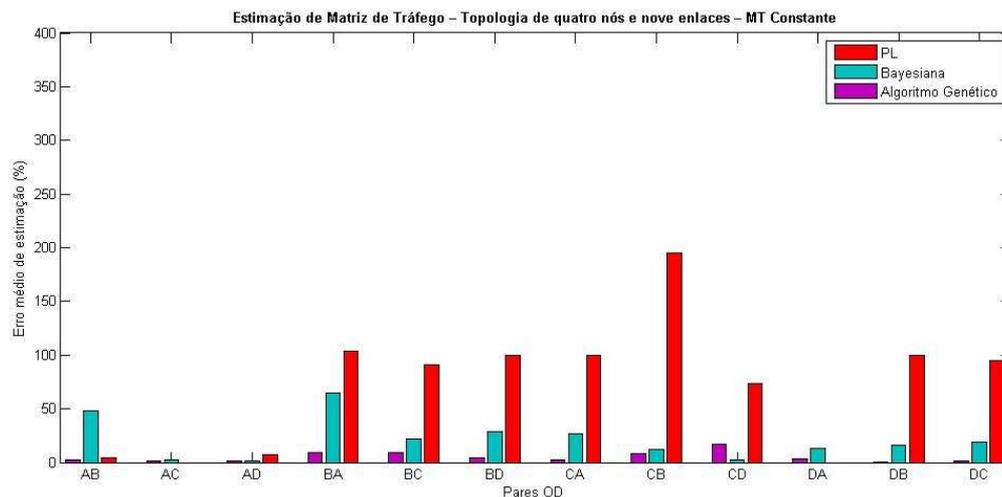
Para essa nova rede percebe-se uma melhora significativa dos métodos de estimação. No método de estimação com algoritmo genético, a estimação apresentou bom resultado. A taxa média para o erro foi de 5 %. Do total, 67% das taxas de erro nos fluxos não ultrapassaram os 5% de erro médio de estimação. O erro máximo de estimação é 17%. Isso prova o desempenho do método com essa característica de rede. Esses dados podem ser analisados na Tabela 6.9.

A PL se favoreceu com o maior número de enlaces na rede. O erro médio apresentou uma taxa de 72,49% em erro de estimação, mais ainda infere valores zero para alguns fluxos OD estimados. O método Bayesiano também apresentou bom resultado para a nova rede.

**Tabela 6.9 – Estimação para topologia de quatro nós e nove enlaces – MT Constante**

<i>(Constante)</i>	Original MT		PL		Bayesiana		AG	
	MT Priori	Erro (%)	MT Estimada	Erro (%)	MT Estimada	Erro (%)	MT Estimada	Erro (%)
AB: 318	376,018	18,245	305,160	4,038	163,368	48,626	311,5249	2,036
AC: 289	16,990	94,12	289,000	0,00	280,611	2,90	283,9019	1,76
AD: 312	11,319	96,37	287,841	7,74	308,076	1,26	316,5377	1,45
BA: 294	365,583	24,35	599,000	103,74	103,564	64,77	266,7649	9,26
BC: 292	93,954	67,82	559,000	91,44	227,341	22,14	266,4308	8,76
BD: 267	3,837	98,56	0,000	100,00	344,903	29,18	278,6241	4,35
CA: 305	129,335	57,59	0,000	100,00	387,300	26,98	298,6524	2,08
CB: 289	37,405	87,06	852,841	195,10	253,731	12,20	265,1907	8,24
CD: 324	232,814	28,14	563,160	73,81	315,869	2,51	268,9208	17,00
DA: 283	437,625	54,64	283,000	0,00	246,017	13,07	273,3044	3,43
DB: 277	187,805	32,20	0,000	100,00	232,621	16,02	277,9828	0,35
DC: 291	1,060	99,64	568,000	95,19	235,729	18,99	294,2855	1,13
	<b>Erro Médio:</b>	<b>63,23</b>	<b>Erro Médio:</b>	<b>72,59</b>	<b>Erro Médio:</b>	<b>21,56</b>	<b>Erro Médio:</b>	<b>4,99</b>

A variação em estimação dos três métodos pode ser analisada também na Figura 6.7. No método Bayesiano vimos que cerca de 90% das estimativas estão subestimadas (estimativa para um dado par OD é menor que o valor real MT). Para estas topologias pequenas, o método usando AG executa melhor do que o método Bayesiano em termos de erro da média. Isso prova a eficiência em termos médio da proposta.

**Figura 6.7 – Estimação de Matriz de Tráfego – Topologia de quatro nós e nove enlaces – MT Constante**

Percebe-se que o método proposto, utilizando AG, não é dependente da distribuição de Poisson. A variação de estimação aconteceu independente de tratar de dados

constantes ou dados gerados baseados em Poisson. O fator “número de enlaces” apresentou uma caracterização de influência maior no método.

Por último, a Tabela 6.10 apresenta igualmente os estatísticos para essa simulação. A variância está bem próxima a zero, indicando a proximidade das soluções com os dados reais. O erro médio alcançou o valor de 5% e a amplitude apresentou valor bem baixo, 16%.

**Tabela 6.10 – Estatísticos da estimação para topologia de quatro nós e nove enlaces – MT Constante**

ESTATÍSTICO	<i>PL</i>	<i>Bayesiana</i>	<i>AG</i>
	MT Estimada	MT Estimada	MT Estimada
Média	0,726	0,216	0,050
Variância	0,350	0,036	0,002
Desvio-Padrão	0,592	0,191	0,049
Coef. de Variação	0,815	0,886	0,987
Mínimo	0,000	0,013	0,004
Máximo	1,951	0,648	0,170
Amplitude	1,951	0,635	0,166

