

Relatório Técnico – Projeto “Redes Bayesianas”

- **Requisitos:** Os requisitos de *software* e *hardware* utilizados para o desenvolvimento do projeto foram, respectivamente: Sistema Operacional Windows 10 Pro x64 (2015), 4GB de memória RAM e processador Intel(R) Core(TM) i5-3230M CPU @ 2.60GHz 2.60GHz.
- **Interface gráfica:** O projeto foi desenvolvido utilizando o *Eclipse Java EE IDE for Web Developers*, versão 4.5.2. (2016). Para implementar a interface gráfica, foi utilizada a extensão *WindowBuilder*, disponível especialmente para o *Eclipse*. A **Figura 1** ilustra a interface gráfica da ferramenta *Bayesian Networks – Tuberculosis Problem*. Por se tratar da resolução de um problema baseado em uma rede bayesiana específica (ilustrada na aba Representação da rede bayesiana), a ferramenta permite realizar qualquer inferência unicamente para a rede descrita em [1]. Dessa forma, na aba **Network** o usuário pode visualizar a representação pictórica da rede, conferir as tabelas de probabilidade condicionais na aba **Tables**, ou realizar consultas de inferência sobre a rede, na aba **Inference**.

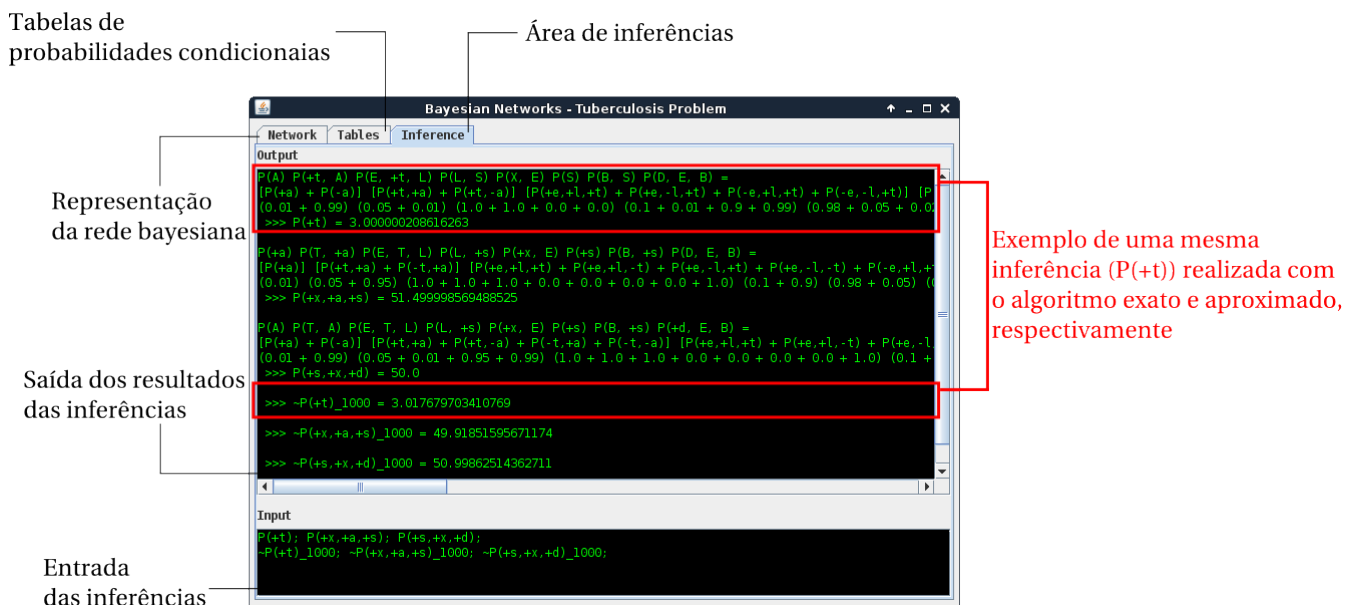


Figura 1: Interface gráfica da ferramenta *Bayesian Networks – Tuberculosis Problem*.

- **Funcionamento:** A principal utilidade da ferramenta desenvolvida é fornecer um meio de realizar inferências sobre a rede bayesiana e comparar os resultados de dois métodos diferentes: um exato (Enumeração) e um aproximado (*Likelihood Weighting*). Para o caso de inferências utilizando o método exato, a entrada deve seguir o padrão “ $P(<evidências>)$ ”, onde “ $<evidências>$ ” deve ser a lista de evidências (separadas por vírgula (,)) e indicadas com os respectivos valores das variáveis. Por exemplo, para saber a probabilidade de uma pessoa ser fumante (+s) dado o resultado positivo no Raio-X (+x) e ter dispnéia (+d), a inferência ficaria “ $P(+s,+x,+d)$ ”. Note que a probabilidade alvo (nesse caso, a probabilidade positiva de um fumante (+s)) deve ser sempre a primeira evidência da inferência realizada. Para o caso do método aproximado, a mesma inferência ficaria “ $\sim P(+s,+x,+d)_{1000}$ ”, por exemplo. Nesse caso, o método aproximado é diferenciado por um acento til (~), simbolizando aproximação, e pelo número de amostras (no caso, 1000) a serem realizadas durante a execução do método, que deve ser pelo menos 1. Também é possível realizar múltiplas inferências de uma só vez, separando as entradas por ponto e vírgula (;). Na **Figura**

