Exam DM840 Cheminformatics (2022 (Exam Jan 2023))

Time and Place

Time: Thursday, January 25, 2022, starting time 08:30.
Place: The exam takes place physically (IMADA meeting room 3).

Even though the expected total examination time per student is about 27 minutes (see below), it is not possible to calculate the exact examination time from the placement on the list (to be put in blackboard), since students earlier on the list may not show up. I will during the exam update a webpage such that you can see which students have already had their exam, and for whom the exam is currently i progress. Be careful, as it can happen that there can be no-shows, and we expect you to be available then even if you are next in line. In principle, all students who are taking the exam on a particular date are supposed to "show up" when the examination starts, i.e., at the time the first student is scheduled. This is partly because of the way external examiners are paid, which is by the number of students who show up for examination. For this particular exam, we do not expect many no-shows, so checking the online list (link will be send) up one to two hours before the estimated time of the exam should be safe.

Procedure

The exam is in English.

The procedure is as follows:

You will first randomly chose a question and give a presentation (technically you will decide to turn around a specific piece of paper with the question number on it). This will be followed by an the examinator and censor asking question on the topic you chose. After that, the second part of the exam starts, where the examinator and censor will ask questions on all other topics listed below.

When it is your turn for examination, you will draw a question. Note that you have no preparation time. The list of questions can be found below. Then the actual exam takes place.

The whole exam (without the censor and the examiner agreeing on a grade) lasts approximately 25-30 minutes. You should start by presenting material related to the question you drew. Aim for a reasonable high pace and focus on the most interesting material related to the question. You are not supposed to use note material, textbooks, transparencies, etc. Aim for a reasonable high pace and focus on the most interesting material related to the question.

While most of the exam is oral, you must also ensure that you can write text and draw figures and present this to the examinator and censor via Zoom (similar to a blackboard in a non-virtual exam). For presenting you may either share your screen or you may your camera and just record the paper that you will be writing on. Please test your setup beforehand, such that there are no unexpected delays. You should only start from empty paper. The only two figures that you might use in addition are attached at the end of this document.

You are allowed to use keywords for each question, such that you can remember what you want to present during your presentation. As a guideline you are expected to not have more than 10 keywords per question on the list, that you are allowed to use during the oral exam. Note that this list is expected to be put on your table during

your oral exam. We will not accept that you use any material during the (second) part of the exam, where the examiner and censor ask questions.

We, the examinator and the censor, will supplement with specific questions when appropriate, and after a while, we will end the discussion of the exam question that you drew and turn to material from other parts of the curriculum. Note that all of this as well as discussion between examinator and censor about the grade is included in the approx minutes, so **do not count on more than 10-14 minutes for your own presentation**.

Some of the questions below are quite broad, so you must select the material you choose to cover. You will of course also be evaluated based on your selection of material. If you only present the simplest material, you limit the grade you can obtain. On the other hand, a good presentation of the simple material is better than a very poor presentation of the harder material. We might of course still ask you questions about material that you have decided to skip.

Curriculum

The curriculum consists of all documents listed below. The following list of articles, book chapters, and slide sets is a superset of what you should know. Note that "mandatory" might of course not necessarily mean, that you need to know the complete content of the articles/slides, nothing that has not been discussed during the lecture you do not have to know. Furthermore, there is also mandatory material not marked as mandatory below.

Slides / notes:

week36-2022.GenerativeChemistries.pdf
week36-2022.slideset1.final.pdf
week36-2022.slideset2.final.pdf
week37-2022-canonicalization.pdf
week38-2022.Isomorphism-Ullmann.mandatory.pdf
week38-2022-mckay-notes.pdf
week38-2022-morgan.pdf
week38-2022-strategies-slides.pdf
week38-SMILES-notes.mandatory.(or_weininger_article).pdf
week39-2022-canon-addon.pdf
week39-2022.MCB.horton.dePina.mandatory.pdf
week39-2022-MOD-framework-addon.pdf
week39-2022-traces.partially-mandatory.pdf
week39.ring-perception-2022-notes(summary_of_mandatory_article).pdf
week40-2022.cycle-basis.not.final.pdf
week40-traces.partially-mandatory.pdf
week41-2022-petrinet.mandatory.pdf
week41.cycle-basis.mandatory-2.not.final.pdf
week41-MCB-pina.runtime.voluntary.pdf
week41-petri-examples.slide-available-in-oral-exam.pdf
week43-AC-OGF-detailed.pdf
week44-AC-EGF-detailed.pdf
week45-2022-kshortest.pdf
week45-AC-MGF-voluntary.pdf
week46-AC-complex.justanaddon.pdf
week46-boltzmann.voluntary.pdf
week46-stereo-combinatorics.pdf
week47-cayley-semigroup.pdf
week47-mol-des.partially-mandatory.pdf
week47.MultistepEnzymeNOG.pdf
week47-pca-notes-voluntary_notes_follow_the_article.pdf
week48-Process.Algebras_-_Gillespie.Simulation.pdf

Book Chapters:

AnalysisChapter5AnalyticCombinatorics.pdf
Palsson.pdf
week36-2021.jla.thesis.GraphGrammars.mandatory.pdf
week36-2021-partially-mandatory--Chemoinformatics.Gasteiger.chapter2.pdf

week36-2021-polya-enumeration-teaser_and_graph-theory-roots.voluntrary.pdf
week36-2021-voluntary-Combinatorics-Ancient-and-Modern_-_Chapter8_-_Robin-Wilson_Early-Graph-Theory
week36-2022.jla.thesis.GraphGrammars.mandatory.pdf
week36-2022-partially-mandatory--Chemoinformatics.Gasteiger.chapter2.pdf
week36-2022-polya-enumeration-teaser_and_graph-theory-roots.voluntrary.pdf
week36-2022-voluntary-Combinatorics-Ancient-and-Modern_-_Chapter8_-_Robin-Wilson_Early-Graph-Theory
week37.38.jla.thesis.canonicalisation.mandatory.pdf
week38.39.jla.thesis.strategies.mandatory.pdf
week38.jla.thesis.canonicalisation.mandatory.pdf
week38.jla.thesis.networks-strategies.mandatory.pdf

Original Articles:

(to be added)

In addition the curriculum contains the text of the required assignments and the Weekly Notes/Exercises.

Exam Questions

In the following the list questions that you draw from is given (in bold face), (the list of subquestions is incomplete and just a suggestion.)

- **Canonical Representations and Graph Isomorphism**
  - Morgan's algorithm
  - SMILES notation
  - Canonical labeling of graphs
  - Subgraph and graph isomorphism (Ullmanns algorithm)
  - Graph isomorphism (automorphism-pruning-based approaches like traces, nauty)
  - Relation to (generative) chemistry

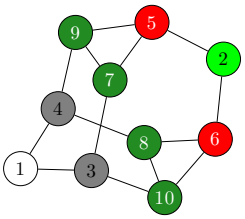
- **Graph Grammars / Generative Chemistries**
  - Graph Grammars
  - Double pushout approach
  - ILP approach in hypergraphs
  - Autocatalysis
- **Ring Perception**
  - Hanser Algorithm
  - Cycle Bases in chemistry
  - Kirchhoff-fundamental cycle basis
  - Minimal cycle basis
  - Horton's algorithm
  - de Pina's algorithm
- **Petri Nets**
  - Properties (liveness/boundedness/reversibility)
  - Reachability graph, coverability graph
  - Invariants and how to determine them (Farkas algorithm)
  - NPc proof of reachable marking
  - Petri nets to model chemical/biological networks
- **Systems Biology, Stochastic Simulations, and Process Algebras**
  - Gillespie
  - Law of Mass Action
  - Flux Balance Analysis and example for an objective function?
  - π-Calculus and modeling of Chemistry with stochastic π-calculus

- **Combinatorics**
  - Generating Functions
  - Symbolic Method
  - Ordinary and Exponential GF
  - Multivariate GF
  - ~~Cylce Index Polynomial and Burnside's Lemma~~
  - Connection to Chemistry (counting isomers, sampling, ...)

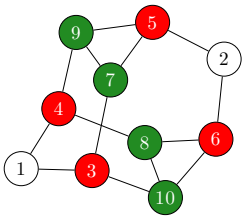
The following is an (incomplete) list of topics that have also been discussed in the course, but they won't appear as an individual presentation question in the oral exam. You of course might still be asked questions on those topics in the second part of the exam.

- **NOT A QUESTION: Process Algebras**
  - π-Calculus
  - Modeling of Chemistry with stochastic π-calculus
  - Modeling of epidemic systems with stochastic π-calculus
- **NOT A QUESTION: Molecular Descriptors and QSAR**
  - Molecular descriptors
  - QSAR
  - Principal component analysis (PCA) / PCR

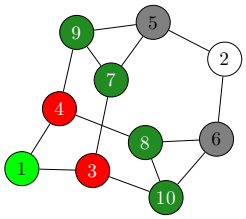
The following two figures is material you can use. You can (but don't have to) use them during the second part (not in the first part) in order to explain.



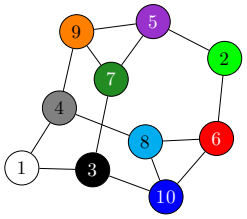
$\pi_{(1)} = [1 \mid 2 \mid 7\ 8\ 9\ 10 \mid 5\ 6 \mid 3\ 4]$



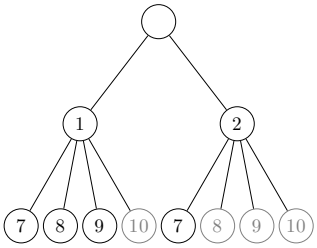
$\pi_{()} = [1\ 2 \mid 7\ 8\ 9\ 10 \mid 3\ 4\ 5\ 6]$

