Statistical learning on period app data to advance personalized health care for women

AMLD2020

Laura Symul

@LauraSymul

lsymul@stanford.edu

Women* menstruate

* women = shortcut for the female biological sex or for “individuals who experience menstruations”

The menstrual cycle is crucial for reproduction

The menstrual cycle is crucial for reproduction

Because of **its crucial role for the survival of the species**, it should not surprise anyone that **the menstrual cycle has far-reaching effects** on our **health**, **behaviors** and **emotional states**, beyond its reproductive dimension and that the **hormonal regulation pathways are redundant, robust and complex**.

The menstrual cycle is crucial for reproduction

Because of **its crucial role for the survival of the species**, it should not surprise anyone that **the menstrual cycle has far-reaching effects** on our **health**, **behaviors** and **emotional states**, beyond its reproductive dimension and that the **hormonal regulation pathways are redundant, robust and complex**.

There are active estrogen receptors in virtually every tissue of the human body

However...

However...

The menstrual cycle has long been an

However...

The menstrual cycle has long been an
Understudied (mostly focusing on the fertility aspect)**,**

However...

The menstrual cycle has long been an
Understudied (mostly focusing on the fertility aspect)**,**
Under-funded (so little money to study endometriosis for example)

However...

The menstrual cycle has long been an
Understudied (mostly focusing on the fertility aspect)**,**
Under-funded (so little money to study endometriosis for example)
& quasi mono-disciplinary (mostly research by obstetricians & gynecologists, where the circadian clock, another fundamental biological rhythm, has been heavily studied from a medical, biological, ecological, modeling perspective)

However...

The menstrual cycle has long been an
Understudied (mostly focusing on the fertility aspect)**,**
Under-funded (so little money to study endometriosis for example)
& quasi mono-disciplinary (mostly research by obstetricians & gynecologists, where the circadian clock, another fundamental biological rhythm, has been heavily studied from a medical, biological, ecological, modeling perspective)
field of research

However…

The menstrual cycle has long been an
Understudied (mostly focusing on the fertility aspect)**,**
Under-funded (so little money to study endometriosis for example)
& quasi mono-disciplinary (mostly research by obstetricians & gynecologists, where the circadian clock, another fundamental biological rhythm, has been heavily studied from a medical, biological, ecological, modeling perspective)
field of research

We still do not know…

However…

The menstrual cycle has long been an
Understudied (mostly focusing on the fertility aspect)**,**
Under-funded (so little money to study endometriosis for example)
& quasi mono-disciplinary (mostly research by obstetricians & gynecologists, where the circadian clock, another fundamental biological rhythm, has been heavily studied from a medical, biological, ecological, modeling perspective)
field of research

We still do not know…

- The causes of pre-menstrual symptoms and pre-menstrual dysphoric disorder

However…

The menstrual cycle has long been an
Understudied (mostly focusing on the fertility aspect)**,**
Under-funded (so little money to study endometriosis for example)
& quasi mono-disciplinary (mostly research by obstetricians & gynecologists, where the circadian clock, another fundamental biological rhythm, has been heavily studied from a medical, biological, ecological, modeling perspective)
field of research

We still do not know…

- The causes of pre-menstrual symptoms and pre-menstrual dysphoric disorder
  - A genetic predisposition?

However…

The menstrual cycle has long been an
Understudied (mostly focusing on the fertility aspect)**,**
Under-funded (so little money to study endometriosis for example)
& quasi mono-disciplinary (mostly research by obstetricians & gynecologists, where the circadian clock, another fundamental biological rhythm, has been heavily studied from a medical, biological, ecological, modeling perspective)
field of research

We still do not know…

- The causes of pre-menstrual symptoms and pre-menstrual dysphoric disorder
  - A genetic predisposition?
  - An hormonal sensitivity? General? Tissue-specific?

However…

The menstrual cycle has long been an
Understudied (mostly focusing on the fertility aspect)**,**
Under-funded (so little money to study endometriosis for example)
& quasi mono-disciplinary (mostly research by obstetricians & gynecologists, where the circadian clock, another fundamental biological rhythm, has been heavily studied from a medical, biological, ecological, modeling perspective)
field of research

We still do not know…

- The causes of pre-menstrual symptoms and pre-menstrual dysphoric disorder
  - A genetic predisposition?
  - An hormonal sensitivity? General? Tissue-specific?
  - An abnormality of the immune system? In the interactions with the immune system?

However…

The menstrual cycle has long been an
Understudied (mostly focusing on the fertility aspect)**,**
Under-funded (so little money to study endometriosis for example)
& quasi mono-disciplinary (mostly research by obstetricians & gynecologists, where the circadian clock, another fundamental biological rhythm, has been heavily studied from a medical, biological, ecological, modeling perspective)
field of research

We still do not know…

- The causes of pre-menstrual symptoms and pre-menstrual dysphoric disorder
  - A genetic predisposition?
  - An hormonal sensitivity? General? Tissue-specific?
  - An abnormality of the immune system? In the interactions with the immune system?
  - An evolutionary strategy to lead infertile couples to break-ups?

However…

The menstrual cycle has long been an
Understudied (mostly focusing on the fertility aspect)**,**
Under-funded (so little money to study endometriosis for example)
& quasi mono-disciplinary (mostly research by obstetricians & gynecologists, where the circadian clock, another fundamental biological rhythm, has been heavily studied from a medical, biological, ecological, modeling perspective)
field of research

We still do not know…

- The causes of pre-menstrual symptoms and pre-menstrual dysphoric disorder
  - A genetic predisposition?
  - An hormonal sensitivity? General? Tissue-specific?
  - An abnormality of the immune system? In the interactions with the immune system?
  - An evolutionary strategy to lead infertile couples to break-ups?
- The drivers of human birth seasonality? Which aspects of fertility are seasonal?

However…

The menstrual cycle has long been an
Understudied (mostly focusing on the fertility aspect)**,**
Under-funded (so little money to study endometriosis for example)
& quasi mono-disciplinary (mostly research by obstetricians & gynecologists, where the circadian clock, another fundamental biological rhythm, has been heavily studied from a medical, biological, ecological, modeling perspective)
field of research

We still do not know…

- The causes of pre-menstrual symptoms and pre-menstrual dysphoric disorder
  - A genetic predisposition?
  - An hormonal sensitivity? General? Tissue-specific?
  - An abnormality of the immune system? In the interactions with the immune system?
  - An evolutionary strategy to lead infertile couples to break-ups?
- The drivers of human birth seasonality? Which aspects of fertility are seasonal?
- How menstrual bleeding is regulated

However…

The menstrual cycle has long been an
Understudied (mostly focusing on the fertility aspect)**,**
Under-funded (so little money to study endometriosis for example)
& quasi mono-disciplinary (mostly research by obstetricians & gynecologists, where the circadian clock, another fundamental biological rhythm, has been heavily studied from a medical, biological, ecological, modeling perspective)
field of research

We still do not know…

- The causes of pre-menstrual symptoms and pre-menstrual dysphoric disorder
  - A genetic predisposition?
  - An hormonal sensitivity? General? Tissue-specific?
  - An abnormality of the immune system? In the interactions with the immune system?
  - An evolutionary strategy to lead infertile couples to break-ups?
- The drivers of human birth seasonality? Which aspects of fertility are seasonal?
- How menstrual bleeding is regulated
  - and, consequently what happens during anovulatory cycles?

However…

The menstrual cycle has long been an
Understudied (mostly focusing on the fertility aspect)**,**
Under-funded (so little money to study endometriosis for example)
& quasi mono-disciplinary (mostly research by obstetricians & gynecologists, where the circadian clock, another fundamental biological rhythm, has been heavily studied from a medical, biological, ecological, modeling perspective)
field of research

We still do not know…

- The causes of pre-menstrual symptoms and pre-menstrual dysphoric disorder
  - A genetic predisposition?
  - An hormonal sensitivity? General? Tissue-specific?
  - An abnormality of the immune system? In the interactions with the immune system?
  - An evolutionary strategy to lead infertile couples to break-ups?
- The drivers of human birth seasonality? Which aspects of fertility are seasonal?
- How menstrual bleeding is regulated
  - and, consequently what happens during anovulatory cycles?
- What drives irregularity in menstrual cycles in the absence of medical conditions and why a specific cycle is longer or shorter.

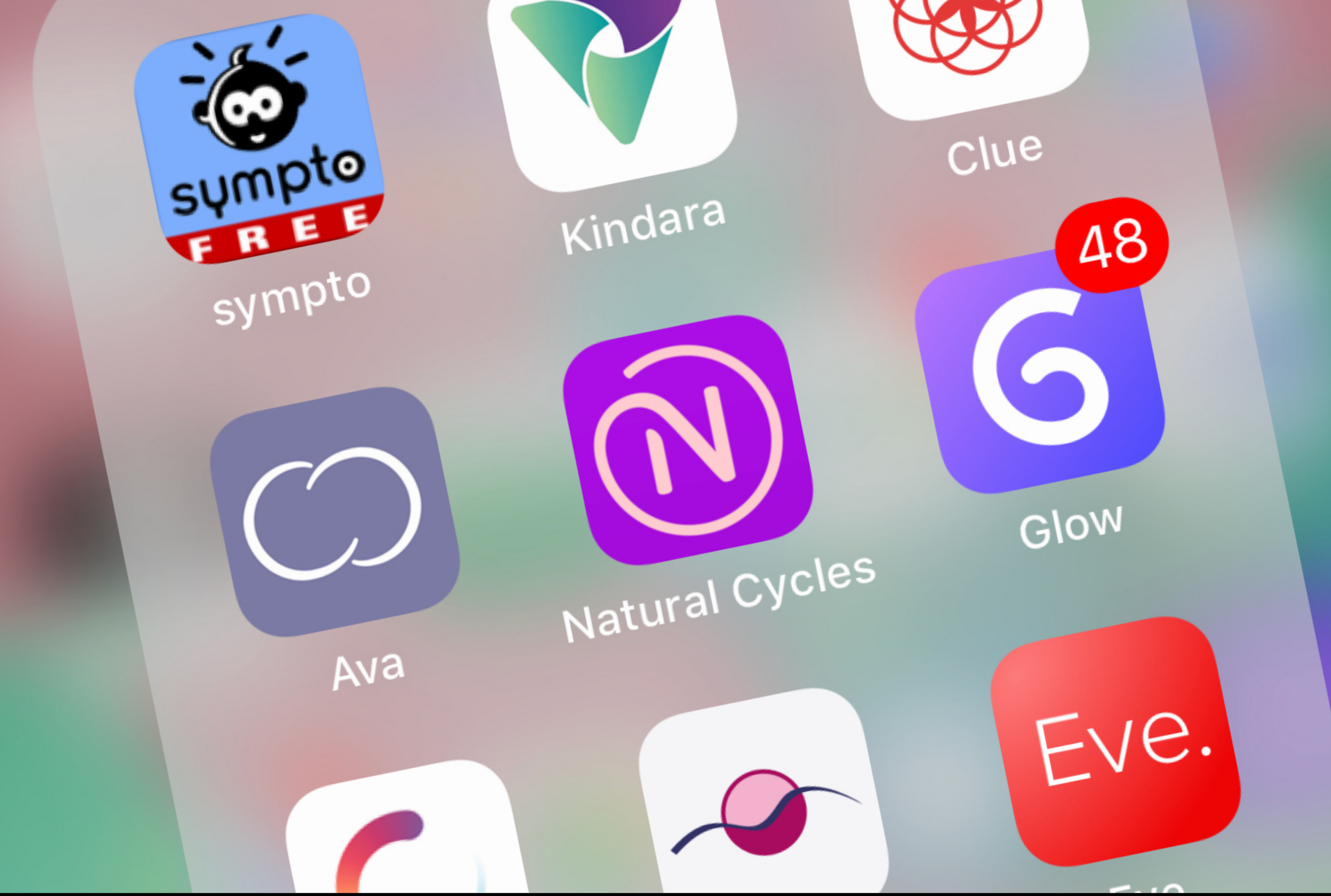
However…

The menstrual cycle has long been an
Understudied (mostly focusing on the fertility aspect)**,**
Under-funded (so little money to study endometriosis for example)
& quasi mono-disciplinary (mostly research by obstetricians & gynecologists, where the circadian clock, another fundamental biological rhythm, has been heavily studied from a medical, biological, ecological, modeling perspective)
field of research

We still do not know…

- The causes of pre-menstrual symptoms and pre-menstrual dysphoric disorder
  - A genetic predisposition?
  - An hormonal sensitivity? General? Tissue-specific?
  - An abnormality of the immune system? In the interactions with the immune system?
  - An evolutionary strategy to lead infertile couples to break-ups?
- The drivers of human birth seasonality? Which aspects of fertility are seasonal?
- How menstrual bleeding is regulated
  - and, consequently what happens during anovulatory cycles?
- What drives irregularity in menstrual cycles in the absence of medical conditions and why a specific cycle is longer or shorter.

