



Revista Eletrônica  
Paulista de Matemática

ISSN 2316-9664  
Volume 14, fev, 2019  
Edição Ermac

**Bruno Afonso Corrêa**

UNESP – Universidade Estadual Paulista “Júlio de Mesquita Filho”  
brunoafonso12@hotmail.com

**Giovana Lanza Okada**

UNESP – Universidade Estadual Paulista “Júlio de Mesquita Filho”  
gigilanzaokada@gmail.com

**Leonardo Mantelle Mello Maldonado**

UNESP – Universidade Estadual Paulista “Júlio de Mesquita Filho”  
leomantellemaldonado@gmail.com

**Ivan Rodrigo Wolf**

UNESP – Universidade Estadual Paulista “Júlio de Mesquita Filho”  
ivan.wolf@gmail.com

**Guilherme Targino Valente**

UNESP – Universidade Estadual Paulista “Júlio de Mesquita Filho”  
valentegt@fca.unesp.br

**Rafael Plana Simões**

UNESP – Universidade Estadual Paulista “Júlio de Mesquita Filho”  
rafael.simoaes@unesp.br

## Geração de redes de co-expressão de genes utilizando informações de expressão diferencial

Generation of co-expression gene networks using differential expression information

### Resumo

As redes de co-expressão de genes são aplicadas para análise dos processos biológicos. O método de correlação simples é amplamente empregado para a criação de redes, porém não é capaz de modelar adequadamente sua complexidade. O presente estudo propõe um método de geração de redes de co-expressão utilizando dados de expressão diferencial de genes. Para isso, foram obtidos de banco de dados públicos informações de expressão de genes para as espécies *Saccharomyces cerevisiae* e *Homo sapiens*. As expressões diferenciais foram calculadas utilizando o pacote DESeq2 do *software* R. A análise topológica da rede foi realizada utilizando o *software* MatLab e as redes geradas foram comparadas com redes validadas experimentalmente. Os resultados revelaram que o método proposto é mais eficiente que a correlação para a geração de redes para a espécie *Saccharomyces cerevisiae*, contudo estudos mais aprofundados ainda são necessários para uso da metodologia para dados da espécie *Homo sapiens*.

**Palavras-chave:** Redes de co-expressão de genes, correlação simples, expressão diferencial de genes.

### Abstract

Co-expression gene network are applied to analyze biological processes. The simple correlation method is widely used to create this kind of network, however it is not able of properly modeling its complexity. Therefore, this present study aims to propose a new method to generate gene co-expression networks by an approach applying gene differential expression. For this purpose, gene expression data was obtained from public data base for the species *Saccharomyces cerevisiae* and *Homo sapiens*. Differential expression was calculated by DESeq2 package for R software. Network topologic analyses were performed using MatLab software. The constructed network was compared with experimental validated networks. The results revealed the proposed method is more efficient than the correlation for generating networks for the *Saccharomyces cerevisiae*, but further studies are still needed to use the methodology for data from *Homo sapiens*.

**Keywords:** Co-expression gene networks, simple correlation, genes differential expression.



# 1 Introdução

A evolução dos recursos tecnológicos permitiu um grande avanço na obtenção de informações de sistemas biológicos. Compreender como essas informações podem estar associadas, ou seja, a função e a dinâmica dos processos biológicos de metabolismo, regulação e sinalização e suas interações em nível celular é crucial para fazer progressos na biologia, medicina e biotecnologia (ROSA; ROCHA; FURLAN, 2007). Entretanto, pelo fato desses dados de sistemas biológicos serem complexos, é requerido o desenvolvimento de métodos baseados em princípios matemáticos e computacionais para integrar esse grande volume de informações (LI et al, 2015). A biologia de sistemas surgiu nesse contexto, sendo um campo de estudos interdisciplinar (KITANO, 2001). Com o advento de biotecnologias como o microarranjo de DNA e sequenciamento de nova geração, o uso da expressão gênica passou a ser indispensável nos estudos biológicos (THOMAS; PORTIER, 2013). Uma maneira eficiente de se realizar interpretações biológicas significativas dos dados de expressão é a partir da construção de redes de co-expressão gênica, uma importante ferramenta da Biologia de Sistemas.

