Graphs and biological networks

Practical Course

Martin Junghanns

Computer Science
University of Leipzig

August 9, 2013

Content

Environment

- Python 2.7.3 for calculating
- SciPy (Spearman Correlation + p-Value)
- Kernel 3.6.10-2.fc16.i686.PAE
- R 2.15.2 for plotting

Data

- `adult.txt` 3 species, 6 individuals per species, 12955 genes
- `develop.txt` 3 species, 12 individuals per species, 12098 genes
- `TFgeodNo4.txt` list of 280 TF

Calculations - Spearman

- Spearman Correlation + p-value

$$\rho = 1 - \frac{6 \sum_i d_i^2}{n(n^2 - 1)} \text{mit } d_i = rk(x_i) - rk(y_i) \tag{1}$$

- For all TF-Pairs and TF-Gene-Pairs (280 x 12xxx)
- Result is matrix A with $a_{ij} = (\rho, \texttt{p-value})$
- Example: TF1 = JRKL, TF2 = ZNF264, Species = Human

$$v_{TF1} = (2.11, 2.08, 2.31, 2.24, 2.03, 2.01) \tag{2}$$

$$v_{TF2} = (2.78, 3.03, 2.91, 2.87, 2.94, 2.66) \tag{3}$$

$$\rho = 0.142857142857 \tag{4}$$

$$p - value = 0.787172011662 \tag{5}$$

Calculations - Topological Overlap

- Topological Overlap

$$\omega_{ij} = \frac{\sum_{u \in U} a_{iu} a_{uj}}{\mathtt{min}(k_i, k_j) + 1 - |a_{ij}|} \tag{6}$$

- U is the topological overlap between Gene i and j
- $k_i = \sum_j |a_{ij}|$ is the weighted connectivity
- For all TF-Pairs (280 x 280)
- Result is matrix Ω

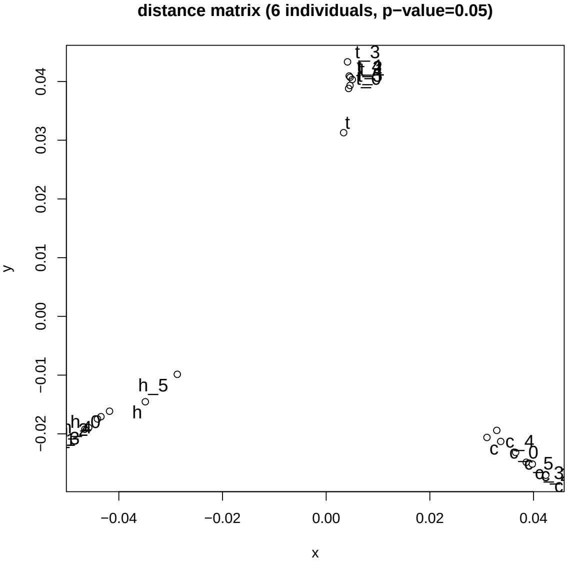
Calculations - Leave-one-out

- Cross-validation method to test the quality of the model
- Calculate A and Ω for all possible vectors of length $n-1$ where $n = |\texttt{Individuals}|$
- Result is $\Omega_1, \Omega_2, \ldots, \Omega_{sn+n}$ Overlap matrices where $s = |\texttt{species}|$