Data?



Millions of women now use apps & wearables to track their cycle

Long term tracking

Long term tracking of 2 app users (top user likely sub-fertile; bottom users: pregnancy)

(Machine) Learning about Cats&Dogs

(Machine) Learning about Cats&Dogs

Deep Learning

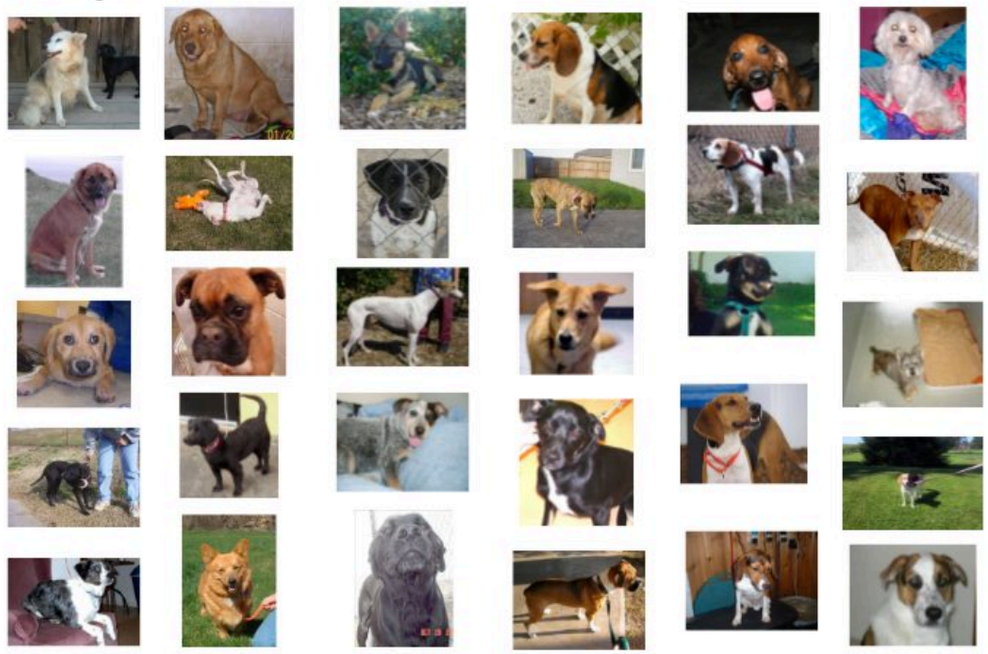
Human-like Learning

(Machine) Learning about Cats&Dogs

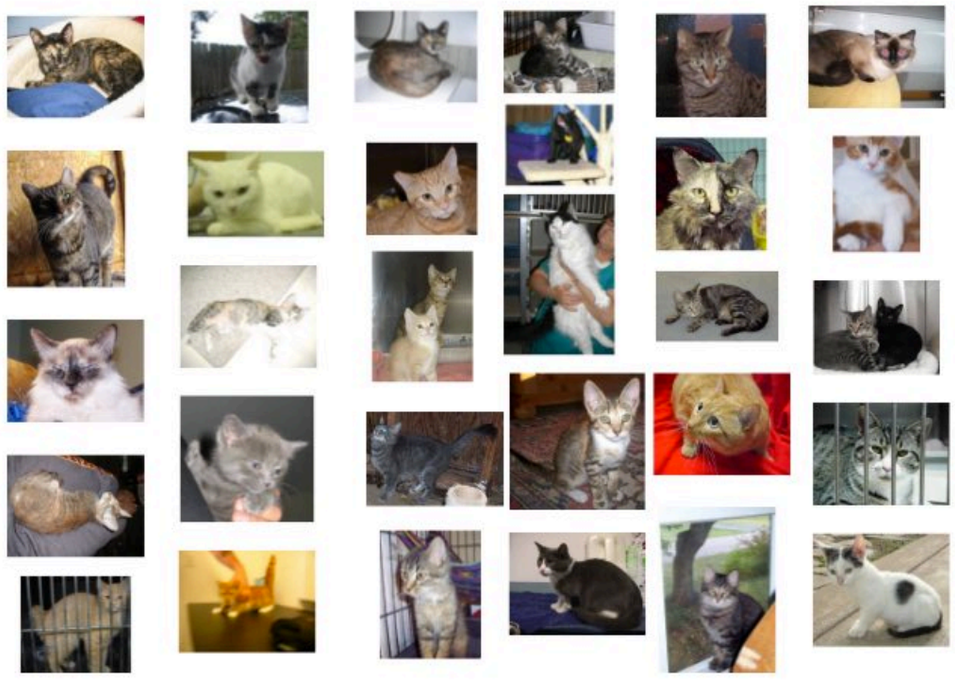
Deep Learning

A TON of labeled examples

Dogs



Cats



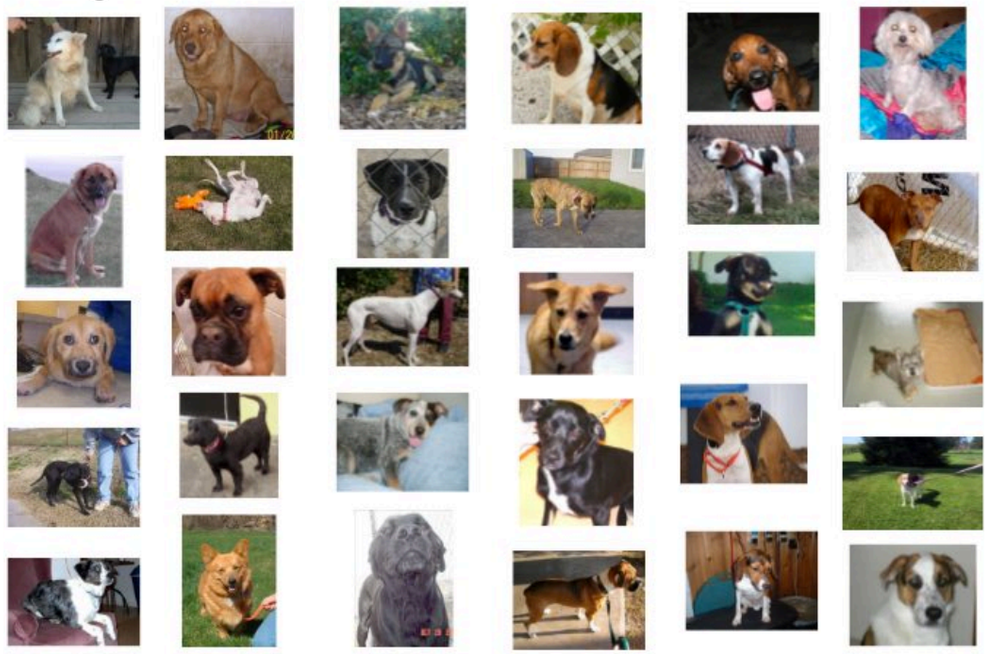
Human-like Learning

(Machine) Learning about Cats&Dogs

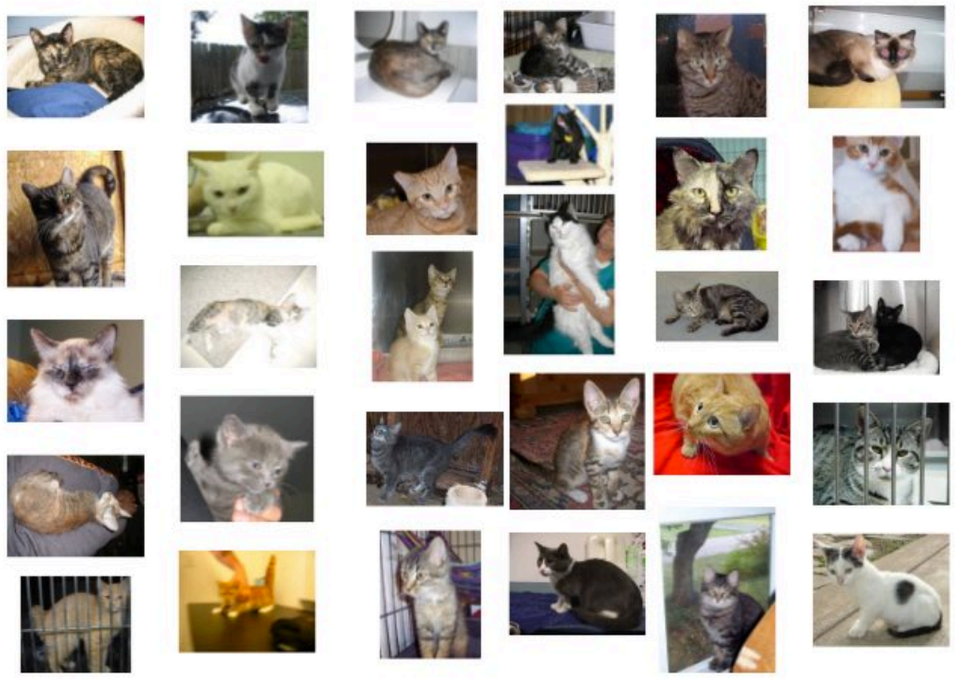
Deep Learning

A TON of labeled examples

Dogs



Cats



Human-like Learning

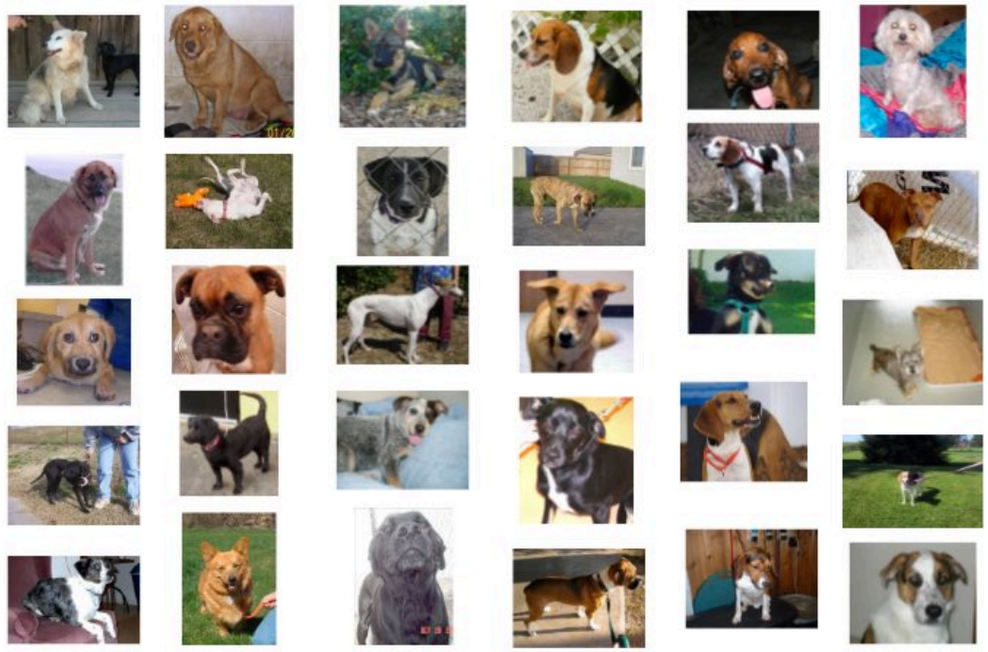
Facts, contrasting information

(Machine) Learning about Cats&Dogs

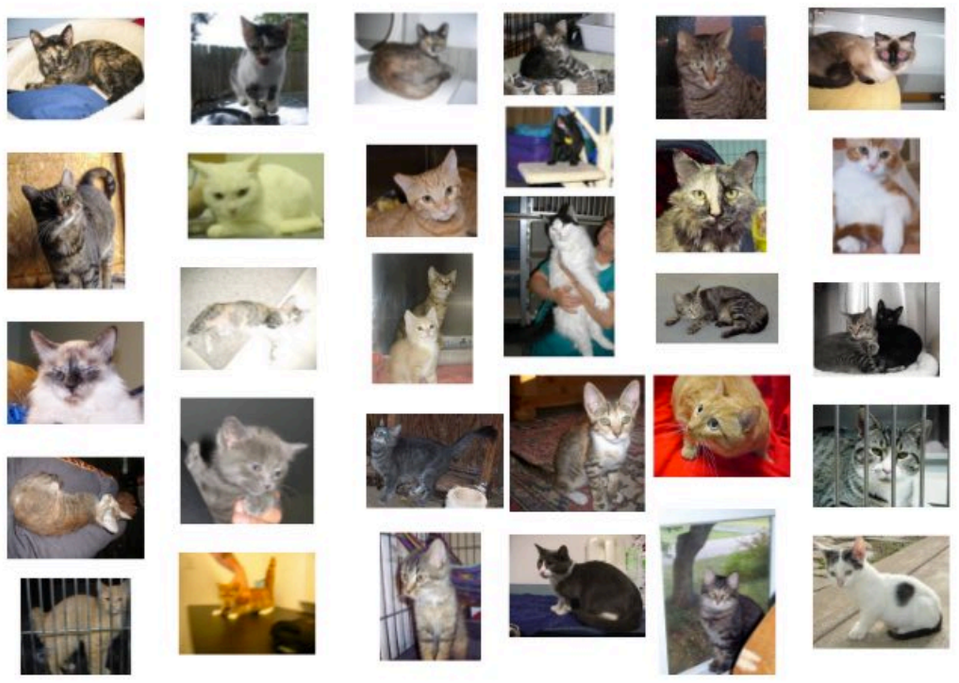
Deep Learning

A TON of labeled examples

Dogs



Cats



Human-like Learning

Facts, contrasting information

Dogs

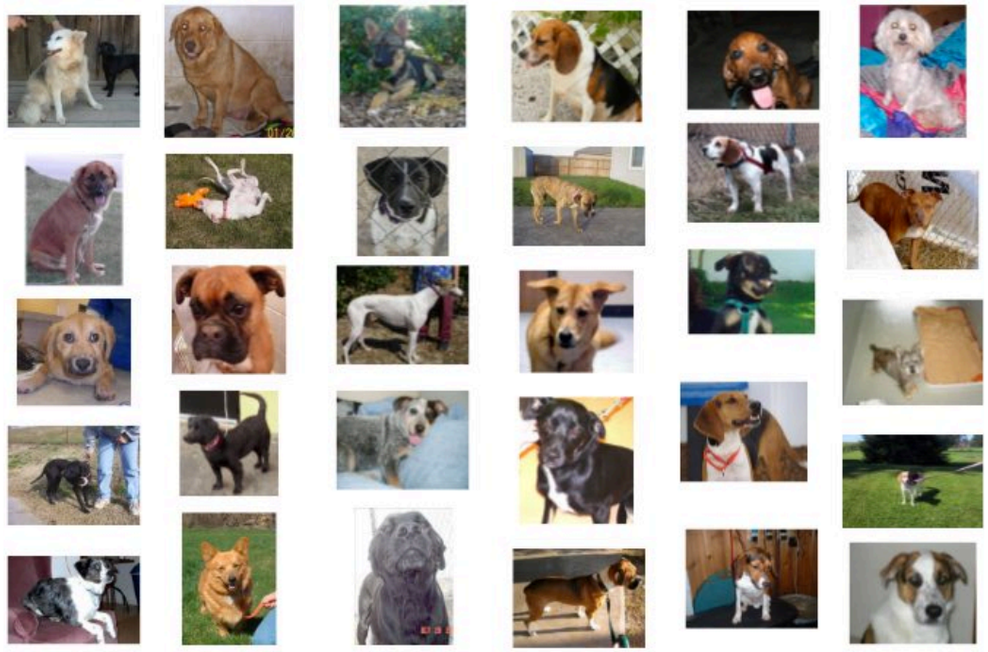
- 4 legged animals (mammals)
- Size: from ~20 cm to ~1m (large diversity in sizes)
- Hair type: short hair, long hair or fluffy curly hair
- Hair Color: mostly shades of brown, but also white or black
- Most dogs have long nose with black truffle
- Ears: mostly falling on the side of head, except when paying attention to smth

(Machine) Learning about Cats&Dogs

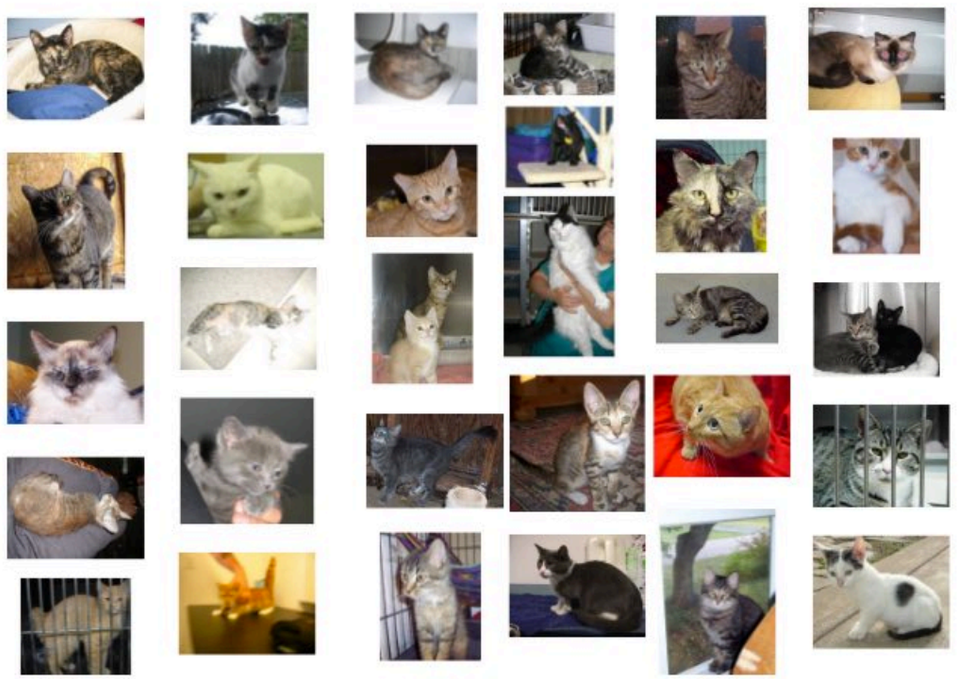
Deep Learning

A TON of labeled examples

Dogs



Cats



Human-like Learning

Facts, contrasting information

Dogs