Muitas abordagens estatísticas para inferência de redes de co-expressão de genes foram desenvolvidas (ALLEN et al, 2012). Os métodos baseados em correlação simples são majoritariamente utilizados, especialmente por sua baixa complexidade numérica (CARTER et al, 2004; LANGFELDER; HORVATH, 2008; ZHANG; HORVATH, 2005). A literatura mostra que esse método produz redes com características topológicas adequadas, na qual o grau de distribuição segue a lei de potência. Entretanto, observa-se que as taxas de arestas verdadeiras positivas são baixas e, quando altas, refletem apenas o fato do método ter sido aplicado sobre uma rede de dimensão imensamente inferior às redes reais (RUAN; DEAN; ZHANG, 2010; ZHANG; HORVATH, 2005). Isso ocorre porque a correlação simples considera apenas correlação de expressões entre pares isolados de genes em uma rede, excluindo a possível influência de um terceiro gene sobre cada par. Logo, ocorre o aumento de arestas falsas positivas pela utilização de tal método. Por isso, redes biológicas tendem a ser extremamente complexas, assim como as informações, expressões e regulações entre os nós.

Nesse sentido, o presente estudo propõe uma metodologia baseada no uso de expressão diferencial de genes para a geração de redes de co-expressão. Essa abordagem tende a ser mais eficiente pois as informações de expressão diferencial são obtidas de uma análise multivariada da expressão dos genes, modelando de forma mais fidedigna a complexidade das redes biológicas (LOVE; HUBER; ANDERS, 2014).

O objetivo do presente estudo é inferir redes de co-expressão de genes para as espécies *Saccharomyces cerevisiae* e *Homo sapiens* utilizando o método proposto e comparar sua performance preditiva com a do método de correlação simples.

## 2 Métodos

### 2.1 Obtenção dos dados de expressão de genes

Para o presente estudo foram utilizados dados genômicos de duas espécies modelos. O primeiro organismo modelo foi a espécie *Saccharomyces cerevisiae*, uma levedura fermentativa. Essa espécie foi selecionada por ser um dos mais populares modelos para diversos experimentos em Biologia e, sendo assim, seu genoma está com melhor qualidade de montagem e anotação. Também há para essa espécie milhares de dados de transcriptoma disponíveis. Os



transcriptomas foram baixados do banco de dados *Sequence Read Archive* (SRA) do *National Center for Biotechnology Information* (NCBI) e o genoma de referência coletado do banco de dados *Ensemble* (ZERBINO et al, 2018).

Os *reads* filtrados foram mapeados sobre seu respectivo genoma utilizando o programa *Bowtie 2* e as expressões em *Rlog* foram calculadas em usando o pacote *DESeq2* do *software R*. Dessa forma, foram obtidos os dados de expressão gênica de todos os genes ( $n_{genes} = 6703$ ) da espécie *Saccharomyces cerevisiae* em 36 experimentos diferentes ( $n_{experimentos} = 36$ ).

O segundo organismo modelo estudado foi a espécie *Homo sapiens*, de forma a incrementar os resultados apresentados no trabalho publicado no V Encontro Regional de Matemática Aplicada (CORRÊA et al, 2018). Dessa maneira, de forma adicional, neste estudo foram analisados dados de expressão gênica de cardiomiócitos da espécie *Homo sapiens* publicados por Pavlovic et al (2018), obtidos diretamente do banco de dados *GEO DataSets* (NCBI). Dos 44125 genes, foram selecionados para o estudo dados de expressão de apenas 21607 genes ( $n_{genes} = 21607$ ), por ser apenas esse o conjunto de genes presentes em redes de co-expressão validadas experimentalmente. Para esse conjunto de genes foram feitas duas análises: uma considerando um total de 10 experimentos ( $n_{experimentos} = 10$ ) e outra analisando 36 experimentos ( $n_{experimentos} = 36$ ). O objetivo da escolha de dois conjuntos experimentais de diferentes tamanhos é avaliar se o número de experimentos tem interferência na performance preditiva das redes geradas.

## 2.2 Geração da matriz de expressão diferencial

A determinação da expressão diferencial dos genes de ambas as espécies foi realizada utilizando o pacote *DESeq2*. A análise de expressão diferencial no *DESeq2* usa um modelo linear generalizado da forma:

