1---

In this colloquium, I would like to present my research results achieved under the supervision of professor tatjana petrov on data-informed parameter synthesis for markov population models

2--

* In many fields of research and application, we interest in population of single or multiple species,
  + For example: how many surviving nodes are there in a distributed system
* We also care about population dynamics, that is how does the population evolve
  + For example: Given a community, how many people get infected before the number of infections stabilized
* In this thesis we study the population dynamics by modeling the population using stochastic population process

3---

* Specifically, we use stochastic population process study population dynamics which contains random evolution.
  + States tracks population sizes of involved species
  + Transitions model the random evolution of the population of each species
* Which population process to use? In this thesis we use Markov population process.
* To simplify the problem, we model the population with discrete-time model
* Thus, the main object of research in this thesis is
  + DTMC with state labels represents population vector, I.e population of each involved species

4---

* As in all model checking problems, we first need to define a model
* In this thesis we use Discrete-time markov chain as the model.

5---

* We also needs to define a temporal logic to specify the property of interest
* Here comes the definition of pctl

6---

Complexity: polynomial to state space size

Looks nice, I will show later why model checking is still computationally expensive

7---

Modelling using dtmc, we have once challenge:

* **Unknown features of the system** 
  + In DTMC, transition matrix is known a priori.
  + However, the population dynamics is usually unknown, or assumed
  + => We introduce *parametric models.*
  + *Pmodels is partially know, since we know where the transition might be positive, or graphical structure, we don’t know yet the exact value*

We also assume greybox study of the system => we can observe steady state data

8---

Formal definition of pdtmc follow Phd thesis of Jnges at katoen group

9---

* There are different definitions of parameter synthesis, here we stick to the definition by Katoen, 2011 in his paper probabilistic model checking landscape that (def..)
* In the paper, Katoen also summarized the following method for parameter synthesis. These methods based on parameter space partitioning
* In this thesis, we use sampling-based method.

10---

* *As we have parameterized transitions, the following questions emerge:*
* *Parameter inference*
  + How to estimate model parameters given experiment data?
* *Parameter synthesis*
  + How to find a set of parameters satisfying a temporal property of interest?

11---

* Number of model states grows exponentially as the number of involved state variable increase
* The exponential growing of state space leads to high computational cost
* Of course during the model construction, some states can be *lumped*
* However, lumping is not always possible
* => need another method to mitigate
* …. (Tanja: this is before the population abstraction – leave it like this, but be aware that it may be asked.. Population abstraction is not exponential, but is combinatorial

12---

* Statistical model checking is a simulation-based approach to model checking
* Statistical model checking trade-off accuracy to computational cost

13---

1. The system itself behaves randomly; thus we model with stochastic process

2. We use data informed methods, thus low quality/noisy data deteriorates inference results. We later show how to deal with this situation.

3. Randomized algorithm, each execution of algorithm gives different results. Furthermore, how do we select hyperparameters? Just try...

14---

Leftside: we see the shape (trend) but not the color

Right side: we see the color but not the shape

15---

* *Data-informed parameter synthesis, this is the main contribution of this thesis*
  + *Find a set of satisfying parameters values*
  + *Approximate or exact their corresponding likelihood to generate experiment data*
  + *From the set of parameter and likelihood values, estimate model parameters, validate its satisfaction to the property of interest*

Sanity check to detect the case that parameter space is complicatedly partitioned (nonconved)

16---