### 6.2.3.1 Topologia de quatro nós com nove enlaces e tráfego de Poisson

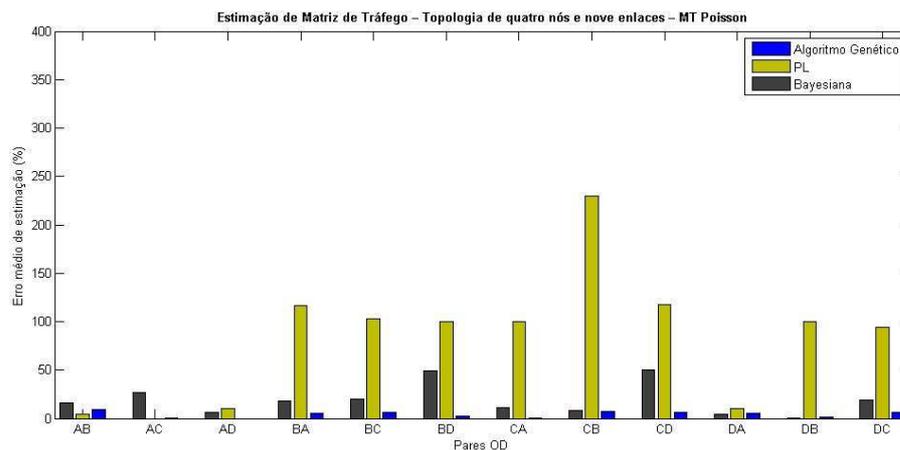
Para encerrar a seção 6.2, que trata das simulações feitas em redes de dados sintéticos, são apresentados agora os resultados obtidos a partir das simulações feitas em uma rede de quatro nós e nove enlaces e a matriz de tráfego original gerada a partir de Poisson.

O método proposto utilizando AG apresentou mais um bom desempenho, conseguindo um erro médio de estimação de apenas 4,25 %. O método Bayesiano apresentou um erro de 19,31 %. Já o método de PL apresentou pela segunda vez uma diminuição do percentual de erro de estimação, com 82, 29% ainda não sendo um valor considerável (Tabela 6.11).

**Tabela 6.11 – Estimação para topologia de quatro nós e nove enlaces – MT Poisson**

<i>(Poisson)</i>	Original MT		PL		Bayesiana		AG	
	<i>MT Priori</i>	Erro (%)	MT Estimada	Erro (%)	MT Estimada	Erro (%)	MT Estimada	Erro (%)
AB: 399	741,688	85,887	379,800	4,812	332,850	16,579	361,3157	9,445
AC: 391	729,075	86,46	391,000	0,00	286,365	26,76	390,9702	0,01
AD: 374	733,497	96,12	334,500	10,56	348,761	6,75	373,9816	0,00
BA: 355	9,146	97,42	770,000	116,90	420,443	18,43	336,2279	5,29
BC: 383	748,835	95,52	779,000	103,39	458,003	19,58	358,7194	6,34
BD: 396	486,872	22,95	0,000	100,00	199,737	49,56	386,6271	2,37
CA: 415	125,555	69,75	0,000	100,00	461,961	11,32	414,7516	0,06
CB: 355	28,860	91,87	1169,200	229,35	384,219	8,23	329,2668	7,25
CD: 370	185,907	49,75	805,500	117,70	555,991	50,27	346,5949	6,33
DA: 374	470,984	25,93	335,300	10,35	358,255	4,21	353,8115	5,40
DB: 424	101,572	76,04	0,000	100,00	420,575	0,81	415,8298	1,93
DC: 414	426,506	3,02	804,700	94,37	334,670	19,16	386,8086	6,57
	<b>Erro Médio:</b>	<b>66,73</b>	<b>Erro Médio:</b>	<b>82,29</b>	<b>Erro Médio:</b>	<b>19,31</b>	<b>Erro Médio:</b>	<b>4,25</b>

Para o último caso, o total de pares OD estimados no método AG mostraram taxas abaixo de 10% no erro da estimação. Mais uma vez pode-se constatar um bom desempenho na estimação. Pode-se notar a regularidade na estimação usando RB. A Figura 6.8 apresenta graficamente a variação na estimação dos três métodos. Completando a série, o método da pesquisa mostra valores abaixo das demais estimações e com níveis aceitáveis e recomendados a redes com poucos nós.

**Figura 6.8 – Estimação de Matriz de Tráfego – Topologia de quatro nós e nove enlaces – MT Poisson**

Nos estatísticos, mais uma vez os valores apresentaram pouca variação de média, variância e desvio padrão. A média foi de 5,4% de erro estimado, com variância de 0,001 e

um desvio-padrão de 3,2% para os erros. O Coeficiente de Variação esteve bem próximo dos métodos apresentados por PL e estimação Bayesiana. A baixa variação em termos de estimação por fluxo pode ser vista na amplitude, apresentando apenas 9,5% de amplitude.

**Tabela 6.12 – Estatísticos da estimação para topologia de quatro nós e nove enlaces – MT Poisson**

ESTATÍSTICO	<i>PL</i>	<i>Bayesiana</i>	<i>AG</i>
	MT Estimada	MT Estimada	MT Estimada
Média	0,823	0,193	0,042
Variância	0,443	0,026	0,001
Desvio-Padrão	0,665	0,161	0,032
Coef. de Variação	0,809	0,834	0,761
Mínimo	0,000	0,008	0,000
Máximo	2,294	0,503	0,094
Amplitude	2,294	0,495	0,094

Para resumir os estatísticos apresentados para cada caso, a Tabela 6.13 faz um resumo das principais características estatísticas do método baseado em algoritmo genético.

**Tabela 6.13 – Estatísticos da estimação - Resumo**

<b>Método: Algoritmo Genético</b>	<b>Média</b>	<b>Variância</b>	<b>Desvio Padrão</b>	<b>Coef. de Variação</b>	<b>Amplitude</b>
Topologia com 4N-7E - Constante	0,062	0,001	0,040	0,650	0,136
Topologia com 4N-7E - Poisson	0,064	0,001	0,043	0,672	0,131
Topologia com 4N-5E - Constante	0,063	0,004	0,064	1,015	0,166
Topologia com 4N-5E - Poisson	0,066	0,001	0,036	0,548	0,128
Topologia com 4N-9E - Constante	0,049	0,002	0,049	0,986	0,166
Topologia com 4N-9E - Poisson	0,042	0,001	0,032	0,760	0,094