$$K_{ij} \approx NB(\mu_{ij}, \alpha_i). \quad (1)$$

$$\mu_{ij} = s_j q_{ij}. \quad (2)$$

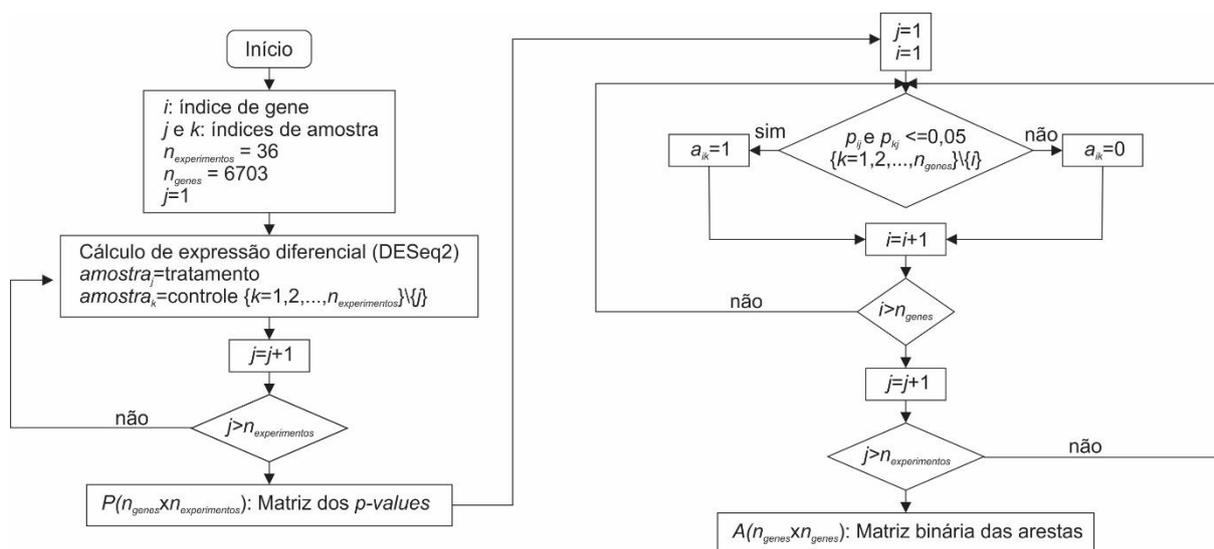
onde, na Equação 1,  $i$  corresponde ao índice do gene e  $j$  corresponde ao índice do experimento,  $K_{ij}$  é o número de *counts* do gene  $i$  em uma amostra  $j$ , a qual é modelado utilizando uma distribuição binomial negativa com média  $\mu_{ij}$  e um parâmetro de dispersão  $\alpha_i$ . A Equação 2 mostra que a média  $\mu_{ij}$  é composta de um fator de tamanho específico da amostra  $s_j$  e um parâmetro  $q_{ij}$  proporcional à concentração real esperada de fragmentos de um gene  $i$  para a amostra  $j$  (LOVE; HUBER; ANDERS, 2014).

## 2.3 Matriz de adjacências

Para determinar a significância estatística (*p-value*) da expressão diferencial de todos os genes  $i$  em dado experimento  $j$ , todos os demais experimentos  $k$  (sendo  $k = \{1, 2, \dots, n_{experimentos}\} \setminus \{j\}$ ) foram considerados como condição controle. Dessa maneira foi possível obter uma matriz de probabilidades,  $P(n_{genes} \times n_{experimentos})$ , onde cada linha corresponde a um gene  $i$  e cada coluna a um experimento  $j$ . A partir dessa matriz, foi gerada uma matriz de adjacências da rede,  $A(n_{genes} \times n_{genes})$ , onde cada elemento tem um valor binário. Sendo assim, em um dado experimento  $j$  os elementos da matriz são definidos da seguinte forma:

$$\begin{cases} a_{ik} = 0 \leftrightarrow p_{ij} \leq 0,05 \wedge p_{kj} \leq 0,05 \\ a_{ij} = 0 \leftrightarrow p_{ij} > 0,05 \vee p_{kj} > 0,05 \end{cases} \quad (3)$$

A Equação 3 mostra que são considerados como genes co-expressos aqueles com expressão diferencial estatisticamente significativa ( $p \leq 0,05$ ) simultaneamente em um ou mais experimentos. O fluxograma apresentado na Figura 1 ilustra o método para a geração de redes de co-expressão.



**Figura 1** - Fluxograma ilustrativo do método para a geração de redes de co-expressão utilizando informações da expressão diferencial de genes.

## 2.4 Representações gráficas das redes e histogramas de graus de conectividade dos nós

O *software* MatLab foi utilizado para produzir a representação gráfica das redes inferidas. A representação gráfica da rede para a espécie *Saccharomyces cerevisiae* levou em consideração todos os seus 6703 genes. Já as representações gráficas da rede de co-expressão de genes da espécie *Homo sapiens* foram produzidas utilizando uma seleção aleatória de 6703 genes para fins de comparação com os resultados para a espécie *Saccharomyces cerevisiae*. Todas as representações gráficas das redes foram produzidas utilizando a função *Graph* do *software* MatLab.

Ainda com o auxílio do *software* MatLab, determinou-se os histogramas dos graus de conectividade dos genes. Os dados histogramas foram ajustados por uma função do tipo exponencial decrescente. Esse ajuste é importante para caracterizar as redes geradas como redes livres de escala, na qual a probabilidade de que um nó tenha  $k$  arestas é dado por  $P(k) \sim k^{-\gamma}$ , onde  $\gamma$  é grau de conectividade (BARABÁSI; OLTVAI, 2004).