- 4 legged animals (mammals)
- Size: from ~20 cm to ~1m (large diversity in sizes)
- Hair type: short hair, long hair or fluffy curly hair
- Hair Color: mostly shades of brown, but also white or black
- Most dogs have long nose with black truffle
- Ears: mostly falling on the side of head, except when paying attention to smth

Cats

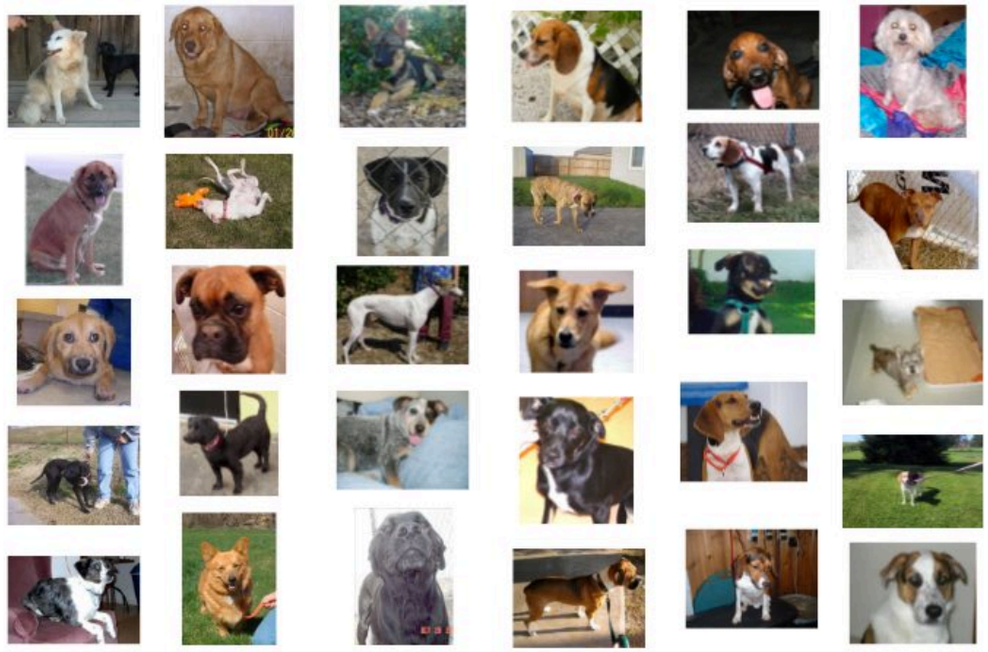
- 4 legged animals (mammals)
- Size: 25cm ± 5cm (smaller variance in size than dogs)
- Hair type: mostly short hair, sometimes long, can be fluffy (but not curly)
- Hair Color: white, black, gray, warm brown. “Tigger pattern"
- Most cats have round faces with short noses
- Ears are pointy and mostly straight up.

(Machine) Learning about Cats&Dogs

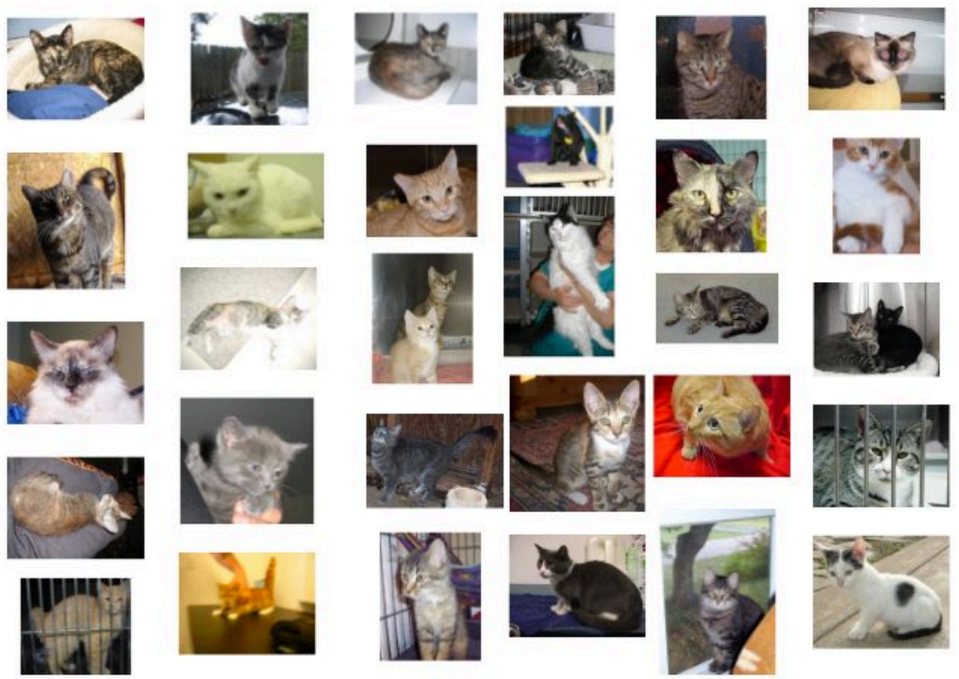
Deep Learning

A TON of labeled examples

Dogs



Cats



Human-like Learning

Facts, contrasting information

Dogs

- 4 legged animals (mammals)
- Size: from ~20 cm to ~1m (large diversity in sizes)
- Hair type: short hair, long hair or fluffy curly hair
- Hair Color: mostly shades of brown, but also white or black
- Most dogs have long nose with black truffle
- Ears: mostly falling on the side of head, except when paying attention to smth

Cats

- 4 legged animals (mammals)
- Size: 25cm ± 5cm (smaller variance in size than dogs)
- Hair type: mostly short hair, sometimes long, can be fluffy (but not curly)
- Hair Color: white, black, gray, warm brown. "Tigger pattern"
- Most cats have round faces with short noses
- Ears are pointy and mostly straight up.