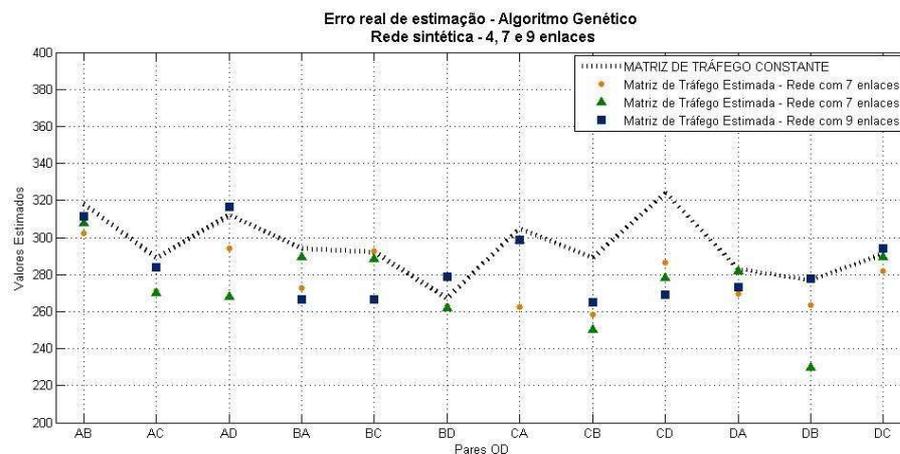
\* N – Número de nós

\*\* E – Número de enlaces

Como se pode ver, houve uma mínima variação ocorrida na média, variância, desvio padrão, coeficiente de variação e amplitude no método com AG. O fator que tem que levar em conta, em alguns casos, diz respeito ao número de simulação necessária a atingir os objetivos. A homogeneidade das estimações é maior nesse método, embora o coeficiente de variação fora a medida de dispersão que mais se assemelhou com os resultados de PL e Bayesiano.

Assim como os demais métodos, os resultados foram mais satisfatórios, utilizando tráfegos gerados a partir de uma distribuição de Poisson. A variância merece destaque, apresentando resultados entre 0,1% e 0,4% de desvios em torno da média aritmética.

Para finalizar, é apresentado na Figura 6.9, um panorama geral dos resultados das estimações com AG. A distância do ponto estimado e a linha central refletem o erro dessa estimativa. É fácil também visualizar se os erros aconteceram para mais ou para menos em relação ao tráfego original.



**Figura 6.9 - Matriz de tráfego original x erros de estimação – tráfego constante e utilizando AG**

### 6.3 Resultados utilizando dados reais

Nessa subseção serão analisados resultados obtidos com a aplicação do método baseado em AG, em duas redes de dados reais: RNP e Géant. Verificar a estimativa utilizando redes reais de *backbone* é de grande importância, principalmente para a engenharia de tráfego e suas funções. Testou-se, também, o método de inferência Bayesiana. O método de PL não foi utilizado no processo de estimativa de redes de dados reais pelo problema já constatado nas topologias pequenas estudadas anteriormente, que pela atribuição a muito pares OD do valor zero, buscam compensar isso com atribuição a vários outros pares OD com valores elevados [Medina et. al., 2002], o que o torna inviável até mesmo em testes. A atribuição do valor de carga do enlace muito pequeno e do valor de carga do enlace muito grandes (ou seja, tão grande quanto à contagem de todo o enlace) obviamente não corresponde a uma distribuição em que pode apresentar conjuntos de pares OD uniformemente distribuído na faixa de largura de uma banda.

### 6.3.1 primeiro caso: Rede Nacional de Ensino e Pesquisa - RNP

Os resultados desta seção analisam o processo de estimação da matriz de tráfego utilizando inferência Bayesiana e AG na rede da RNP. A Tabela 6.14 apresenta uma matriz de tráfego original  $X$  na rede, objetivo da estimação. Os dados de quantidade de tráfego de tráfego foram coletados nos primeiros meses do ano de 2005. Abrange apenas 10 (dez) PoPs principais da referida rede.

**Tabela 6.14 – Matriz de tráfego real (exemplo) - RNP**

	BA	CE	DF	MG	PE	PR	RJ	RS	SC	SP
BA	5,33E+06	8,85E+07	2,07E+08	2,30E+08	1,92E+07	4,24E+07	4,25E+07	7,58E+06	2,13E+07	1,82E+08
CE	1,00E+08	1,21E+07	1,15E+09	5,02E+08	6,58E+07	8,10E+07	1,59E+09	2,06E+07	2,29E+07	8,07E+07
DF	2,23E+09	2,71E+09	2,96E+09	6,28E+09	6,90E+08	5,68E+08	3,91E+09	2,52E+08	2,18E+09	2,96E+09
MG	4,00E+08	5,29E+08	1,67E+09	9,04E+07	2,77E+08	4,67E+08	3,12E+08	3,62E+07	1,29E+08	8,10E+08
PE	4,46E+07	4,23E+07	1,05E+08	5,80E+07	6,12E+03	2,48E+07	5,59E+06	9,84E+05	9,53E+06	2,42E+06
PR	6,74E+07	9,23E+07	1,21E+08	1,98E+08	8,43E+08	7,40E+05	3,86E+07	6,87E+06	2,59E+07	4,59E+07
RJ	9,76E+07	5,86E+10	4,49E+09	7,12E+08	2,41E+07	8,13E+07	1,43E+09	8,26E+06	1,16E+08	6,80E+08
RS	2,03E+06	2,44E+06	4,15E+07	1,85E+07	1,28E+06	1,88E+06	1,74E+06	0,00E+00	1,01E+07	1,14E+06
SC	1,98E+07	1,42E+07	3,09E+08	4,62E+07	5,79E+06	1,84E+07	6,09E+06	1,67E+07	9,12E+04	9,26E+06
SP	2,67E+08	8,49E+07	2,41E+09	5,47E+08	1,37E+07	4,65E+07	5,49E+08	6,53E+06	1,01E+07	4,02E+08

Os valores dos fluxos OD dessa matriz real usada como modelo de comparação para estimação, variam de **983.549** a **5.8564E+010** quantidade de tráfego.

Os estatísticos dessa base de dado são apresentados na Tabela 6.15. É possível analisar a grande variação existente em uma matriz de tráfego real. Essa inconstância na quantidade de tráfego entre origem e destino é comum em PoPs, apresentando em um dado momento fluxos muito grandes e sua grande maioria tem fluxos menores referente. Pode acontecer também o inverso. Pode-se analisar também a dispersão dos valores através de seu desvio-padrão.