## 2.5 Comparação entre os métodos de geração de redes

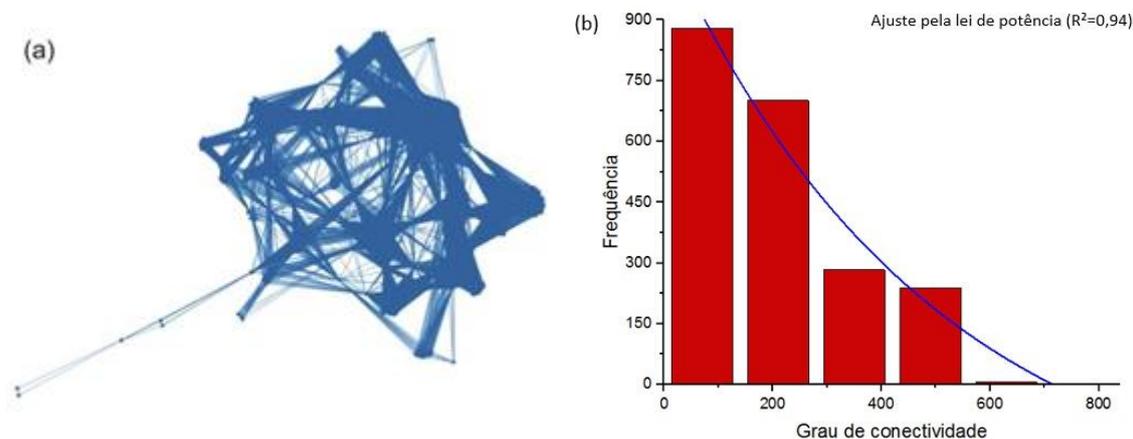
Para a comparação da performance preditiva do método proposto com o método de correlação simples para a geração de redes de co-expressão, foi feita uma intersecção das arestas das redes obtidas pelos diferentes métodos com as arestas de redes validadas experimentalmente. Para a espécie *Saccharomyces cerevisiae*, utilizou-se uma rede validada obtida do

banco de dados DREAM (MARBACH et al, 2010). Para a espécie *Homo sapiens*, utilizou-se a rede validada e publicada no trabalho de Stuart et al (2003). A interseção gerou uma matriz de confusão, caracterizada pelos parâmetros: verdadeiros positivos (*VP*), verdadeiros negativos (*VN*), falsos positivos (*FP*) e falsos negativos (*FN*).

## 3 Resultados

### 3.1 *Saccharomyces cerevisiae*

O método de geração de redes de co-expressão utilizando dados de expressão diferencial gerou uma rede com 231.452 vértices, com topologia bidimensional ilustrada na Figura 2(a). O histograma dos graus de conectividade desses nós foi bem ajustado à lei de potência, com coeficiente de determinação  $R^2=0,94$ , conforme apresentado na Figura 2(b). Esses resultados revelam que o método proposto produziu redes com topologia adequada, ou seja, redes livres de escala (CORRÊA et al, 2018).



**Figura 2** - (a) Representação gráfica bidimensional da rede de *Saccharomyces cerevisiae* gerada pelo método utilizando informação de co-expressão de genes. (b) Histograma dos graus de conectividade dos nós da rede.

Pela Figura 2(b) é possível observar que muitos genes possuem um grande número de conexões (>400), impossibilitando um melhor ajuste do histograma à lei de potência. Tais genes são conhecidos como *hubs* e são essenciais em sistemas biológicos (BARABÁSI; OLTVAI, 2004).

A Tabela 1 mostra os parâmetros da matriz de confusão para ambos os métodos de geração de redes da *Saccharomyces cerevisiae*: a expressão diferencial e a correlação simples. A matriz de confusão é importante para avaliar a eficácia dos métodos. Do ponto de vista da biologia de sistemas, é desejável que o método para a geração de redes de co-expressão produza um alto número de *VP* em uma rede com um reduzido número de arestas ( $n_{arestas}$ ), ou seja, é importante minimizar a razão  $n_{arestas}/VP$ .