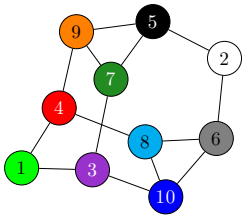


$\pi_{(2)} = [2 \mid 1 \mid 7\ 8\ 9\ 10 \mid 3\ 4 \mid 5\ 6]$

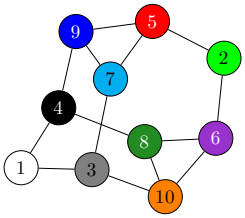


$\pi_{(1,7)} = [1 \mid 2 \mid 7 \mid 10 \mid 8 \mid 9 \mid 6 \mid 5 \mid 4 \mid 3]$



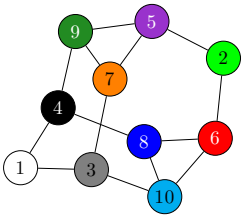


$\pi_{(2,7)} = [2 \mid 1 \mid 7 \mid 10 \mid 8 \mid 9 \mid 4 \mid 3 \mid 6 \mid 5]$

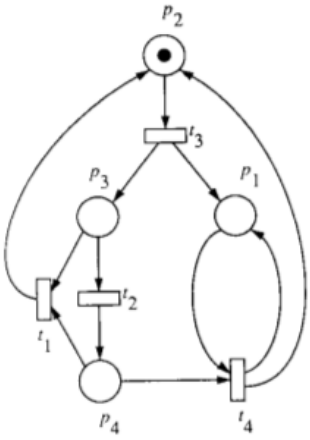


$\pi_{(1,8)} = [1 \mid 2 \mid 8 \mid 9 \mid 7 \mid 10 \mid 5 \mid 6 \mid 3 \mid 4]$

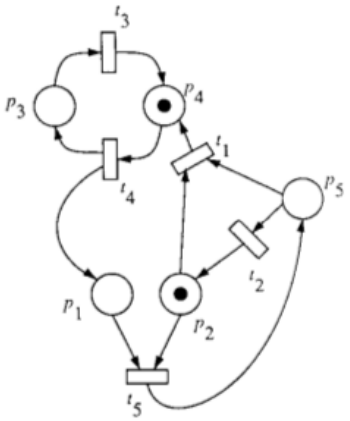
Colour order



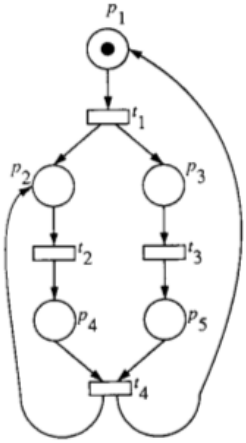
$\pi_{(1,9)} = [1 \mid 2 \mid 9 \mid 8 \mid 10 \mid 7 \mid 6 \mid 5 \mid 3 \mid 4]$


p_2
t_3
p_3
p_1
t_2
t_1
p_4
t_4

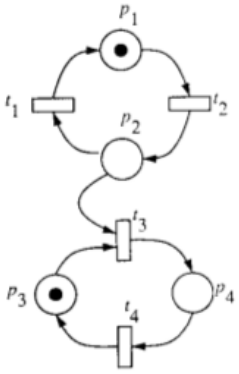
(a)


t_3
p_3
p_4
t_1
t_4
p_5
t_2
p_1
p_2
t_5

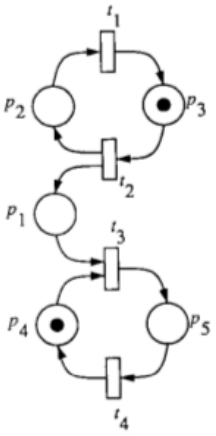
(b)


p_1
t_1
p_2
p_3
t_2
t_3
p_4
p_5
t_4

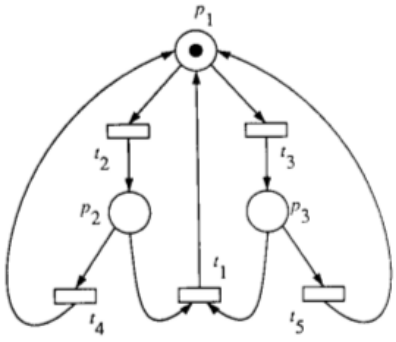
(c)


p_1
t_1
t_2
p_2
t_3
p_3
p_4
t_4

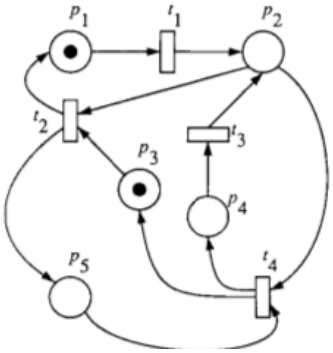
(d)

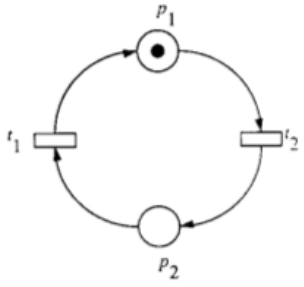

t_1
p_2
p_3
t_2
p_1
t_3
p_4
p_5
t_4

(e)


p_1
t_2
t_3
p_2
p_3
t_1
t_4
t_5

(f)


p_1
t_1
p_2
t_2
t_3
p_3
p_4
p_5
t_4


p_1
t_1
t_2
p_2