**Tabela 6.15 – Estatísticos de matriz de tráfego real exemplo - RNP**

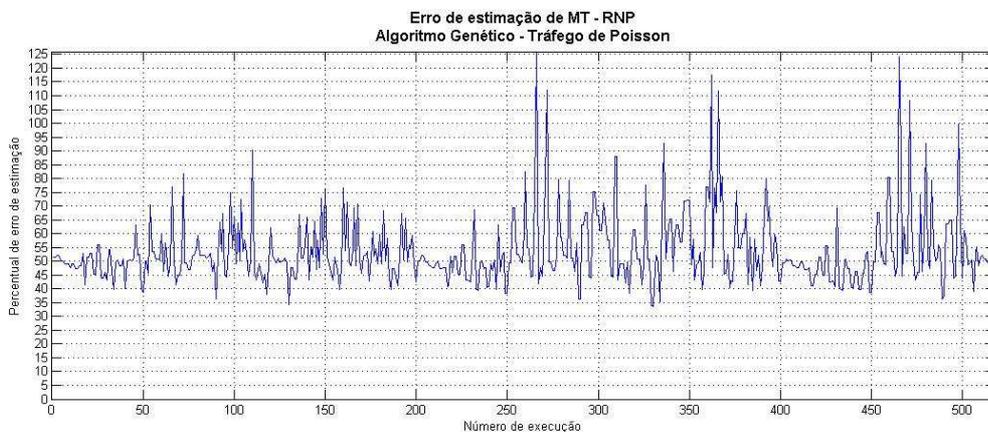
ESTATÍSTICO	MT real
Média	1,13E+09
Variância	3,85741E+19
Desvio-Padrão	6,21E+09
Mínimo	983.549
Máximo	5,86E+10
Amplitude	5,86E+10

### 6.3.1.1 Aplicação da estimação com Inferência Bayesiana e Algoritmo Genético

O modelo proposto nesse trabalho e o método de inferência Bayesiana apresentaram resultados totalmente adversos na estimação da matriz de tráfego da rede RNP, com base na matriz de tráfego real. Ambos tiveram taxas de erro elevadas demais, o que não correspondeu ao esperado. Com Algoritmo Genético, o mínimo, apresentado nas simulações, de erro médio de estimação foi de 145,94 %. Com a utilização do método Bayesiano esse erro ficou mais elevado. O alto valor da variância, para esses dados reais, tem relação direta com esses resultados. Os modelos se tornam sensíveis a grande variação de quantidade de fluxo OD.

No entanto, aproveitando esse cenário (tamanho da rede da RNP) para avaliar o método é proposto um tráfego gerado com distribuição de Poisson. Os resultados médios foram significativamente melhores. A melhor estimação para inferência Bayesiana foi de 15,1% e o erro médio para as estimatóes foi de 25,3 %. Já proposta com AG, conseguiu seu melhor resultado com 33.64 % de erro para estimação e 53.51% na média de erro.

A Figura 6.10 apresenta os resultados dessas simulações usando o algoritmo genético.



**Figura 6.10 – Estimação da MT utilizando AG e o erro de estimação  
Tráfego de Poisson – RNP**

Com o uso de um tráfego gerado com distribuição de Poisson, notou-se uma considerável melhora em termos de estimação em ambos os métodos.

### 6.3.2 segundo caso: Rede Gigabit de Pesquisa Pan-Européia - Géant

Semelhante a seção anterior é analisada a estimação de matriz de tráfego utilizando inferência Bayesiana e AG na rede Géant. A Tabela 6.16 mostra um conjunto de valores que representam um tráfego coletado na rede. A matriz de tráfego é representada por informações fornecidas para a comunidade acadêmica com o propósito de desenvolvimento da pesquisa. São 23 (vinte e três) componentes de rede, onde a grande maioria são roteadores de rede. São 506 pares de fluxo OD. Os valores do tráfego entre esses fluxos variam de 1000 a  $1.0373e+012$ . Evidentemente a rede Géant é um caso típico de rede de grande porte.

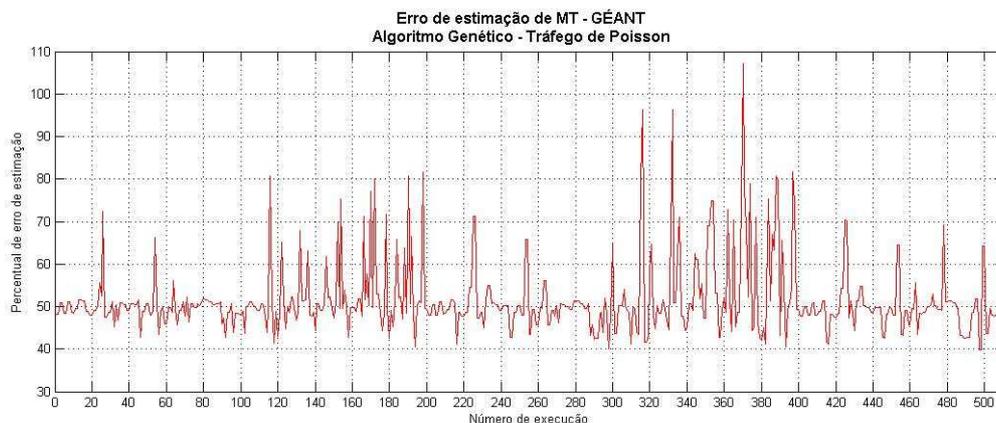
#### 6.3.2.1 Aplicação da estimação com Inferência Bayesiana e Algoritmo Genético

Mais uma vez, o algoritmo genético proposto no trabalho e o método de inferência Bayesiana apresentaram resultados desfavoráveis a utilização dos métodos com essas características de rede. As taxas de erro apresentadas superam os índices de 100% de erro de estimação, descredenciando o uso nessas condições.

Ainda semelhante à metodologia aplicada ao caso anterior (rede da RNP), optou-se por avaliar os métodos com um tráfego gerado a partir de uma distribuição de Poisson.