**Tabela 1** - Comparação entre os métodos para geração de redes da *Saccharomyces cerevisiae*

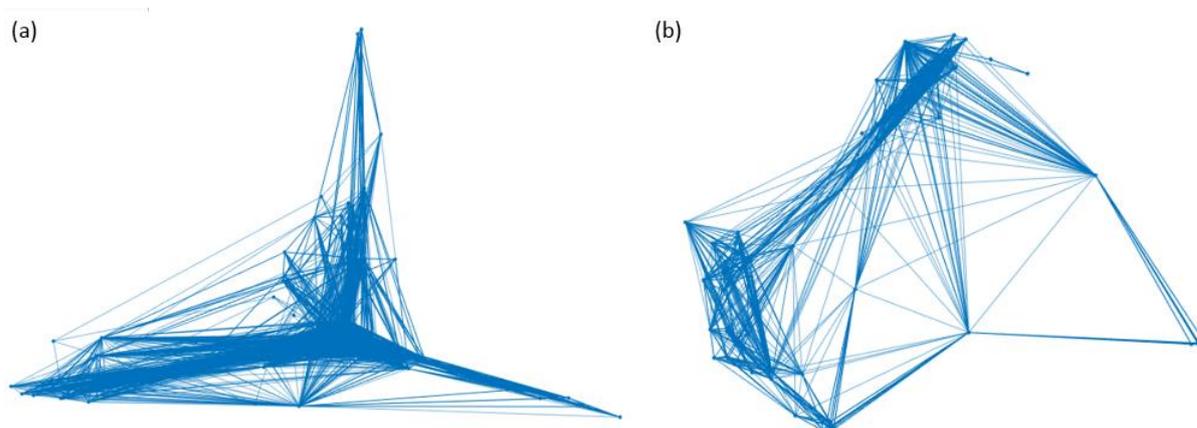
| Parâmetros | Expressão diferencial | Correlação Simples |
|------------|-----------------------|--------------------|
| <i>VP</i>  | 175                   | 10                 |
| <i>VN</i>  | 22.218.203            | 22.425.608         |
| <i>FP</i>  | 231.367               | 23.952             |

|                  |         |         |
|------------------|---------|---------|
| $FN$             | 12.008  | 12.173  |
| $n_{arestas}$    | 231.542 | 23.962  |
| $n_{arestas}/VP$ | 1.323,1 | 2.396,2 |

Os resultados mostram que o método de geração de redes de co-expressão utilizando informações de expressão diferencial de genes é mais eficiente que o método de correlação simples, identificando um número significativamente maior de arestas verdadeiras positivas e apresentando uma razão  $n_{arestas}/VP$  44,8% menor em relação à rede gerada pelo método de correlações simples.

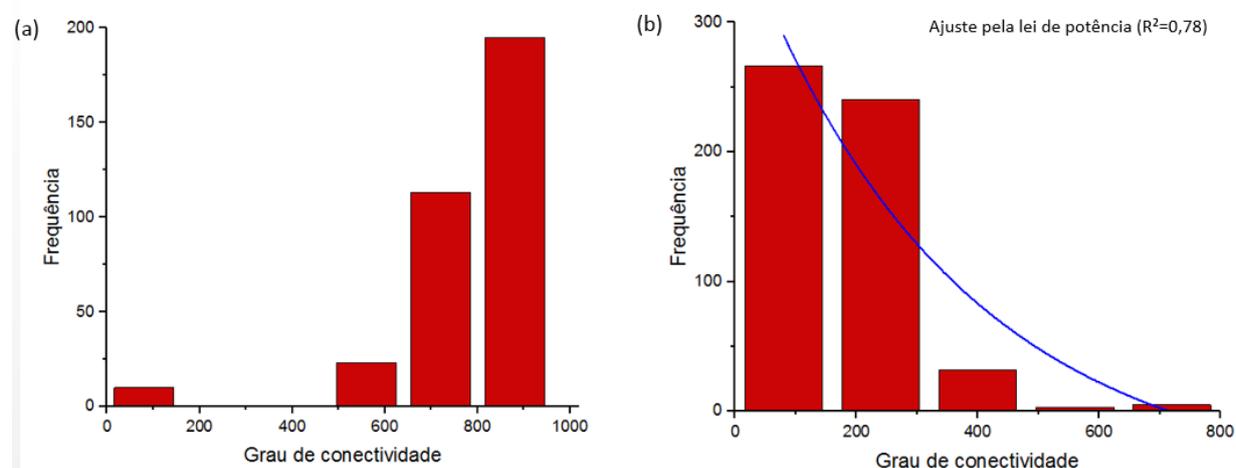
### 3.2 *Homo sapiens*

Para a rede de co-expressão do *Homo sapiens* considerando 10 experimentos, o método de expressão diferencial gerou 5.636.584 arestas, já para 36 experimentos, 2.498.55. Ambas as redes são ilustradas na Figura 3, onde foram selecionados 6.703 genes de forma aleatória a fim de permitir a comparação de suas topologias com as redes da espécie *Saccharomyces cerevisiae*.



**Figura 3** - Representação gráfica bidimensional da rede de co-expressão de genes da espécie *Homo sapiens* gerada pelo método de expressão diferencial utilizando: (a) 10 experimentos e; (b) 36 experimentos.

