Bioinformatics Resources
Exercise Sheet 6
due on July 3rd, 9 a.m.
Total number of points: 9P

Task 1
Given the following ER diagram, and assuming that you only want to use two tables to model it in a relational database: What are your table definitions? The datatypes are not of interest, but do not forget to specify primary keys etc. How would you model it with three tables? Which approach do you prefer in this case and why? (3P)

ER: (men)-1--<marriedTo>--->1-(women) that is, two entities, men and women, connected in a 1-1 relationship marriedTo.

Task 2
2a) Given the following two tables in a relational database, what is the SQL statement to select the gate of a flight that travels from Munich to Tokyo, with one stop. (1P)
Table Airport
IATA_Code | City
--- | ---
MUC | Munich
LGA | New York
JFK | New York
EWR | New York
NRT | Tokyo
HND | Tokyo
TXL | Berlin
...

Table Connections
Flight_Number | From_Code | To_Code | Airline | Gate
--- | --- | --- | --- | ---
42 | LGA | NRT | ANA | A2
1337 | MUC | TXL | LH | C3
...

2b) Now model this data in neo4j. Give the corresponding create statements for some sample data, as well as the cypher query to answer the question from above. What is the query if you want your layover to be in New York and the airline for the second flight to be ANA? (3P)

Task 3
You have a logfile with pairs of IPs and timestamps for accessing your website. Write the map and reduce functions to count the number of accesses in a given...
time interval $t_i$. Give the input and output of both functions and describe what is done with the inputs to get to the output. You can make use of mathematical operations such as $\in$ as well. (2P)

Good luck,
J. Reeb & L. Richter