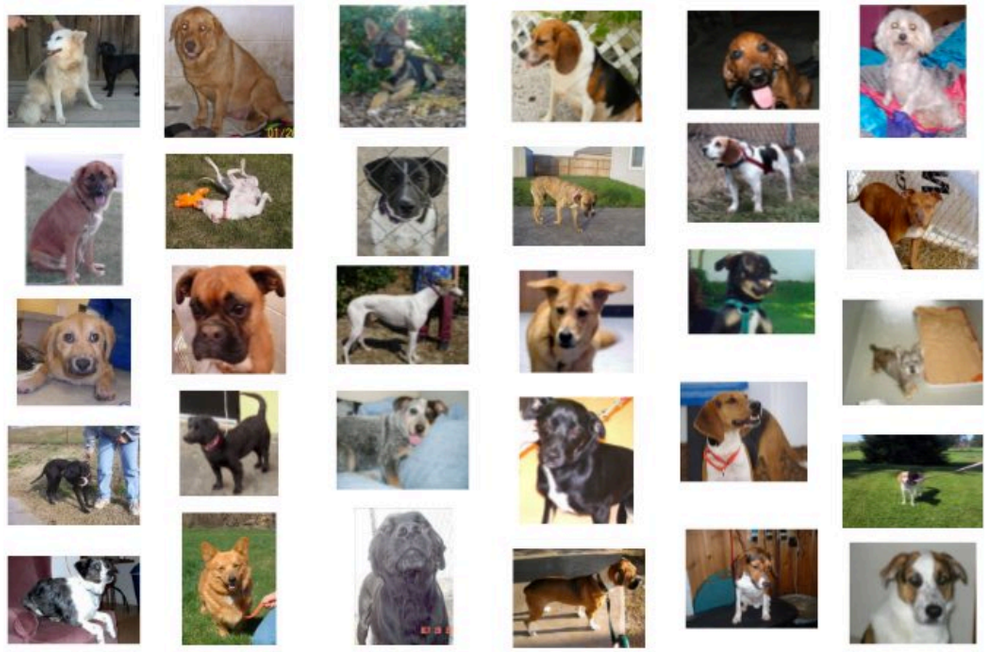
A few examples

(Machine) Learning about Cats&Dogs

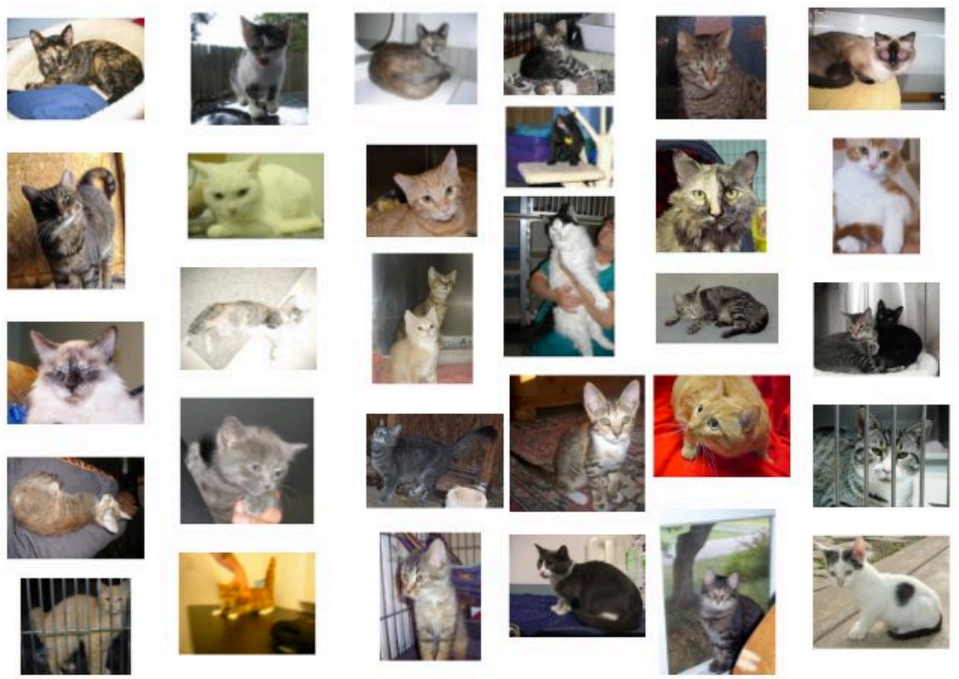
Deep Learning

A TON of labeled examples

Dogs



Cats



Human-like Learning or **Model-based Stat L**

Facts, contrasting information

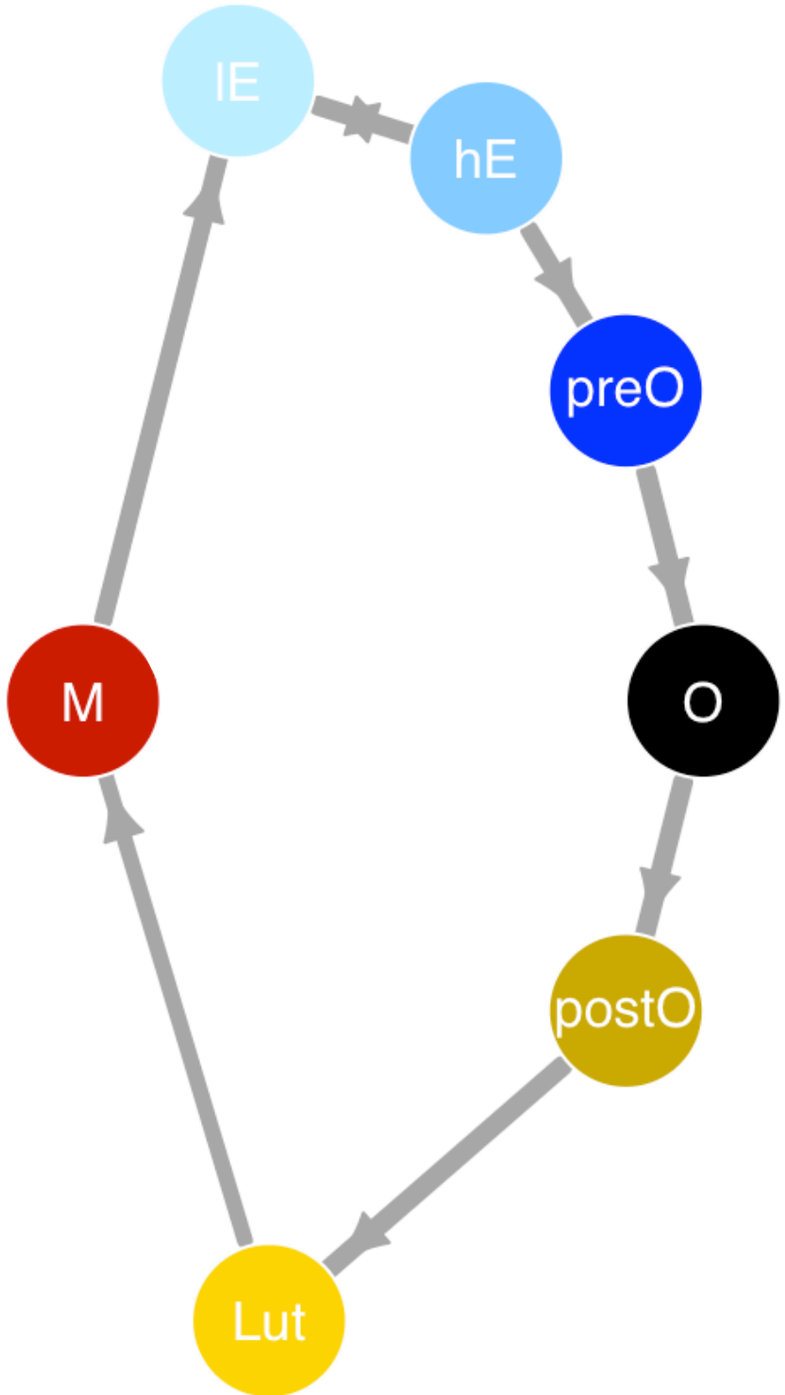
Dogs

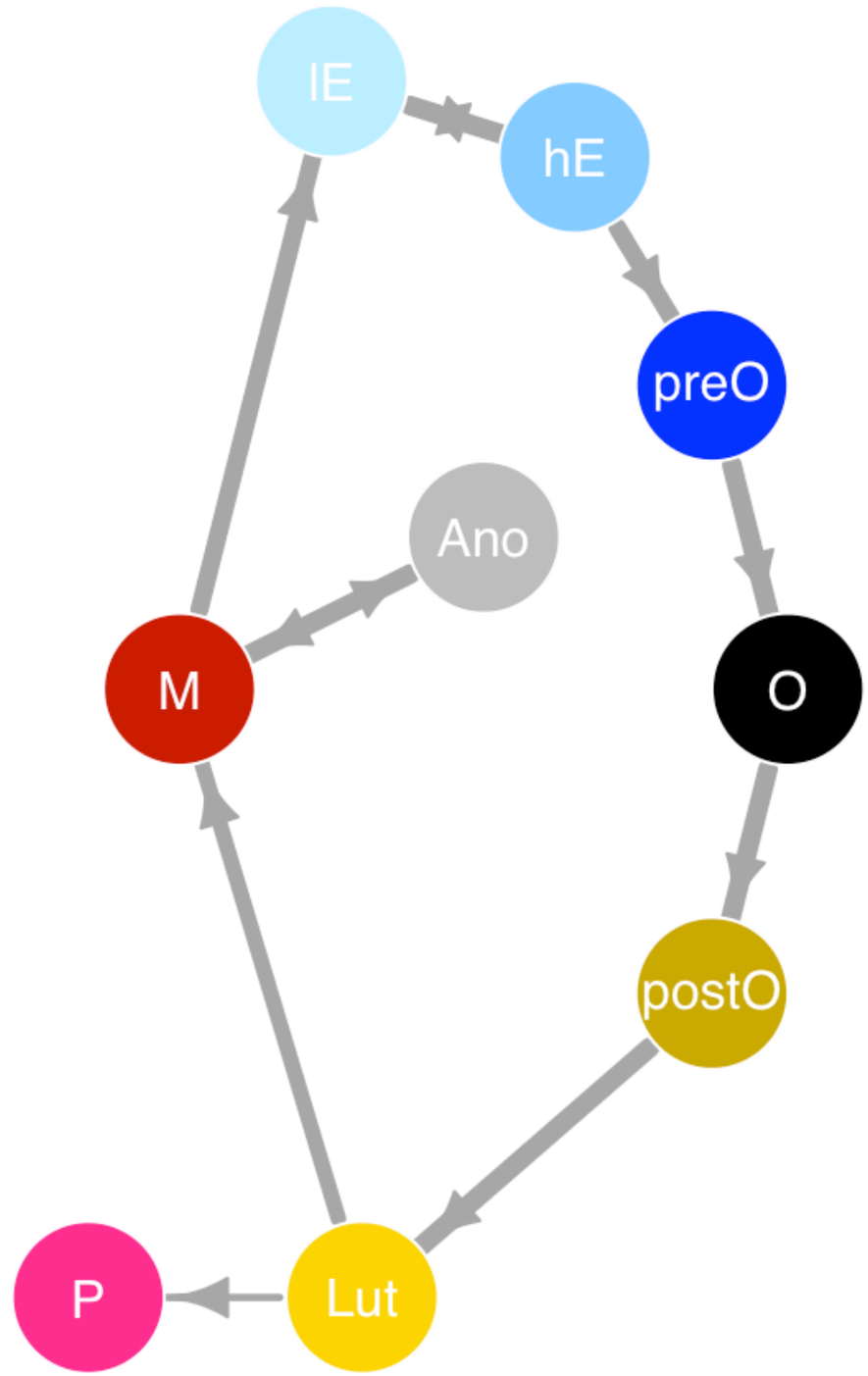
- 4 legged animals (mammals)
- Size: from ~20 cm to ~1m (large diversity in sizes)
- Hair type: short hair, long hair or fluffy curly hair
- Hair Color: mostly shades of brown, but also white or black
- Most dogs have long nose with black truffle
- Ears: mostly falling on the side of head, except when paying attention to smth

