

# Estudo e aplicação de medidas de similaridade semântica na visualização das interações gênicas em Esquizofrenia

Lulu Wu, Roberto Carlos Lopes da Costa-Filho, Evandro Eduardo Seron Ruiz

Departamento de Computação e Matemática, FFCLRP – USP  
{lulitawu, costafilhorcl}@gmail.com, evandro@usp.br

Atualmente existe uma grande tendência para a integração de dados com o objetivo de selecionar genes candidatos relacionados a doenças complexas, como a Esquizofrenia (EZ). As interações dos genes responsáveis pela EZ podem ser representadas através de redes gênicas, com nós e arestas semanticamente definidos. Métodos de análise e visualização de redes complexas são usados frequentemente como abordagens de investigação em redes gênicas. No entanto, os sistemas de análise e visualização de redes em Biologia apresentam baixos índices de especialização e muitas representações gráficas não respeitam as convenções biológicas. Neste projeto avaliamos os genes responsáveis pela EZ. Nossa tese é que a comparação de anotações dentro de uma ontologia, conhecida como “similaridade semântica” (SS), pode avaliar como duas entidades são relacionadas [Pesquita *et al.*, 2009]. A SS é basicamente a distância entre dois termos ao seu ancestral comum. Este ancestral refere-se ao termo com maior conteúdo informacional (CI), ou seja, *mica*. Assim  $SS(g_1, g_2) = CI(mica)$ . Neste caso estudamos o relacionamento entre genes usando termos da Gene Ontology (GO). As medidas de SS entre genes relacionados à EZ auxiliam a conceitualização de agrupamentos gênicos com funções biológicas similares. Utilizamos duas medidas de SS, uma proposta por Resnik [Resnik *et al.*, 1995] e outra proposta por Lin [Lin *et al.*, 1998]. As medidas de SS são calculadas por um software desenvolvido localmente, como descrito em [Costa-Filho *et al.*, 2010]. Um exemplo de caso de uso para este software: dada uma lista de genes como dados de entrada, obtemos uma matriz de distância usando SS entre todos os genes da lista. Elaboramos uma metodologia de visualização de agrupamentos gênicos utilizando as duas medidas de SS citadas. Utilizamos da técnica de *Multidimensional Scaling* (MDS) para a projeção bidimensional das medidas de similaridade entre genes, assim podemos visualizar a distância semântica entre os genes da EZ. Acreditamos que esta metodologia pode auxiliar a avaliação entre causa e efeito dos genes em relação a doenças como a EZ.

**Agradecimento:** L. Wu e R. C. L. Costa-Filho agradecem FAPESP e PIBIT/CNPq respectivamente.

## Referências :

- Pesquita C, Faria D, Falcão AO, Lord P, Couto FM. Semantic Similarity in biomedical ontologies. *PLoS Computational Biology*. 2009 Jul;5(7):e1000443+.
- Resnik P. Using information content to evaluate semantic similarity in a taxonomy. *In International Joint Conference on Artificial Intelligence*. 1995.
- Lin D. An information-theoretic definition of similarity. In: *Proc. of the 15th International Conference on Machine Learning*. 1998; San Francisco, CA: Morgan Kaufmann. 296–304.
- Costa-Filho, RCL e Ruiz, EES Uso de Ontologias para identificação de associações gênicas e priorização de genes nas relações gene-doença. CBIS'2010.