Os resultados médios se assemelham com os apresentados no outro caso de rede real. A melhor estimação para inferência Bayesiana foi de 19,54% e o erro médio para as estimatóes foi de 28,7%. Já com o uso de AG, o melhor resultado obtido foi 39,77 % para o erro de estimação e 51,50% para média desses erros.

A Figura 6.11 apresenta os resultados dessas simulações usando o algoritmo genético.



**Figura 6.11 – Estimação da MT utilizando AG e o erro de estimação Tráfego de Poisson – GÉANT**

Tabela 6.16 – Matriz de tráfego real (exemplo) - GÉANT

1,10E+09	5,21E+08	6,86E+07	3,03E+07	4,06E+06	5,54E+05	2,33E+08	3,90E+06	2,29E+07	5,59E+06	3,20E+06	6,22E+08	3,86E+07	6,25E+06	5,29E+06	5,29E+08	6,84E+08	7,87E+07	1,08E+07	1,02E+07	1,29E+09	1,37E+07	7,88E+05
4,41E+09	5,96E+10	5,76E+06	5,25E+10	7,28E+06	9,57E+07	1,87E+09	7,64E+06	1,29E+07	1,42E+10	6,44E+06	2,92E+11	1,67E+08	4,43E+06	2,80E+06	1,28E+07	3,87E+10	2,29E+10	1,41E+07	1,32E+07	4,47E+10	1,11E+07	3,51E+06
3,60E+10	6,54E+10	1,06E+06	2,71E+11	9,31E+10	2,30E+05	1,53E+10	3,27E+06	4,13E+08	7,40E+09	7,55E+09	3,06E+10	4,90E+10	6,10E+10	1,54E+06	4,85E+07	7,13E+11	5,65E+06	1,18E+07	1,20E+07	1,19E+11	2,73E+10	4,94E+05
1,66E+09	7,11E+11	6,43E+10	7,68E+10	1,71E+07	2,69E+06	1,45E+10	5,07E+09	1,40E+09	1,00E+03	9,96E+10	5,59E+11	3,24E+09	1,29E+11	7,78E+06	3,77E+09	8,66E+10	1,20E+07	3,65E+08	2,27E+09	5,89E+07	2,75E+09	8,03E+06
7,97E+07	8,42E+06	4,36E+05	5,43E+10	1,00E+03	1,02E+05	1,03E+06	4,30E+07	2,41E+10	6,27E+06	1,88E+06	1,58E+08	2,15E+07	1,41E+05	5,69E+06	9,71E+08	7,77E+11	3,70E+06	5,22E+05	5,21E+06	1,58E+06	1,74E+07	3,97E+05
2,42E+06	1,30E+07	7,98E+04	2,64E+05	6,47E+04	3,45E+04	5,88E+05	5,40E+04	6,42E+05	3,48E+06	3,69E+04	1,05E+07	2,91E+05	2,91E+04	5,00E+04	3,31E+06	1,15E+08	7,20E+04	2,95E+05	3,67E+04	3,40E+05	2,12E+05	2,46E+04
1,12E+08	6,82E+07	1,66E+07	1,01E+08	7,06E+06	6,02E+05	2,37E+06	2,54E+06	1,51E+07	1,20E+07	3,25E+06	8,17E+07	6,47E+07	3,18E+06	3,05E+06	9,65E+06	7,21E+07	5,51E+06	1,22E+07	1,03E+07	3,10E+07	1,62E+07	1,50E+06
3,51E+10	5,05E+06	4,40E+06	1,04E+07	4,12E+07	7,19E+04	7,83E+05	6,31E+09	9,22E+10	2,43E+06	7,57E+05	2,50E+09	1,19E+07	1,18E+05	1,04E+07	6,89E+08	2,57E+11	1,55E+06	4,24E+05	7,80E+06	4,83E+06	4,45E+06	5,39E+06
6,77E+07	1,19E+07	1,55E+06	1,84E+10	1,83E+07	5,71E+04	4,94E+06	5,99E+10	3,48E+11	6,94E+06	2,24E+06	2,44E+08	2,85E+07	4,40E+05	1,49E+07	1,27E+09	4,84E+11	1,90E+06	6,77E+06	1,19E+09	2,66E+06	2,06E+07	7,31E+05
1,78E+07	1,39E+08	5,79E+07	1,04E+03	8,85E+09	3,56E+07	6,19E+07	3,79E+08	1,39E+10	5,10E+05	8,36E+09	4,85E+08	5,18E+05	2,45E+09	4,30E+06	1,10E+08	3,34E+08	2,61E+07	9,05E+09	3,14E+07	7,95E+07	2,09E+07	1,02E+07
9,18E+07	1,85E+10	5,08E+06	4,77E+09	1,01E+07	2,67E+05	6,35E+06	4,76E+06	4,35E+07	1,09E+10	6,13E+10	6,20E+08	1,88E+09	1,16E+06	2,42E+06	1,93E+08	6,46E+11	5,36E+06	3,51E+06	7,71E+06	5,94E+06	7,40E+06	3,60E+05
1,10E+08	6,56E+10	1,66E+07	6,36E+09	3,56E+10	4,52E+07	6,21E+07	1,19E+07	4,10E+09	4,01E+07	7,00E+06	9,71E+11	2,90E+08	8,74E+06	2,09E+07	8,47E+07	4,24E+10	1,45E+11	3,70E+07	4,28E+07	1,48E+10	6,18E+07	2,06E+07
6,65E+07	3,32E+08	2,72E+07	6,36E+03	5,47E+07	1,06E+06	8,98E+07	1,87E+07	5,48E+07	6,61E+03	1,20E+07	2,54E+09	9,48E+05	1,68E+07	2,28E+09	3,89E+07	4,68E+08	2,20E+07	6,75E+07	6,62E+07	6,62E+07	6,70E+07	1,65E+07
6,72E+06	1,92E+07	4,55E+06	6,27E+06	1,96E+05	2,25E+03	1,48E+06	3,33E+05	2,01E+06	1,25E+06	4,32E+05	1,76E+07	3,18E+06	1,00E+03	3,43E+04	1,42E+07	1,90E+08	5,08E+04	2,99E+05	6,82E+04	1,57E+06	1,56E+05	7,45E+04
3,42E+06	1,59E+06	1,20E+06	8,84E+06	4,00E+06	1,74E+05	8,78E+05	2,49E+06	3,39E+07	1,61E+06	4,53E+05	1,86E+07	1,38E+07	8,09E+04	1,00E+03	1,16E+07	4,63E+07	4,33E+05	8,12E+06	7,24E+06	2,46E+06	3,39E+05	8,82E+04
9,72E+07	5,62E+06	3,03E+07	1,31E+07	9,41E+08	1,28E+05	8,03E+06	4,69E+08	9,79E+08	4,06E+06	1,04E+08	6,93E+08	1,28E+07	2,15E+08	2,99E+05	5,45E+07	8,09E+07	6,56E+05	2,76E+07	3,00E+04	5,22E+06	6,71E+05	3,57E+05
4,76E+07	1,57E+08	4,71E+07	5,66E+10	7,57E+11	2,05E+06	7,53E+07	7,39E+06	3,60E+07	5,14E+10	1,04E+12	8,26E+08	1,01E+09	3,89E+09	9,26E+06	1,56E+08	1,62E+10	1,41E+07	2,73E+07	5,35E+07	5,77E+10	3,32E+07	6,01E+06
2,21E+07	1,66E+08	9,56E+06	3,24E+09	1,87E+06	7,81E+03	3,08E+06	6,12E+05	6,09E+06	3,52E+09	2,51E+05	6,14E+09	4,10E+06	1,24E+05	1,10E+05	2,20E+07	1,14E+08	9,54E+03	5,22E+05	4,03E+06	6,71E+06	2,00E+06	1,95E+05
2,39E+07	1,46E+07	1,57E+06	8,03E+07	1,01E+06	9,07E+05	3,44E+07	1,65E+06	4,39E+06	5,20E+06	2,80E+06	1,42E+08	2,88E+07	3,73E+05	4,55E+06	5,03E+07	3,28E+07	2,76E+06	1,00E+03	2,99E+06	5,29E+06	3,05E+06	2,56E+05
5,95E+06	2,03E+07	3,21E+06	2,22E+07	2,32E+07	4,61E+04	9,94E+06	4,72E+06	2,80E+07	3,41E+06	1,06E+06	4,58E+07	3,95E+07	7,76E+05	3,46E+06	9,27E+06	6,85E+07	2,74E+06	9,18E+06	1,81E+05	1,26E+07	5,74E+04	7,08E+05
4,78E+07	2,04E+08	2,92E+07	1,03E+10	5,02E+06	1,25E+05	1,77E+07	7,56E+06	1,40E+07	8,43E+06	3,50E+06	4,25E+07	3,41E+07	1,98E+06	5,30E+06	1,92E+07	5,68E+09	1,13E+07	4,75E+06	1,00E+07	1,27E+07	2,66E+09	1,91E+06
3,98E+06	1,94E+07	1,39E+06	4,40E+07	1,89E+07	2,00E+05	7,78E+06	6,95E+06	1,76E+07	7,95E+06	1,87E+06	7,32E+07	3,25E+07	1,45E+06	2,50E+04	3,03E+06	6,92E+07	3,00E+06	4,34E+06	1,00E+03	5,63E+06	1,00E+03	9,75E+05
2,37E+06	2,83E+06	1,61E+06	1,45E+07	1,58E+06	1,28E+04	6,19E+06	4,12E+05	3,28E+06	3,42E+06	3,85E+05	1,86E+07	4,29E+06	7,56E+05	2,52E+05	3,09E+06	7,46E+07	4,12E+06	6,19E+06	2,27E+06	5,04E+06	3,36E+06	1,05E+03