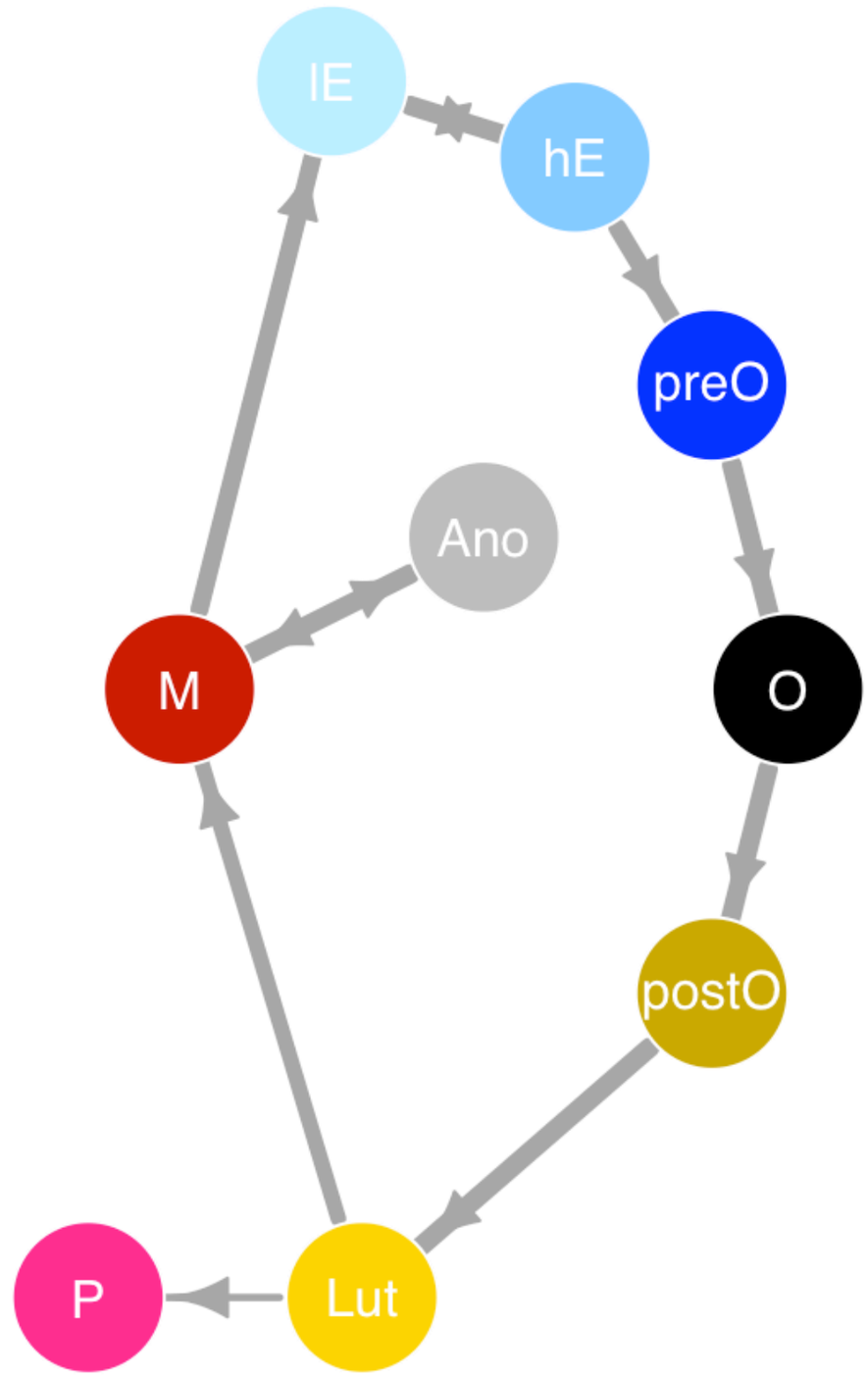
Cats

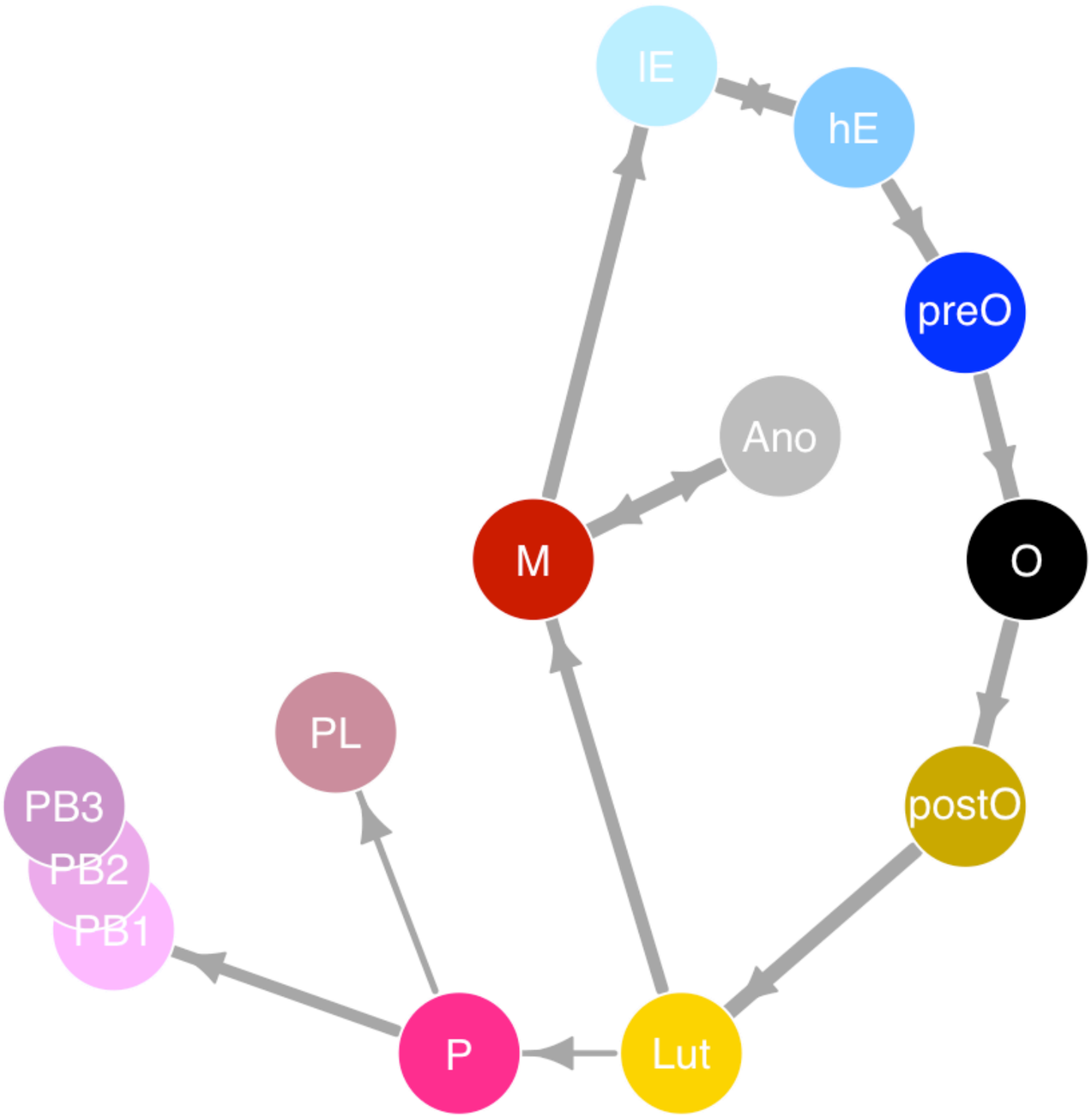
- 4 legged animals (mammals)
- Size: 25cm ± 5cm (smaller variance in size than dogs)
- Hair type: mostly short hair, sometimes long, can be fluffy (but not curly)
- Hair Color: white, black, gray, warm brown. "Tigger pattern"
- Most cats have round faces with short noses
- Ears are pointy and mostly straight up.