1, o resultado de uma mesma inferência utilizando o método exato e aproximado está destacado na cor vermelho, respectivamente. É possível observar que para o método exato muitos cálculos são realizados para se obter a probabilidade final, tratando-se de um somatório de combinações de variáveis ocultas (que não são evidências de entrada da inferência). Já para o método aproximado, por se tratar de uma contagem de amostras aleatórias, apenas o resultado final (todas as probabilidade finais são precedidas do símbolo “>>>”) é exibido.

- **Implementação:** A ferramenta foi desenvolvida utilizando a linguagem Java. Seu código fonte possui 5 *packages*: gui (contém a classe de interface gráfica *MainWindow.java*), main (contém a classe inicializadora da ferramenta, *Main.java*), variable (contém as classes estruturais da rede bayesiana, *Variable.java*, *Connection.java*, *Evidence.java*, *ConditionalProbability.java* e *Result.java*), inference (contém a classe dos algoritmos de inferência, *Inference.java*) e images (contém a imagem representativa da rede bayesiana para este problema). Em ordem de execução, a classe *Main* instancia a classe *MainWindow*, que por sua vez faz a comunicação entre o usuário e a classe de inferência, que se comunica com a estrutura da rede e a classe *Result*. A rede bayesiana é formada por uma lista de variáveis (*Variable*) que possuem uma lista de conexões (*Connection*) de entrada (variáveis que influenciam essa variável) e de saída (variáveis que são influenciadas por essa variável). Como cada variável também possui uma lista de probabilidades condicionais (*ConditionalProbability*), é possível caminhar por todas as variáveis obedecendo as regras de ligações (conexões) da rede.
- **Testes realizados:** A avaliação da ferramenta foi realizada com o objetivo de comparar os diferentes métodos de inferência de probabilidade, considerando a influência do número de amostras para a precisão (diferença para o valor exato) da probabilidade. Para isso, foram realizados testes com o número de amostras igual a 1, 10, 50, 100, 1000 e 1000000, e foram comparados os erros de cada teste em relação à probabilidade exata da inferência $P(+t)$. Esse conjunto de testes foi processado uma única vez para cada abordagem diferente. A seguir, os erros de execução para cada algoritmo são exibidos no **Gráfico 1**.

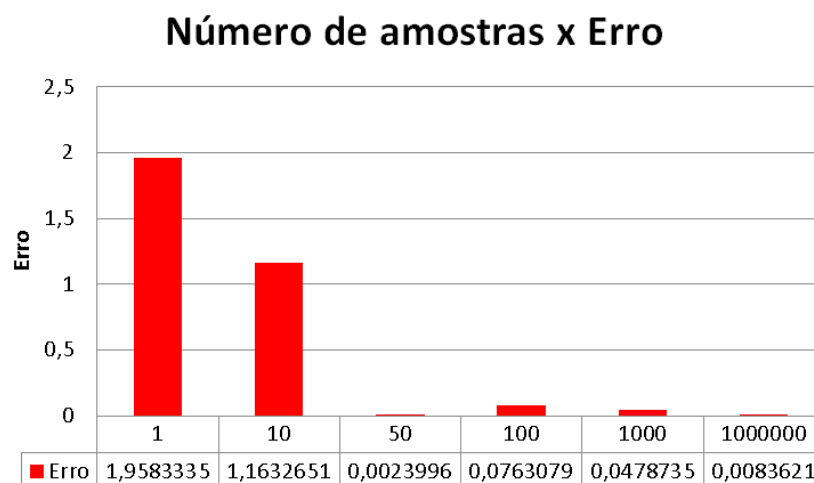


Gráfico 1: Relação do erro com base no número de amostras.

- **Conclusões:** O **Gráfico 1** permite concluir que conforme o número de amostras aumenta, o erro da probabilidade indicada pelo método aproximado diminui, tornando-se quase irrelevante. Baseado nisso e no fato do método aproximado demandar muito esforço computacional para redes complexas, o método aproximado torna-se uma interessante alternativa de método de inferência, por ser mais simples de

implementar, ser menos custoso computacionalmente (relativo à quantidade de amostras) e fornecer resultados bem próximos dos exatos.

Referências Bibliográficas:

[1] Lauritzen S., Spiegelhalter D., (1988). "***Local Computation with Probabilities on Graphical Structures and their Application to Expert Systems (with discussion)***". Journal of the Royal Statistical Society: Series B (Statistical Methodology), 50(2), 157-224.