Embora sejam necessárias melhorias para gerar resultados mais precisos na estimação, o algoritmo proposto apresenta bons resultados de estimação, mesmo para redes de grande porte. O gargalo está no tipo de tráfego a ser considerado. Os resultados em dados reais não são aceitáveis. Com o uso de tráfego gerado com distribuição de Poisson, o algoritmo apresenta uma estabilidade nas estimações.

Todas as simulações foram executadas na mesma máquina, com Processador DualCore Intel Core Duo T2300E, 1666 MHz, memória do sistema de 3063 MB, sistema operacional Windows XP Professional e a ferramenta de programação foi o MatLab R2008a.

#### **6.4 Comentários**

Os resultados dos testes obtidos a partir de redes com dados sintéticos trouxeram boas referências para o método proposto na pesquisa, apresentando sempre valores de estimação recomendados e aceitáveis o que contribui para a defesa da importância da pesquisa. Pode-se constatar que o método não é sensível a modelos de distribuição, o que é muito importante para não limitá-lo no uso dessas redes de pequeno porte.

No entanto, sua aplicação em redes com dados reais não apresentou resultado aceitável. O método não é apropriado à estimação de MT com dados de fluxo muito dispersos. A aplicação de um tráfego gerado a partir de uma distribuição de Poisson veio provar que, mesmo para redes grandes, é possível obter bons valores para os erros de estimação, ou seja, o problema apresentado na proposta é sensível apenas ao tipo de tráfego e não ao tamanho da rede.

# Capítulo 7

## Conclusões e Trabalhos Futuros

São feitas as considerações finais a respeito do trabalho de pesquisa. Algumas questões são levantadas e são feitas sugestões de melhoramento ou mudança de foco de pesquisa.

### 7.1 Conclusões

Neste trabalho de pesquisa, procurou-se defender a importância e relevância do estudo das técnicas de estimação de matriz de tráfego em uma rede de telecomunicação. O ponto de partida para o interesse no tema foi de que os métodos de estimação existentes não resolvem satisfatoriamente todos os casos de estimação. Limites de número de enlace, dependência de informações prévias, de tipos de distribuição são alguns exemplos de problemas enfrentados nas estimações.

Depois veio a motivação da utilização de algoritmo genético para essas estimações, que se deveu, principalmente, ao fato de inexistir, na literatura, pesquisas com essa metodologia. Outro motivador foi à flexibilidade proposta por esses algoritmos. Aliado a essa escolha, a idéia de criar um processo para inicializar a população inicial, trouxe outra técnica, o modelo gravitacional, muita usada na literatura, buscando trabalhar com um espaço de busca global, facilitando bastante a resolução do problema de estimação de MT e que contribuiu para geração e obtenção dos bons resultados.