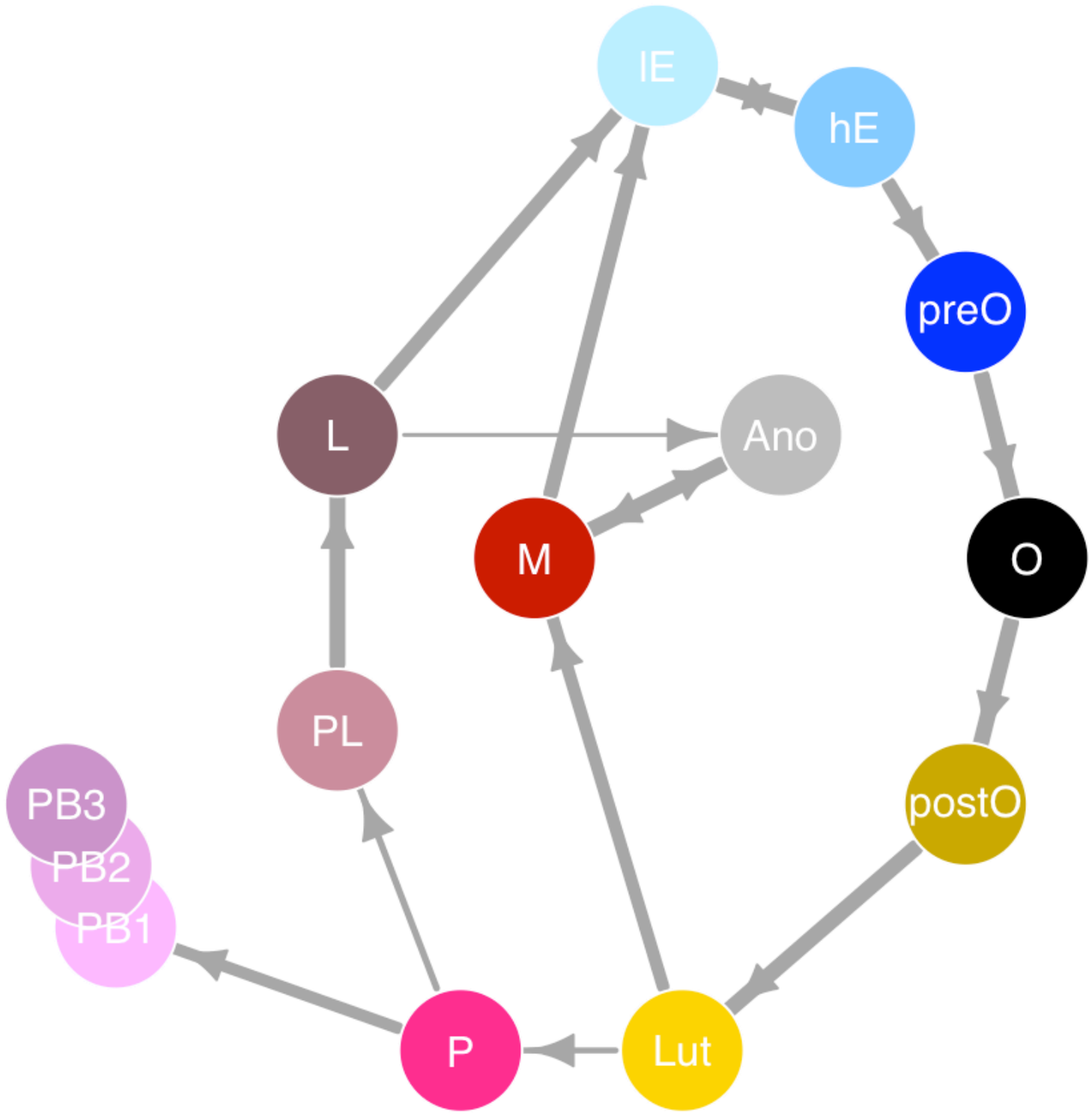
A few examples

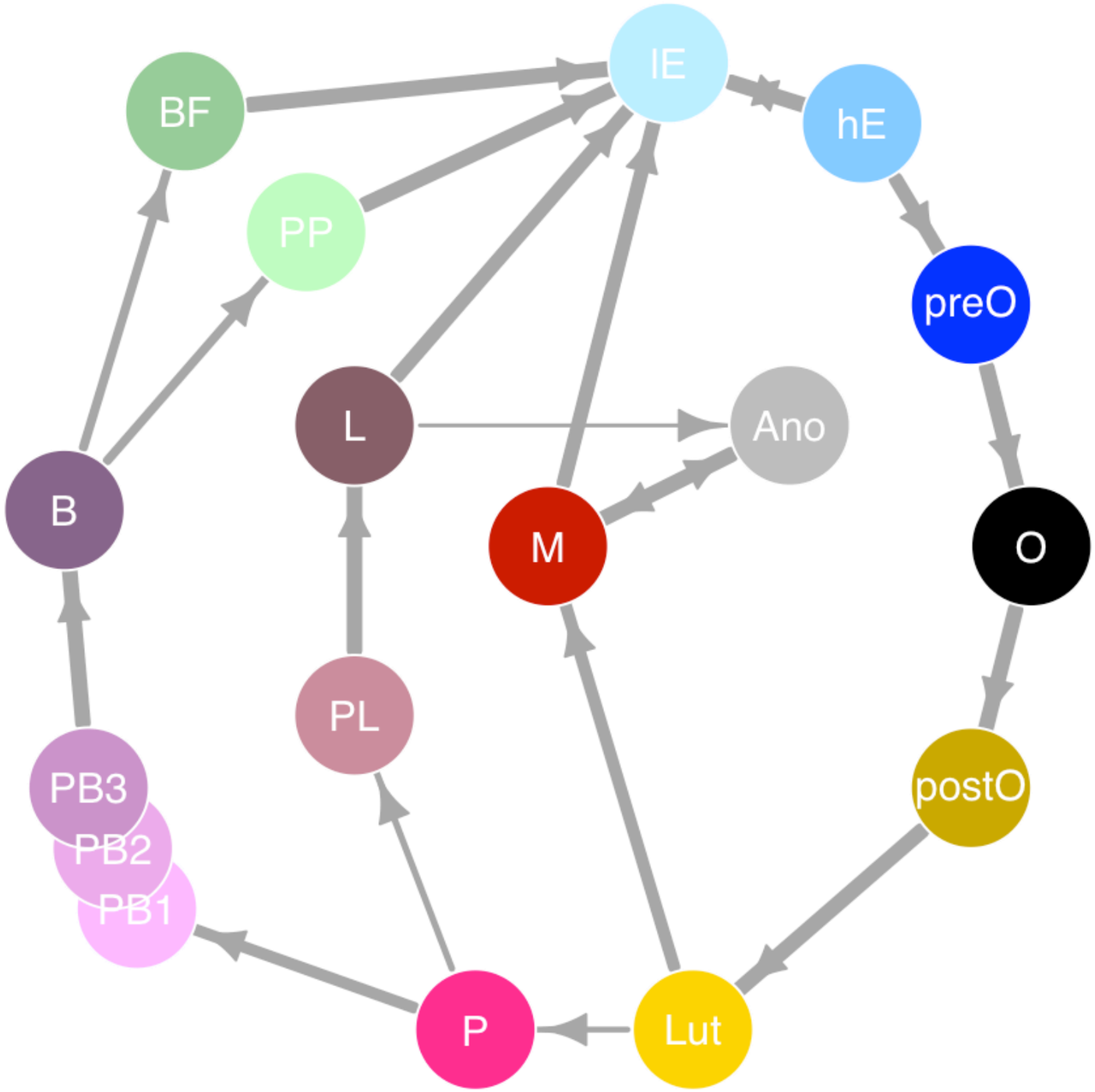

IE
hE
preO
M
O
postO
Lut

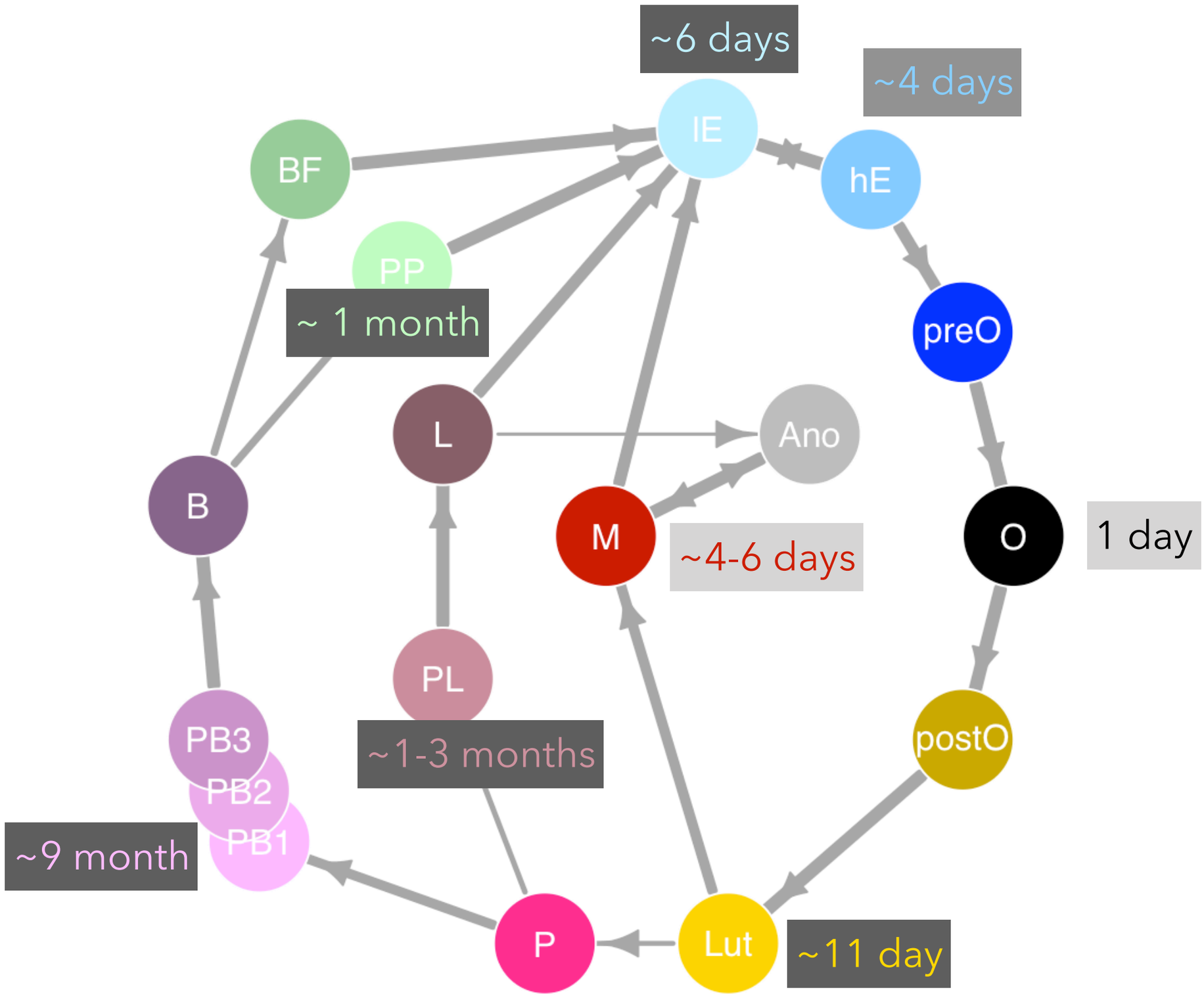

IE
hE
preO
Ano
M
O
postO
P
Lut

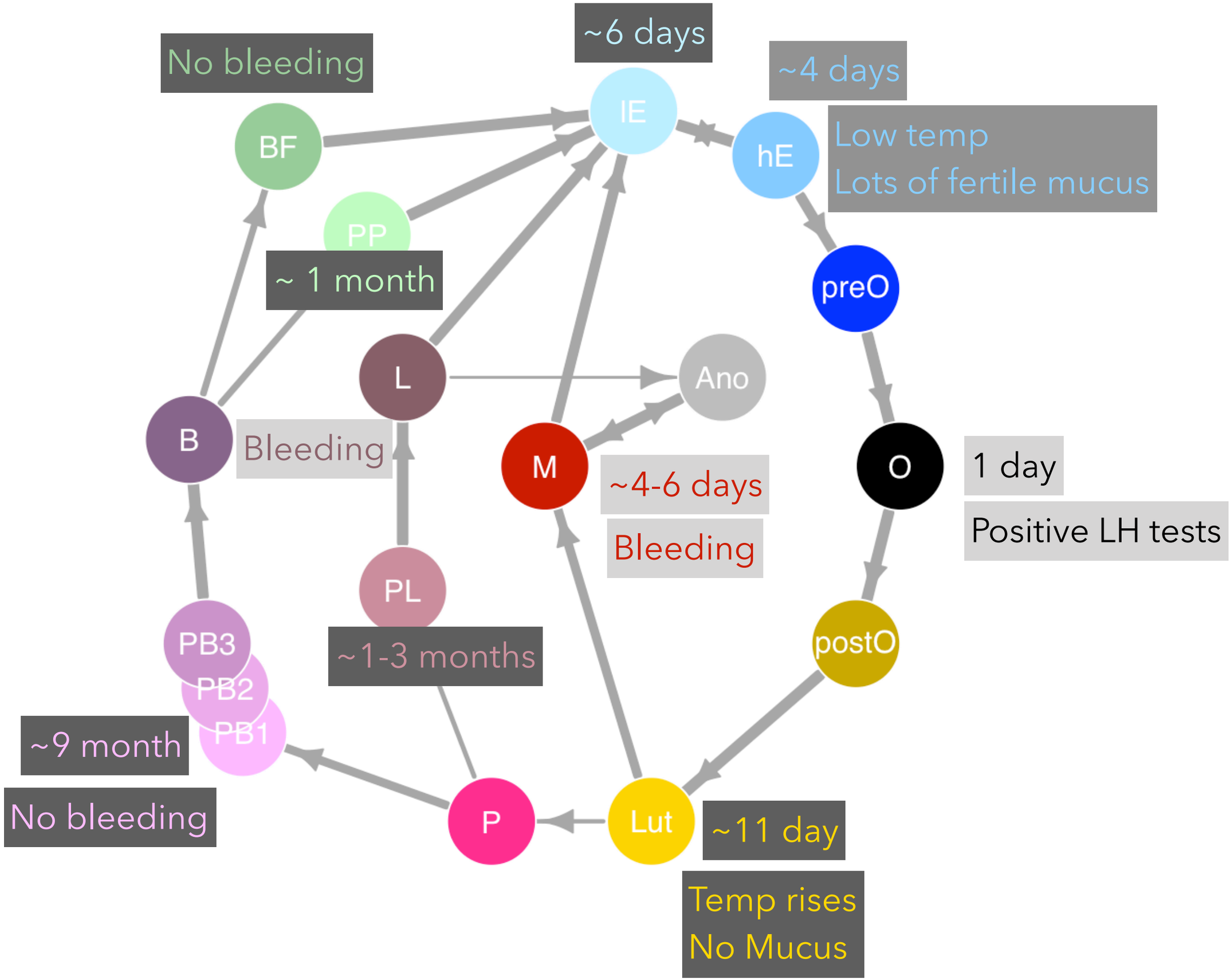

IE
hE
preO
Ano
M
O
postO
P
Lut

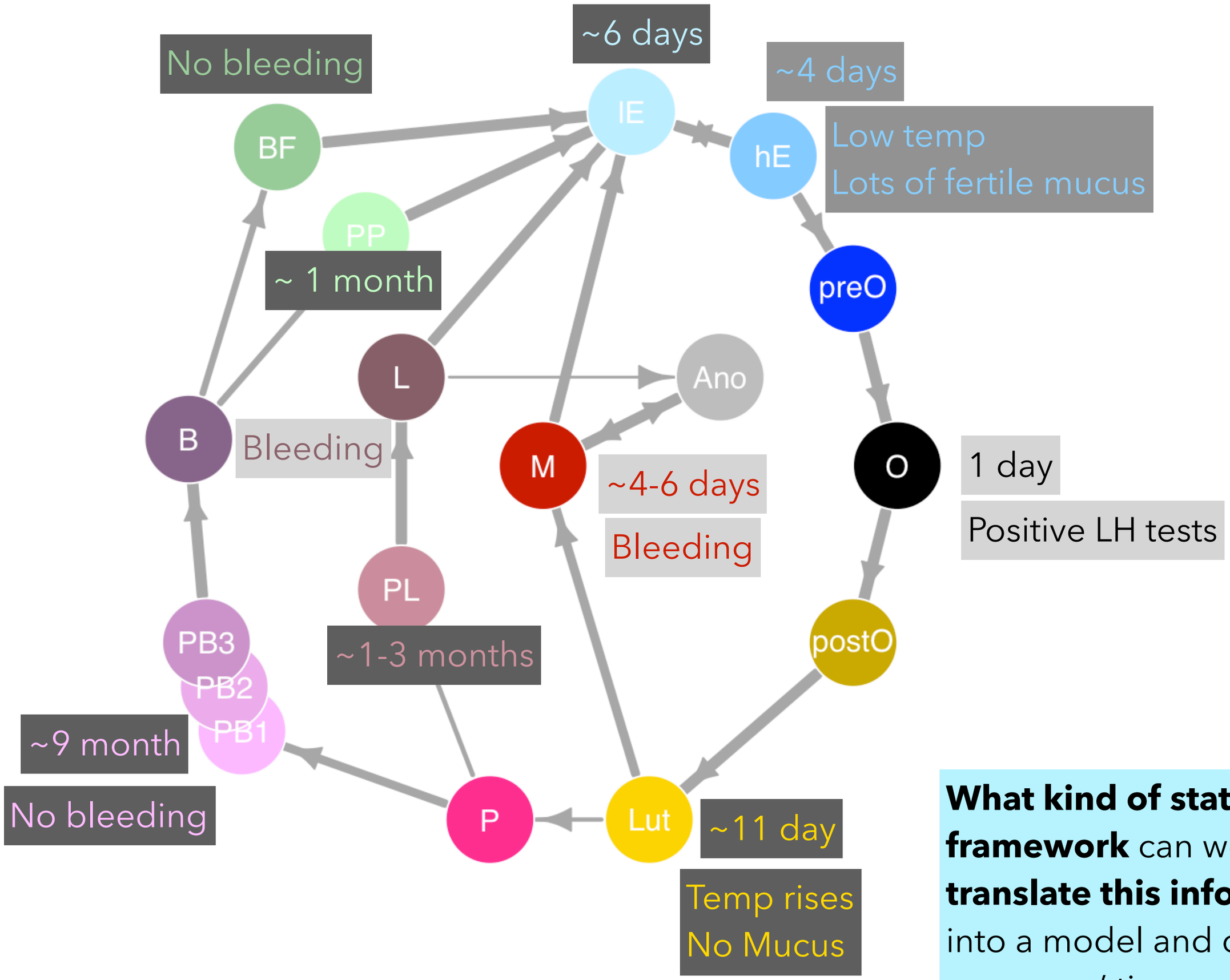

IE
hE
preO
Ano
M
O
PL
PB3
PB2
PB1
postO
P
Lut


IE
hE
preO
L
Ano
M
O
PL
postO
PB3
PB2
PB1
P
Lut


IE
BF
hE
PP
preO
L
Ano
B
M
O
PL
PB3
PB2
PB1
postO
P
Lut


~6 days
~4 days
IE
BF
hE
PP
~ 1 month
preO
L
Ano
B
M
~4-6 days
O
1 day
PL
~1-3 months
PB3
PB2
PB1
~9 month
postO
P
Lut
~11 day