Os histogramas dos graus de conectividade dos genes e seus ajustes à lei de potência são apresentados na Figura 4.



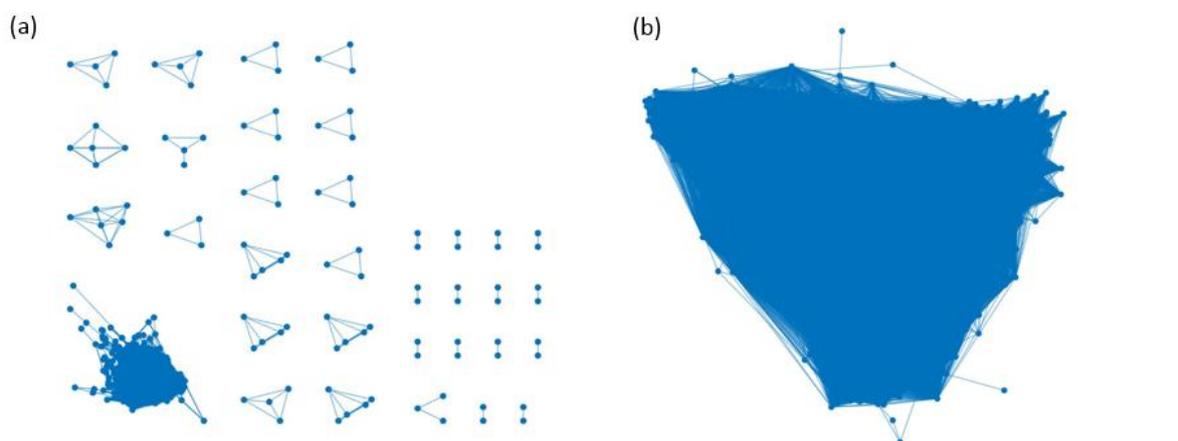
**Figura 4** - Histograma dos graus de conectividade dos nós das redes inferidas pelo método de expressão diferencial para a espécie *Homo sapiens* utilizando: (a) 10 experimentos e (b) 36 experimentos.

Pelas representações gráficas e histogramas das redes de co-expressão para a espécie *Homo sapiens* utilizando o método de expressão diferencial, é possível perceber que um menor número de experimentos ( $n_{\text{experimentos}} = 10$ ) não foi capaz de modelar adequadamente a rede, sendo que o histograma para esse conjunto de dados não pode ser ajustado à lei da potência. Já o histograma para os resultados considerando 36 experimentos ( $n_{\text{experimentos}} = 36$ ) puderam ser ajustados à lei de potência com  $R^2=0,78$ .

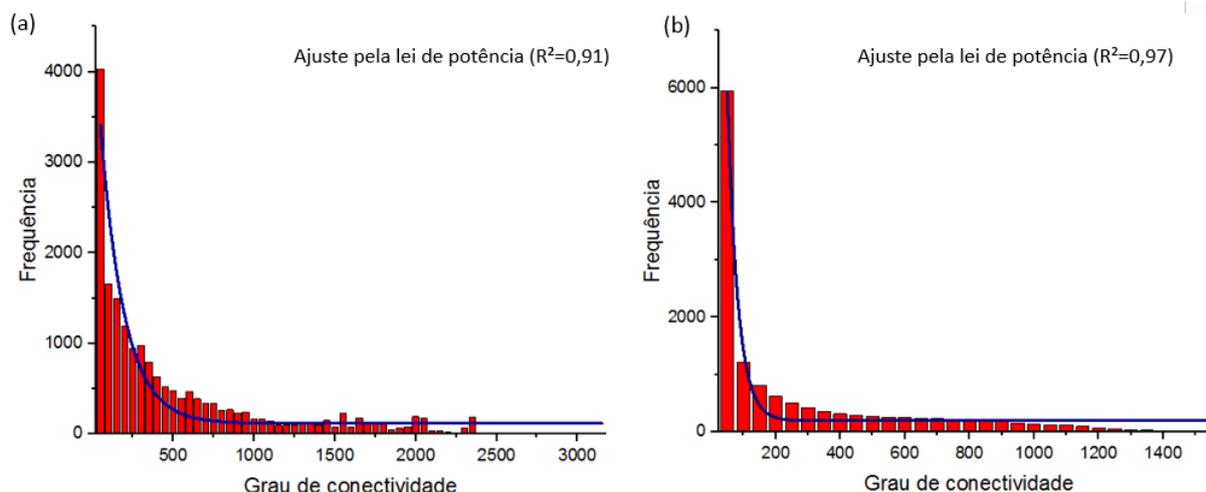
Avaliando os resultados de forma conjunta (histogramas e topologia da rede), é possível afirmar que o aumento no número de experimentos foi significativo para a melhoria da qualidade da rede, aproximando-a de uma rede livre de escala. A rede obtida com 10 experimentos possui um emaranhado de conexões na parte central, confirmando a informação do histograma de que ela possui um grande número de genes com muitas conexões, o que não é consistente e esperado em redes biológicas reais.