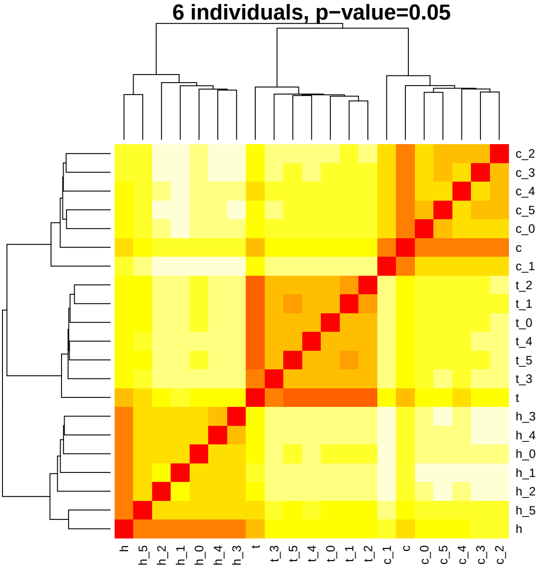
Calculations - Distance Matrix

- Calculate the TF-pair-wise distance between each pair of resulting Ω matrices
- $\omega_{i,j}^{s,x}$ is distance value for TF-pair (i, j) species s and leave-out individual x with $x \in \{1, 2, \ldots, n, \emptyset\}$
- $d((s_1, x_1), (s_2, x_2)) = \frac{1}{n^2} \sum_{i \in [n]} \sum_{j \in [n]} \left(\left| \omega_{i,j}^{s_1,x_1} - \omega_{i,j}^{s_2,x_2} \right| \right)^{\alpha}$ with $\alpha \in [1, \infty)$
- Result is Distance Matrix D $((sn + n)) \times (sn + n))$

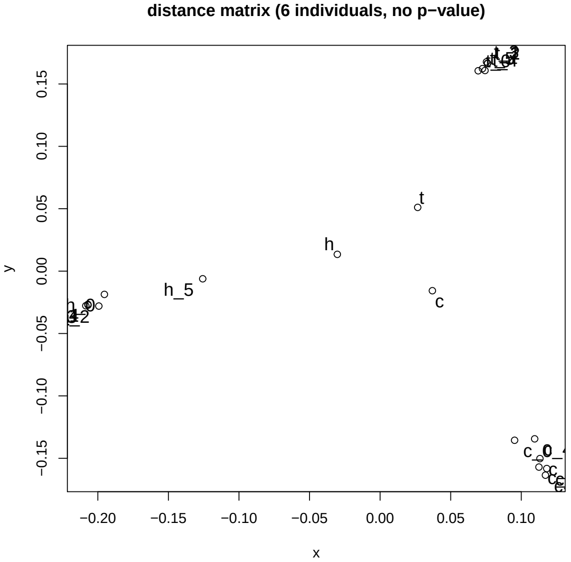
Results


distance matrix (6 individuals, p−value=0.05)
t_3
t_4
t_0
t
h_5
h
h_0
c_4
c_0
c
c_5
c_3
y
0.04
0.03
0.02
0.01
0.00
−0.01
−0.02
−0.04
−0.02
0.00
0.02
0.04
x

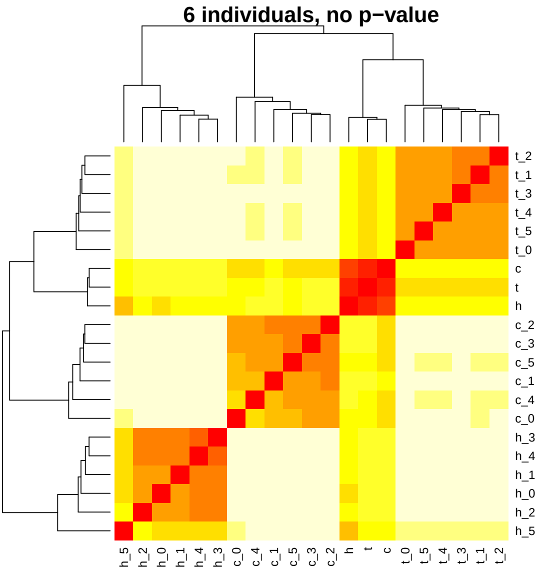
Results


6 individuals, p−value=0.05
c_2
c_3
c_4
c_5
c_0
c
c_1
t_2
t_1
t_0
t_4
t_5
t_3
t
h_3
h_4
h_0
h_1
h_2
h_5
h
h
h_5
h_2
h_1
h_0
h_4
h_3
t
t_3
t_5
t_4
t_0
t_1
t_2
c_1
c
c_0
c_5
c_4
c_3
c_2

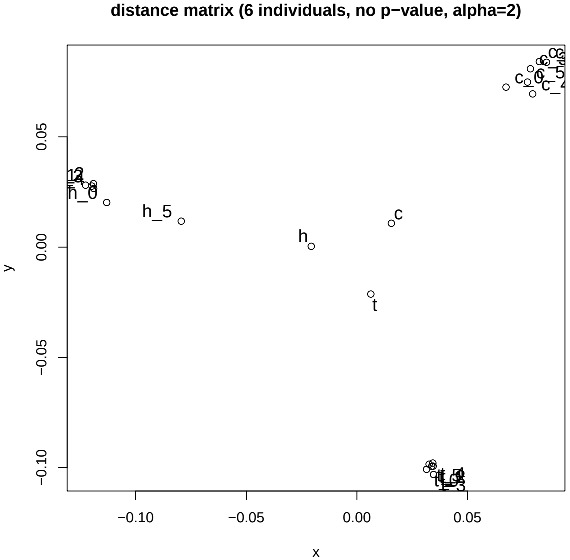
Results


distance matrix (6 individuals, no p−value)
y
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
x
−0.20
−0.15
−0.10
−0.05
0.00
0.05
0.10
t
h
h_5
c
c_0