~6 days
No bleeding
~4 days
IE
BF
hE
Low temp
Lots of fertile mucus
PP
~ 1 month
preO
L
Ano
B
Bleeding
M
O
1 day
~4-6 days
Positive LH tests
Bleeding
PL
PB3
~1-3 months
postO
PB2
PB1
~9 month
No bleeding
P
Lut
~11 day
Temp rises
No Mucus



What kind of statistical/ML framework can we use to **translate this information** into a model and decode app users' time-series?

Short introduction to **hidden semi-Markov models** (hsMm)

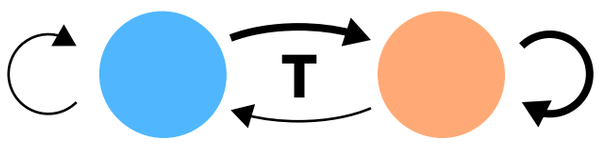
Short introduction to **hidden semi-Markov models** (hsMm)

Hidden Markov Models

Short introduction to **hidden semi-Markov models** (hsMm)

Hidden Markov Models

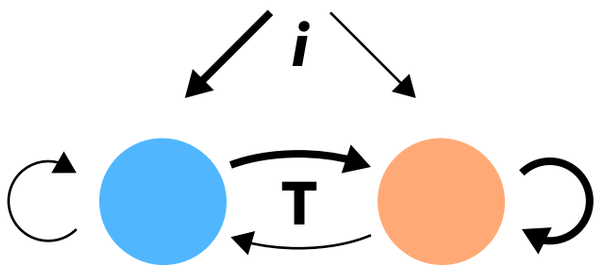
Hidden Markov Models



- **Transition** probabilities (matrix)
  - To switch states
  - To stay in same state

Short introduction to **hidden semi-Markov models** (hsMm)

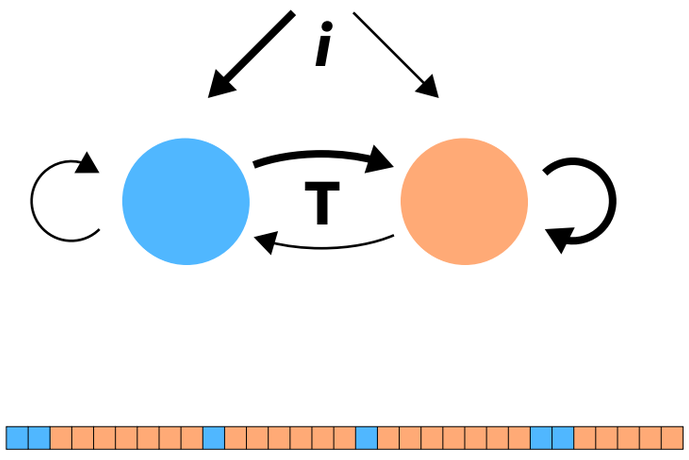
Hidden Markov Models



- **Transition** probabilities (matrix)
  - To switch states
  - To stay in same state

- **Initial state** probability vector

Hidden Markov Models

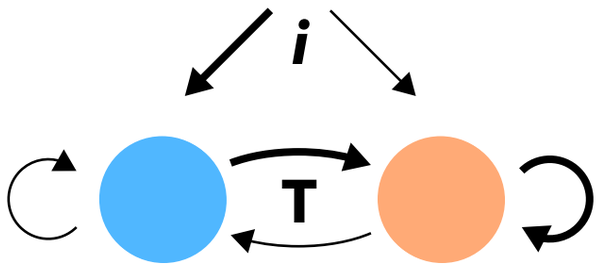


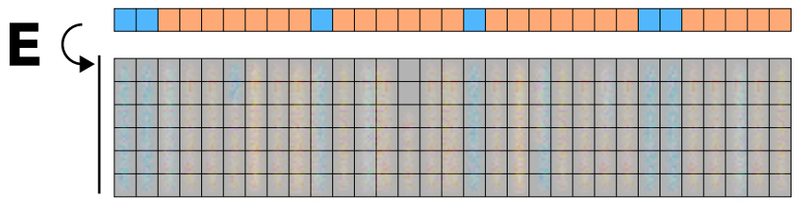
- **Transition** probabilities (matrix)
  - To switch states
  - To stay in same state

- **Initial state** probability vector

Short introduction to **hidden semi-Markov models** (hsMm)

Hidden Markov Models



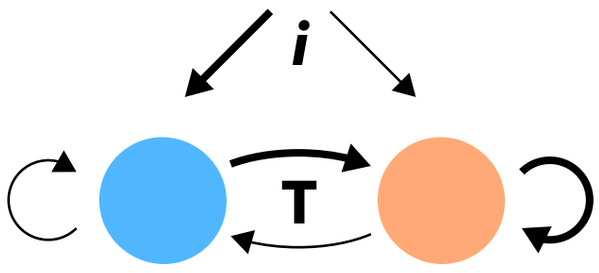


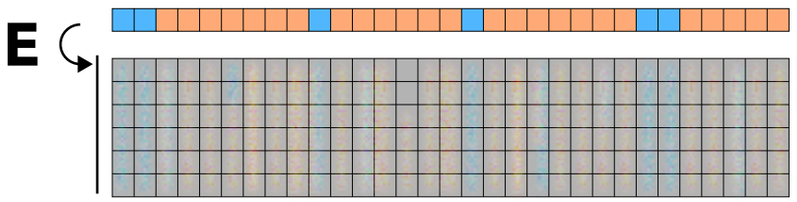
- **Transition** probabilities (matrix)
  - To switch states
  - To stay in same state

- **Initial state** probability vector
- **Emission probabilities** for the observed variables

Short introduction to **hidden semi-Markov models** (hsMm)

Hidden Markov Models





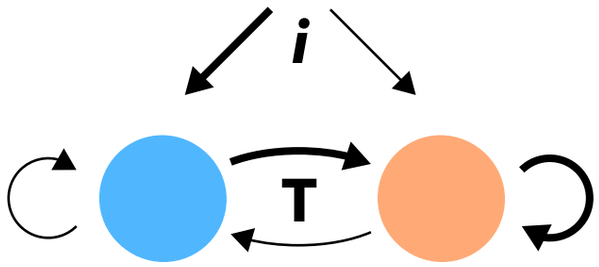
- **Transition** probabilities (matrix)
  - To switch states
  - To stay in same state

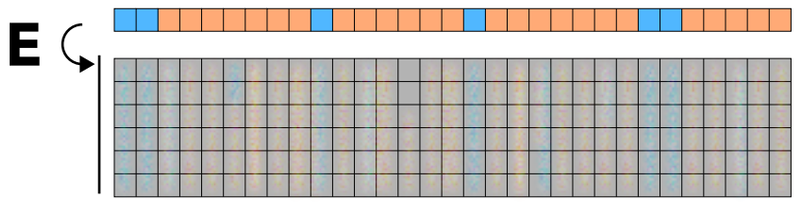
- **Initial state** probability vector
- **Emission probabilities** for the observed variables

Hidden semi-Markov Models

Short introduction to **hidden semi-Markov models** (hsMm)

Hidden Markov Models

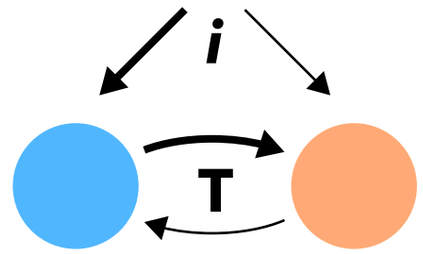




- **Transition** probabilities (matrix)
  - To switch states
  - To stay in same state

- **Initial state** probability vector
- **Emission probabilities** for the observed variables

Hidden semi-Markov Models

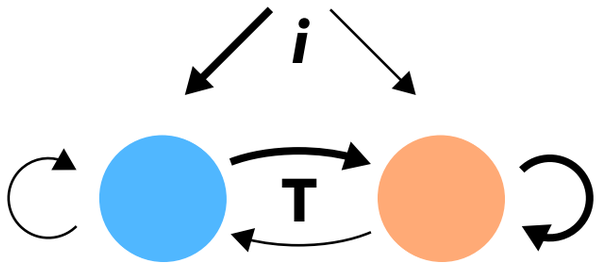


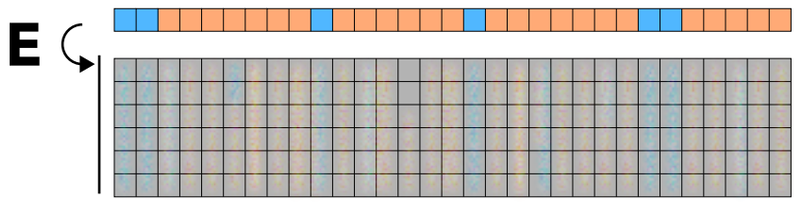
- **Transition** probabilities (matrix)
  - To switch states
  - ~~To stay in same state~~

- **Initial state** probability vector

Short introduction to **hidden semi-Markov models** (hsMm)

Hidden Markov Models

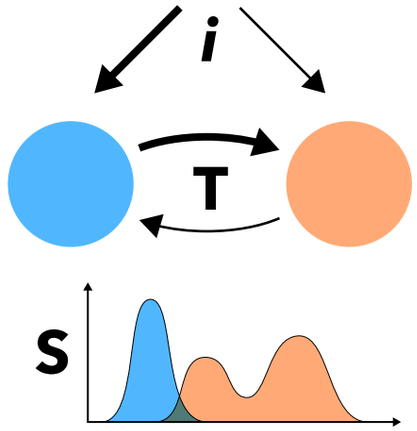




- **Transition** probabilities (matrix)
  - To switch states
  - To stay in same state

- **Initial state** probability vector
- **Emission probabilities** for the observed variables

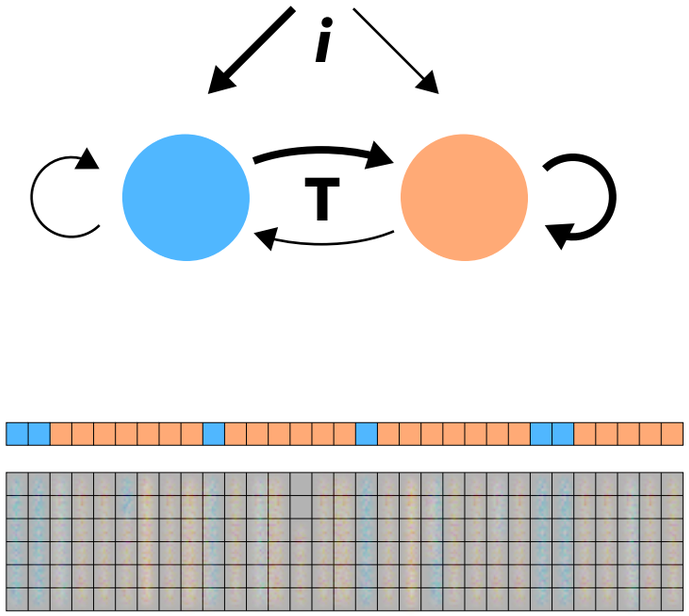
Hidden semi-Markov Models



- **Transition** probabilities (matrix)
  - To switch states
  - ~~To stay in same state~~
- **Sojourn** distributions:
  the time spent in a given state
- **Initial state** probability vector