Esses resultados indicam, de forma indireta, que quanto maior o número de genes do organismo, mais experimentos são necessários para inferir uma rede que se assemelhe a uma rede biológica real. Essa informação é consistente quando observa-se que os dados de expressão da espécie *Saccharomyces cerevisiae* permitiram a inferência de uma rede com topologia satisfatória para seus 6.703 genes com apenas 36 experimentos ( $R^2=0,94$ ). Contudo, os dados de expressão de 21.607 genes da espécie *Homo sapiens* em 36 experimentos produziram uma rede cujo coeficiente de determinação do ajuste do histograma à lei de potência foi significativamente menor ( $R^2=0,78$ ). A avaliação dessa tendência necessita de estudos mais aprofundados e pode ser tema de trabalhos futuros.

As Figuras 5 e 6 apresentam, respectivamente, as representações gráficas das redes e histogramas gerados pelo método de correlação simples utilizando dados de expressão em 10 experimentos (Figura 5-a e Figura 6-a) e 36 experimentos (Figura 5-b e Figura 6-b). Nota-se que ambos os histogramas se adequam à lei de potência, indicando que ambas as redes possuem topologia adequada ( $R^2=0,91$  e  $R^2=0,97$ ).



**Figura 5** - Representação gráfica bidimensional das redes inferidas pelo método de correlação simples para a espécie *Homo sapiens* utilizando: (a) 10 experimentos e; (b) 36 experimentos (b).



**Figura 6** - Histograma dos graus de conectividade dos nós das redes inferidas pelo método de correlação simples para a espécie *Homo sapiens* utilizando: (a) 10 experimentos e; (b) 36 experimentos.

Os resultados mostram que as redes de co-expressão inferidas pelo método de correlação simples para a espécie *Homo sapiens* possuem topologia adequada para ambos conjuntos de dados ( $n_{\text{experimentos}} = 10$  e  $n_{\text{experimentos}} = 36$ ), entretanto é observada a mesma tendência apresentada nos resultados prévios, onde o uso de um conjunto de dados com maior número de experimentos produz redes com topologia mais adequadas às redes livre de escala, ou seja, às redes biológicas reais.

A Tabela 2 apresenta os parâmetros da matriz de confusão para as redes inferidas para a espécie *Homo sapiens* utilizando 10 e 36 experimentos, para ambas as metodologias avaliadas neste estudo.

**Tabela 2** - Comparação entre as matrizes de confusão das redes inferidas para a espécie *Homo sapiens* geradas com 10 e 36 experimentos pelo método de expressão diferencial e correlação simples.

| Métodos                 | Expressão Diferencial        |                              | Correlação Simples           |                              |
|-------------------------|------------------------------|------------------------------|------------------------------|------------------------------|
|                         | $n_{\text{experimentos}}=10$ | $n_{\text{experimentos}}=36$ | $n_{\text{experimentos}}=10$ | $n_{\text{experimentos}}=36$ |
| VP                      | 1.197                        | 225                          | 2.031                        | 3.472                        |
| VN                      | 227.754.579                  | 230.891.096                  | 228.968.923                  | 231.784.955                  |
| FP                      | 5.635.387                    | 2.498.326                    | 4.420.499                    | 1.604.467                    |
| FN                      | 29.258                       | 30.774                       | 28.968                       | 27.527                       |
| $n_{\text{arestas}}$    | 5.636.584                    | 2.498.551                    | 4.422.530                    | 1.607.939                    |
| $n_{\text{arestas}}/VP$ | 4.708,93                     | 11.104,67                    | 2.177,51                     | 463,12                       |

Observando os resultados apresentados é possível afirmar que, para a espécie *Homo sapiens*, o método de correlação simples apresentou melhor performance preditiva, o que é consistente com os resultados apresentados das topologias das redes inferidas. Esse resultado se deve principalmente ao baixo número de experimentos utilizados nas análises, conforme discutido previamente. Observa-se ainda pela Tabela 2 que as redes geradas pelo método de expressão diferencial são maiores que as redes geradas pelo método de correlação simples, indicando que um critério mais rigoroso no valor de *p-value* de expressão diferencial utilizado para estabelecer as conexões pode ser aplicado, sendo esta uma proposta para estudos futuros.