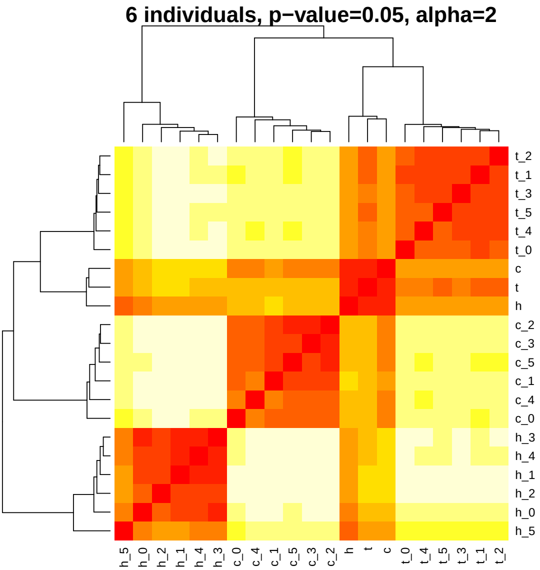
Results


6 individuals, no p−value
t_2
t_1
t_3
t_4
t_5
t_0
c
t
h
c_2
c_3
c_5
c_1
c_4
c_0
h_3
h_4
h_1
h_0
h_2
h_5
h_5
h_2
h_0
h_1
h_4
h_3
c_0
c_4
c_1
c_5
c_3
c_2
h
t
c
t_0
t_5
t_4
t_3
t_1
t_2

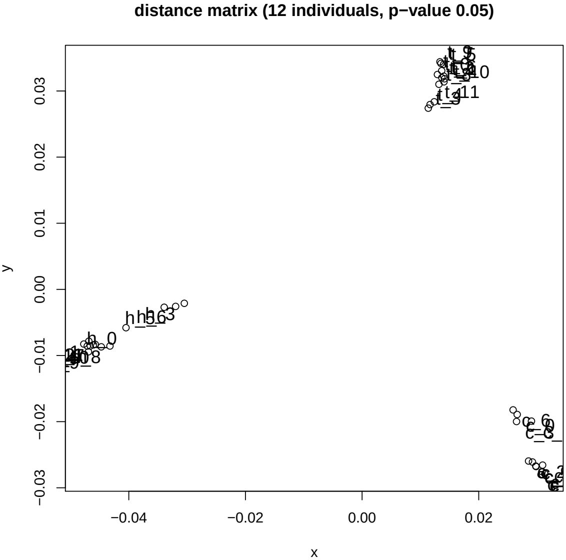
Results


distance matrix (6 individuals, no p−value, alpha=2)
y
x
0.05
0.00
−0.05
−0.10
−0.10
−0.05
0.00
0.05
h_0
h_5
h
c
t
c_0
c_5
c_4
t_0
t_5
t_3

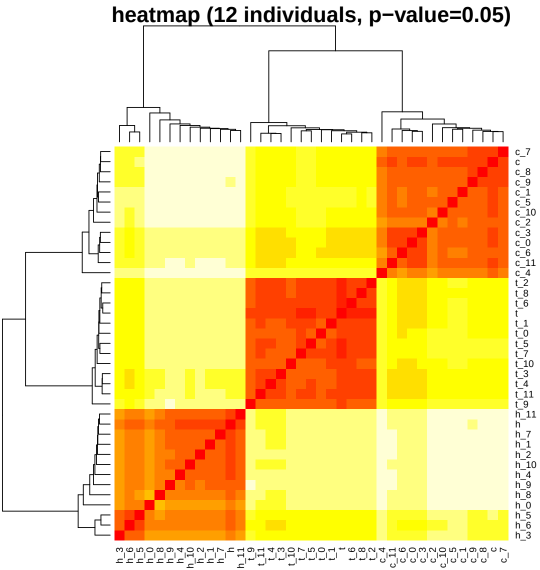
Results


6 individuals, p−value=0.05, alpha=2
t_2
t_1
t_3
t_5
t_4
t_0
c
t
h
c_2
c_3
c_5
c_1
c_4
c_0
h_3
h_4
h_1
h_2
h_0
h_5
h_5
h_0
h_2
h_1
h_4
h_3
c_0
c_4
c_1
c_5
c_3
c_2
h
t
c
t_0
t_4
t_5
t_3
t_1
t_2