No entanto, ficou mais claro que estimar matrizes de tráfego não se trata de um problema simples. O problema de conectividade das redes, das dificuldades em relacionar essa estimação ao tipo de roteamento são problemas que tornaram o assunto desafiador.

A pesquisa analisou dezenas de propostas de modelos para estimação, investigando as técnicas convenientes que já foram implementadas para a resolução do problema de estimação de matriz de tráfego. A especificação e proposta do algoritmo para estimação da matriz de tráfego OD utilizando a idéia de computação evolucionária foi dirigida ao ambiente de redes IP. Comparou-se a eficiência do algoritmo em relação a outras técnicas sob diferentes números de enlace.

Teve-se o cuidado de usar a metodologia em duas situações distintas: em redes de dados sintéticos e em redes de dados reais. Os dados sintéticos são importantes na avaliação do desempenho das técnicas para a estimação da matriz de tráfego e úteis para mostrar como

os métodos se comportam. Gerou-se também matrizes de tráfego sintéticas a partir de distribuições importantes, para expor os pontos fortes e fracos das técnicas avaliadas.

Após estas simulações, algumas suposições podem ser refletidas: o método de PL usado por [MEDINA et al., 2002] não é indicado para redes com muitos nós. Esse método atribui, em sua estimação, a uma grande quantidade de pares OD o valor zero, indicando não haver tráfego e isso é compensado, pelo método, com atribuição à muitos pares OD de valores elevados demais.

No caso do método Bayesiano, também apresentado por [MEDINA et al., 2002], foram obtidas significativas taxas de erro de estimação com redes pequenas, independente do método de geração do tráfego de simulação inicial: constante ou Poisson. Nas redes de topologia grande e com estimação de tráfego real, o método não apresentou bons resultados de taxas de erro na estimação. Uma das razões pelas quais a abordagem Bayesiana parece ter um desempenho pior nessas redes é que a convergência do algoritmo MCMC é estocástica. Por conseguinte, não há nenhuma garantia para fazer melhorias reais na qualidade das estimativas de uma iteração para a próxima. Já com um tráfego gerado por Poisson nessas mesmas redes as estimações alcançaram percentuais próximos aos apresentados em redes pequenas. Segundo [MEDINA et al., 2002], o método Bayesiano deve ser melhor executado se os pares OD seguirem a uma distribuição de Poisson.

O método proposto no estudo teve um ótimo desempenho em redes de topologia de quatro nós, com diferentes quantidades de conectividades (Ex.: cinco, sete e nove enlaces). Para as redes com matriz de tráfego real, o método não resolveu o problema de estimação. O algoritmo com apenas um método de avaliação parece não ser a solução para esse tipo de rede. Sabe-se que a Internet, por exemplo, tem comportamento diverso, onde alguns pares de PoP têm fluxos muito grandes e a maioria dos pares de PoP tem fluxo substancialmente menores entre eles. Um ponto de partida para o melhoramento do algoritmo é implementar uma solução multiobjetiva. Por fim, para o tráfego em grandes redes gerado por Poisson as estimações alcançaram percentuais razoáveis, carecendo de ajustes para buscar melhores resultados.

## 7.2 Trabalhos Futuros

Como trabalhos futuros sugerem-se:

(i) buscar a existência de algum fator que influencie o fluxo de um pacote através da rede, isto é, os fatores que poderiam provocar que um pacote vá de um PoP origem em particular para um PoP destino em particular.

(ii) buscar no *design* de rede algo que afete o caminho e o ponto de saída de um pacote. Podem ser investigados os protocolos de roteamento, o encaminhamento das políticas aplicadas, entre outros.

(iii) buscar um modelo de escolha que favoreça o uso da estimação em grandes redes de *backbone*, desde que se possam utilizar fatores adicionais.

## Referências Bibliográficas

- BACK, T., FOGEL, D.B., e MICHALEWICZ Z. (2000). **Evolutionary Computation 2 – Advanced Algorithms and Operators**. Institute of Physics Publishing, 2000.
- BAKER, J.E.. (1987). Reducing bias and inefficiency in the selection algorithm. In Genetic Algorithms and their Applications: Proceedings of the Second International Conference on Genetic Algorithms, pages 14–21. Erlbaum, 1987.
- CAO J., DAVIS D., VANDER WEIL S., e YU B. (2000). **Time-Varying Network Tomography**. JASA, 2000, pp. 1063-1075.
- COELHO, L. S. (2003). **Fundamentos, potencialidades e aplicações de algoritmos evolutivos**. São Carlos, SP, Brasil: Editora Sociedade Brasileira de Matemática Aplicada e Computacional (SBMAC) / Livraria Espaço, 2003. v. 1. 103 p.
- DE JONG, K.A. (1975). **An analysis of the behavior of a class of genetic adaptive systems**. PhD thesis, University of Michigan, 1975.
- FELDMANN, A., GREENBERG, A., LUND, C., REINGOLD, N., REXFORD, J., e TRUE, F. (2000). **Deriving traffic demands for operational IP networks: Methodology and experience**. IEEE/ACM Transactions on Networking, 9(3):265–279.
- GEN, M. e CHENG R. (2000). **Genetic Algorithms and Engineering Optimization**. Wiley, New York, 2000.
- GOLDBERG, D. E. (1997). **Genetic Algorithms in Search Optimization and Machine Learning**. John Wiley & Sons, United States of America, 1997.
- GOLDBERG, D.E. (1989). **Sizing populations for serial and parallel genetic algorithms**. Proceedings of third International Conference of Genetic Algorithms, pp. 70-79, 1989.
- GOLDSCHMIDT O. (2000). **ISP Backbone Traffic Inference Methods to Support Traffic Engineering**. Internet Statistics and Metrics Analysis (ISMA) Workshop, San Diego, CA, December 2000.
- HAUPT, R. L. e HAUPT, S. E. (2004). **Practical Genetic Algorithms**. Hoboken, NJ, USA: Wiley-Interscience, 2004.
- HOLLAND, J. H. (1975). **Adaptation in Natural and Artificial Systems**. University of Michigan Press, 1975.