## 4 Conclusão

Os resultados apresentados no presente trabalho mostram que o método de geração de redes de co-expressão utilizando informações de expressão diferencial de genes recupera um número significativamente maior de arestas *VP* e minimiza a razão  $n_{arestas}/VP$  da rede gerada para a espécie *Saccharomyces cerevisiae*. Esse aumento de eficiência se deve ao fato da análise de expressão diferencial ser multivariada, modelando com mais eficiência a complexidade das redes biológicas.

Para a espécie *Homo sapiens*, o método de correlação simples apresentou melhor performance preditiva. Os resultados indicam que tal fato se deve ao baixo número de experimentos de expressão utilizados em relação ao elevado número de genes dessa espécie. Essa afirmação é feita com base da tendência observada de proporcionalidade entre o número de genes de uma espécie e número de experimentos necessários para a geração de uma rede de co-expressão de genes livre de escala.

## 5 Agradecimentos

Agradecemos às contribuições dadas pela Comissão Científica e Editorial da revista C.Q.D.

O primeiro autor agradece à Vunesp/SEE pela concessão da bolsa de Iniciação Científica SubPrograma de Extensão Universitária Divulgação do Vestibular Unesp e Inclusão dos Melhores Alunos da Escola Pública na Universidade.

O terceiro autor agradece à PROPE/UNESP pela concessão da bolsa de Iniciação Científica REITORIA/PIBIC.

## 6 Referências

ALLEN, J. D. et al. Comparing statistical methods for constructing large scale gene networks. **PLoS ONE**, v. 7, n. 1, Jan. 2012.

BARABÁSI, A.-L.; OLTVAI, Z. N. Network biology: understanding the cell's functional organization. **Nature Reviews Genetics**, v. 5, n. 2, p. 101–113, Feb. 2004.

CARTER, S. L. et al. Gene co-expression network topology provides a framework for molecular characterization of cellular state. **Bioinformatics** Oxford, v. 20, n. 14, p. 2242–2250, Sept. 2004.

CORRÊA, B. A. et al. Geração de redes de co-expressão de genes utilizando informações de expressão diferencial. In: ENCONTRO REGIONAL DE MATEMÁTICA APLICADA, 5., 2018, Bauru. **Caderno de trabalhos completos e resumos...** Bauru: Unesp, Faculdade de Ciências, 2018. p. 153-158. Disponível em: <<https://www.fc.unesp.br/#!/departamentos/matematica/eventos2341/ermac-2018/caderno-de-trabalhos-e-resumos/>>. Acesso em: 28 jan. 2018.



- KITANO, H. (Ed.). **Foundations of systems biology**. Cambridge: The MIT Press, 2001.
- LANGFELDER, P.; HORVATH, S. WGCNA: an R package for weighted correlation network analysis. **BMC Bioinformatics**, v. 9, p. 559, Dec. 2008.
- LI, W. et al. The EMBL-EBI bioinformatics web and programmatic tools framework. **Nucleic Acid Research**, v. 43, n. W1, p. W580-584, July 2015.
- LOVE, M. I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biology**, v. 15, p. 550, Dec. 2014.
- MARBACH, D. et al. Revealing strengths and weaknesses of methods for gene network inference. **Proceedings of the National Academy of Sciences of the United States of America**, v. 107, n. 14, p. 6286–6291, Apr. 2010.
- PAVLOVIC, B. J. et al. A comparative assessment of human and chimpanzee iPSC-derived cardiomyocytes with primary heart tissues. **Scientific Reports**, v. 8, Oct. 2018.
- ROSA, G. J. de M.; ROCHA, L. B. da; FURLAN, L. R. Estudos de expressão gênica utilizando-se microarrays: delineamento, análise, e aplicações na pesquisa zootécnica. **Revista Brasileira de Zootecnia**, v. 36, p. 186–209, Jul. 2007. Suplemento especial.
- RUAN, J.; DEAN, A. K.; ZHANG, W. A general co-expression network-based approach to gene expression analysis: comparison and applications. **BMC Systems Biology**, v. 4, p. 1-21, Feb. 2010.
- STUART, J. M. et al. A gene-coexpression network for global discovery of conserved genetic modules. **Science**, v. 302, n. 5643, p. 249-255, Oct. 2003.
- THOMAS, R.; PORTIER, C. J. Gene expression networks. In: REISFELD, B.; MAYENO, A. N. (Ed.). **Computational Toxicology**. New York: Humana Press, 2013, v. 2.
- ZERBINO, D. R. et al. Ensembl 2018. **Nucleic Acids Research**, v. 46, n. D1, p. D754–D761, Jan. 2018.
- ZHANG, B.; HORVATH, S. A general framework for weighted gene co-expression network analysis. **Statistical Applications in Genetics and Molecular Biology**, v. 4, n. 1, Aug. 2005.