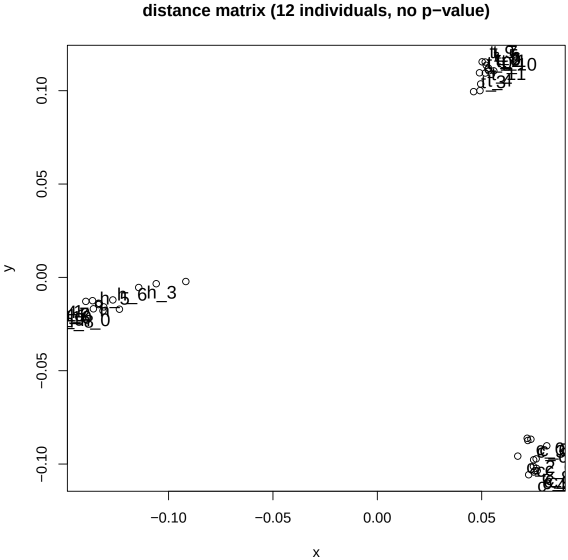
Results


distance matrix (12 individuals, p−value 0.05)
t_9
t_5
t_10
t_11
t_3
t_4
h_5
h_6
h_3
h_0
h_8
c_6
c_0
c_8
0.03
0.02
0.01
0.00
−0.01
−0.02
−0.03
−0.04
−0.02
0.00
0.02
x
y

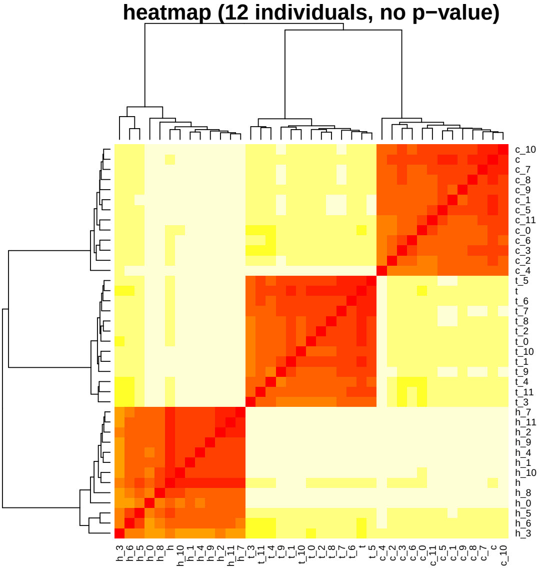
Results


heatmap (12 individuals, p−value=0.05)
c_7
c
c_8
c_9
c_1
c_5
c_10
c_2
c_3
c_0
c_6
c_11
c_4
t_2
t_8
t_6
t
t_1
t_0
t_5
t_7
t_10
t_3
t_4
t_11
t_9
h_11
h
h_7
h_1
h_2
h_10
h_4
h_9
h_8
h_0
h_5
h_6
h_3
h_3
h_6
h_5
h_0
h_8
h_9
h_4
h_10
h_2
h_1
h_7
h
h_11
t_9
t_11
t_4
t_3
t_10
t_7
t_5
t_0
t_1
t
t_6
t_8
t_2
c_4
c_11
c_6
c_0
c_3
c_2
c_10
c_5
c_1
c_9
c_8
c
c_7

Results


distance matrix (12 individuals, no p−value)
y
x
0.10
0.05
0.00
−0.05
−0.10
−0.10
−0.05
0.00
0.05
t_10
t_11
t_3
h_5
h_6
h_3
c_

Results



heatmap (12 individuals, no p−value)
c_10
c
c_7
c_8
c_9
c_1
c_5
c_11
c_0
c_6
c_3
c_2
c_4
t_5
t
t_6
t_7
t_8
t_2
t_0
t_10
t_1
t_9
t_4
t_11
t_3
h_7
h_11
h_2
h_9
h_4
h_1
h_10
h
h_8
h_0
h_5
h_6
h_3