- HOUCK, C. R., JOINES J. A. e KAY, M. G. (1995). **A Genetic Algorithm for Function Optimization: A Matlab Implementation**. Raleigh. North Carolina State University, NC, USA, NCSU-IE TR 95-09, pp. 1-14, 1995.
- KRUIHOF J. (1937), **Telefoonverkeersrekening**. *De Ingenieur*, vol. 52, no. 8, pp. E15-E25, 1937.
- KRUPP R. S. (1979). **Properties of Kruithof's projection method**. *The Bell System Technical Journal*, 58(2):517-538, February 1979.
- LOPES, S. L. (1999). **Algoritmos Genéticos em Projetos de Engenharia: Aplicações e Perspectivas Futuras**. Anais do IV Simpósio Brasileiro de Automação Inteligente, São Paulo, SP, Brasil, Set. 1999, pp. 01-11.
- LÓPEZ-PASTOR, E.T. (2007). **Algoritmo de RWA com considerações de Sobrevivência baseado em Heurística-Algoritmo Genético para Redes IP/WDM**. Tese de Doutorado em Engenharia Elétrica, Publicação PPGENE.TD - 016 A/07, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília: DF, 236p.
- MEDINA, A., TAFT, N., SALAMATIAN, K., BHATTACHARYYA, S., e DIOT, C. (2002). **Traffic matrix estimation: Existing techniques and new directions**. Em Proc. of the ACM SIGCOMM'2002, Pittsburgh, PA, EUA.
- MITCHELL, M. (1996). **An Introduction to Genetic Algorithms**. MIT Press, Massachusetts, London England, 1996.
- NI, J.; TATIKONDA, S.; YEH, E. M. (2006). **A Large-Scale Distributed Traffic Matrix Estimation Algorithm**, Global Telecommunications Conference, GLOBECOM '06. IEEE , vol., no., pp.1-5, Nov. 2006
- NUCCI A., CRUZ R., TAFT N. e DIOT C. (2004). **Design of IGP link weight changes for estimation of traffic matrices**. In Proc. IEEE INFOCOM, Hong Kong, March 2004.
- PAPAGIANNAKI, K., TAFT, N., e LAKHINA, A. (2004). **A distributed approach to measure IP traffic matrices**. Em Proc. of ACM/SIGCOMM Internet Measurement Conference – IMC 2004, Taormina, Itália.
- PAVANI, G. (2003). **Roteamento e Alocação de Comprimentos de Onda com Restrições de Potência usando Algoritmos Genéticos**. Dissertação de mestrado DECOM/FEEC/UNICAMP - Orientador: Hélio Waldman. Data da defesa: 16/09/2003.

- ROUGHAN, M., GRIFFIN, T., MAO, M., GREENBERG, A., e FREEMAN, B. (2004). **IP forwarding anomalies and improving their detection using multiple data sources**. Em Proc. of the ACM SIGCOMM'2004 Workshop on Network Troubleshooting, Portland, OR, EUA.
- ROUGHAN, M., THOURP, M. e ZHANG, Y. (2003). **Traffic engineering with estimated traffic matrices**. In *Proceedings of the USENIX/ACM Internet Measurement Conference* (Miami, USA, October 2003), pp. 248-258.
- SOULE, A., LAKHINA, A., Taft, N., PAPAGIANNAKI, K., SALAMATIAN, K., NUCCI, A., CROVELLA, M., e DIOT, C. (2005). **Traffic matrices: Balancing measurements, inference and modeling**. Em Proc. of the ACM SIGMETRICS'05, Banff, Canadá.
- SOULE, A., NUCCI, A., Cruz, R., Leonardi, E., e Taft, N. (2004). **How to identify and estimate the largest traffic matrix elements in a dynamic environment**. Em Proc. of the ACM SIGMETRICS'04, Nova Iorque, NY, EUA.
- TANOMARU, J. (1995). **Motivação, Fundamentos e Aplicações de Algoritmos Genéticos**. II Congresso de Redes Neurais, Curitiba, Outubro, 1995.
- TEBALDI C., e WEST M., (1998). **Bayesian Inference of Network Traffic Using Link Count Data**. JASA, 1998, pp. 557-573.
- TEIXEIRA, R., DUFFIELD, N., REXFORD, J., e ROUGHAN, M. (2005). **Traffic matrix reloaded: Impact of routing changes**. Em Proc. of the Passive and Active Measurement Workshop - PAM'2005, Boston, MA, EUA.
- UHLIG S., QUOITIN B., LEPROPRE J. e BALON S. (2006). **Providing public intradomain traffic matrices to the research community**. ACM SIGCOMM Computer Communication Review, v.36 n.1, January 2006.
- VARDI, Y. (1996). **Network tomography: Estimating source-destination traffic intensities from link data**. Journal of the American Statistical Association, 91, 365-377, 1996.
- VATON S. e GRAVEY A (2003). **Network tomography: an iterative bayesian analysis**. In Proc. ITC 18, Berlin, Germany, August 2003.
- WEILE, D. S. e MICHIELSSEN, E. (1997). **Genetic Algorithm Optimization Applied to Electromagnetics: A Review**. IEEE Trans. on Antennas and Propagation, vol. 45, no. 3, pp. 343-353, Mar. 1997.

XIANG, Z., MAKLER, S. S., SILVA, E. S. (2008). **New Solution Techniques for the Traffic Matrix Estimation Problem**, in Simpósio Brasileiro de Redes de Computadores, Anais do XXVI Simpósio Brasileiro de Redes de Computadores, 2008, Rio de Janeiro, Rio de Janeiro

ZHANG, Y., ROUGHAN, M., DUFFIELD, N., e GREENBERG, A. (2003a). **Fast accurate computation of large-scale IP traffic matrices from link loads**. Em Proc. of the ACM SIGMETRICS' 03, San Diego, CA, EUA.

ZHANG, Y., ROUGHAN, M., LUND, C., e DONOHO, D. (2003b). **An information-theoretic approach to traffic matrix estimation**. Em Proc. of the ACM SIGCOMM'2003, Karlsruhe, Alemanha.