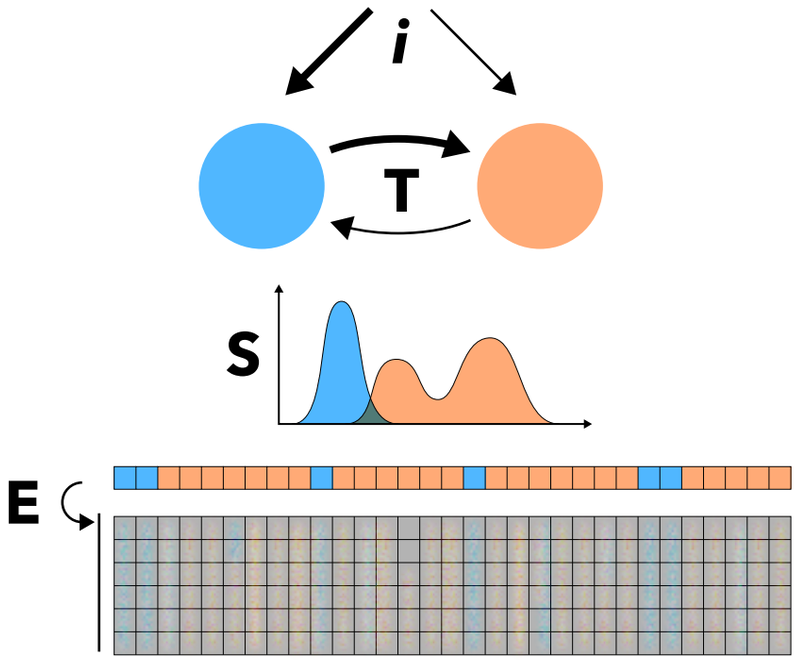
Short introduction to **hidden semi-Markov models** (hsMm)

Hidden Markov Models



- **Transition** probabilities (matrix)
  - To switch states
  - To stay in same state
- **Initial state** probability vector
- **Emission probabilities** for the observed variables

Hidden semi-Markov Models

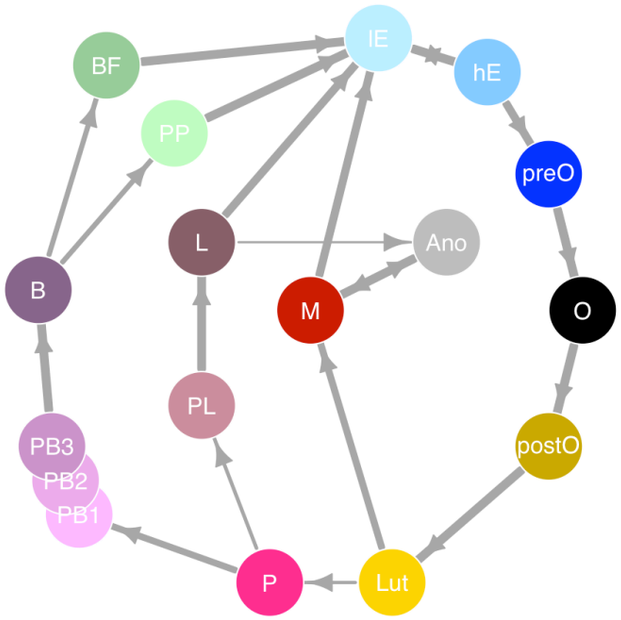


- **Transition** probabilities (matrix)
  - To switch states
  - ~~To stay in same state~~
- **Sojourn** distributions:
  the time spent in a given state
- **Initial state** probability vector
- **Emission probabilities** for the observed variables

Workflow

Workflow

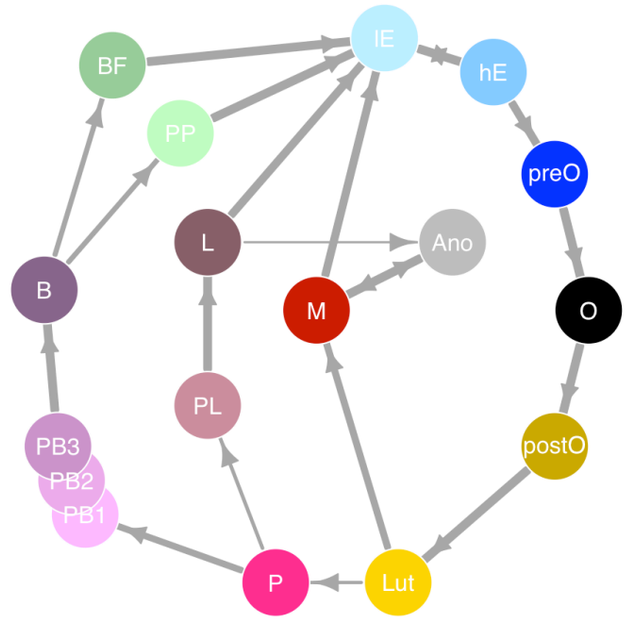
- Initialize the model architecture
  - The transition probabilities
  - The sojourn distributions
  - The emission probabilities



Workflow

- Initialize the model architecture
  - The transition probabilities
  - The sojourn distributions
  - The emission probabilities

- *Manually label a few time-series* [R shiny app]

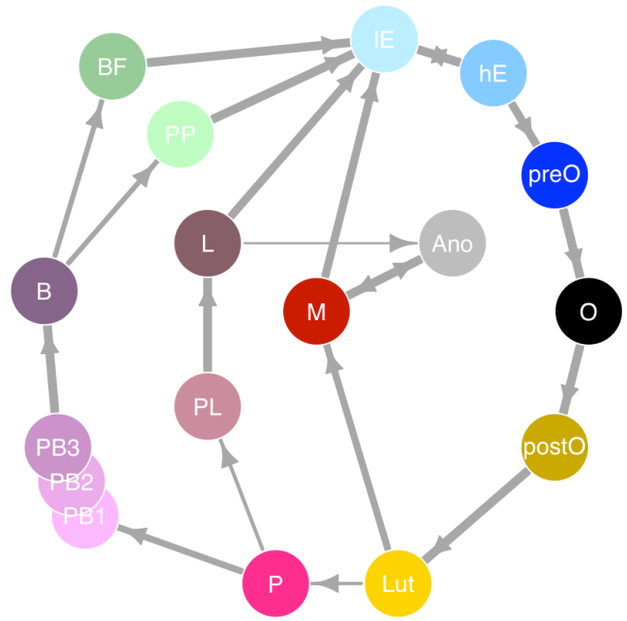


Workflow

- Initialize the model architecture
  - The transition probabilities
  - The sojourn distributions
  - The emission probabilities

- *Manually label a few time-series* [R shiny app]

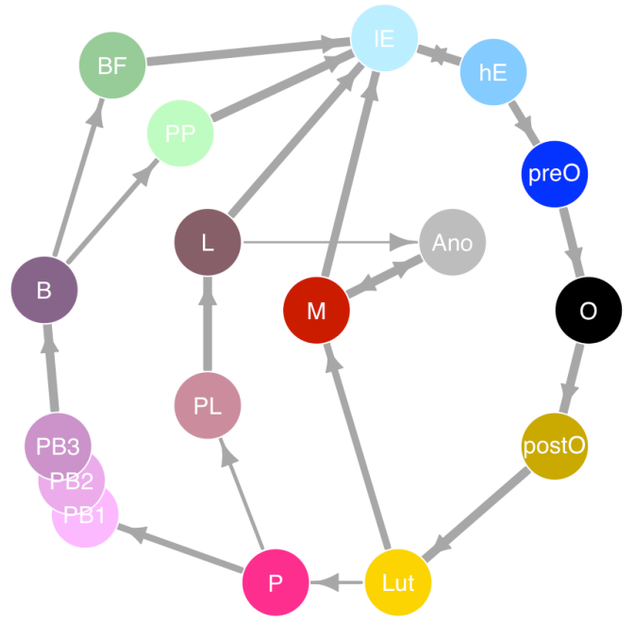
- **Fit the model** (EM approach)
  & **decode** the time-series with the fitted model
- Evaluate the **accuracy** (using the labelled t-s)
  & **confidence** (parametric bootstrap approach)
  in the decoded sequence
- [R shiny app] *Manually label more time-series (where the decoding is wrong)*
  & validate the decoding (where it got it right)



Workflow

- Initialize the model architecture
  - The transition probabilities
  - The sojourn distributions
  - The emission probabilities

- *Manually label a few time-series* [R shiny app]



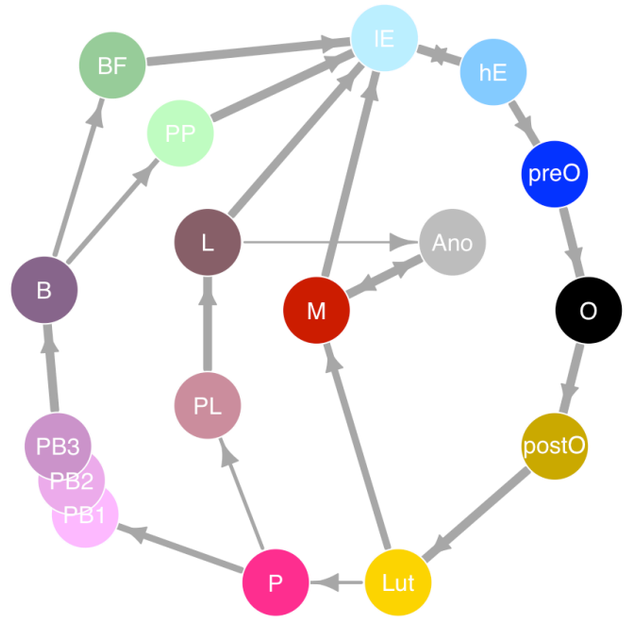
- **Fit the model** (EM approach)
  & **decode** the time-series with the fitted model
- Evaluate the **accuracy** (using the labelled t-s)
  & **confidence** (parametric bootstrap approach)
  in the decoded sequence
- [R shiny app] *Manually label more time-series (where the decoding is wrong)*
  & validate the decoding (where it got it right)

Interactive boosting

Workflow

- Initialize the model architecture
  - The transition probabilities
  - The sojourn distributions
  - The emission probabilities

- *Manually label a few time-series* [R shiny app]



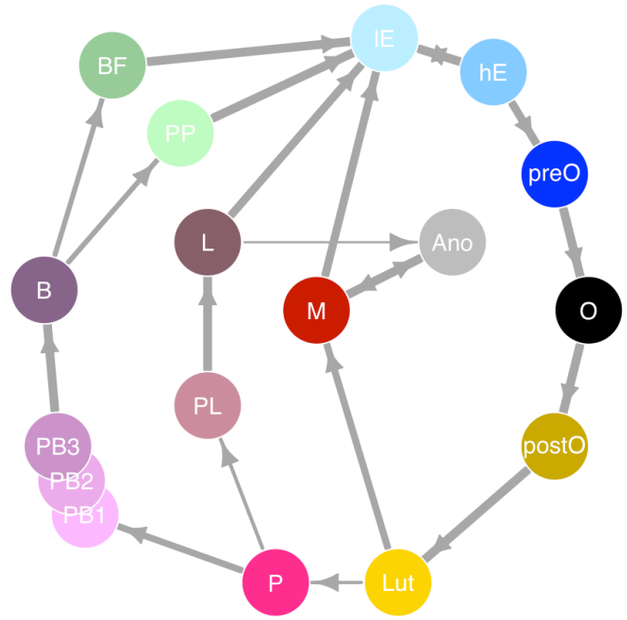
- **Fit the model** (EM approach)
  & **decode** the time-series with the fitted model
- Evaluate the **accuracy** (using the labelled t-s)
  & **confidence** (parametric bootstrap approach)
  in the decoded sequence
- [R shiny app] *Manually label more time-series (where the decoding is wrong)*
  & validate the decoding (where it got it right)

Interactive boosting

- Decode all user's time-series

Workflow

- Initialize the model architecture
  - The transition probabilities
  - The sojourn distributions
  - The emission probabilities



- *Manually label a few time-series* [R shiny app]

- **Fit the model** (EM approach)
  & **decode** the time-series with the fitted model
- Evaluate the **accuracy** (using the labelled t-s)
  & **confidence** (parametric bootstrap approach)
  in the decoded sequence
- [R shiny app] *Manually label more time-series (where the decoding is wrong)*
  & validate the decoding (where it got it right)

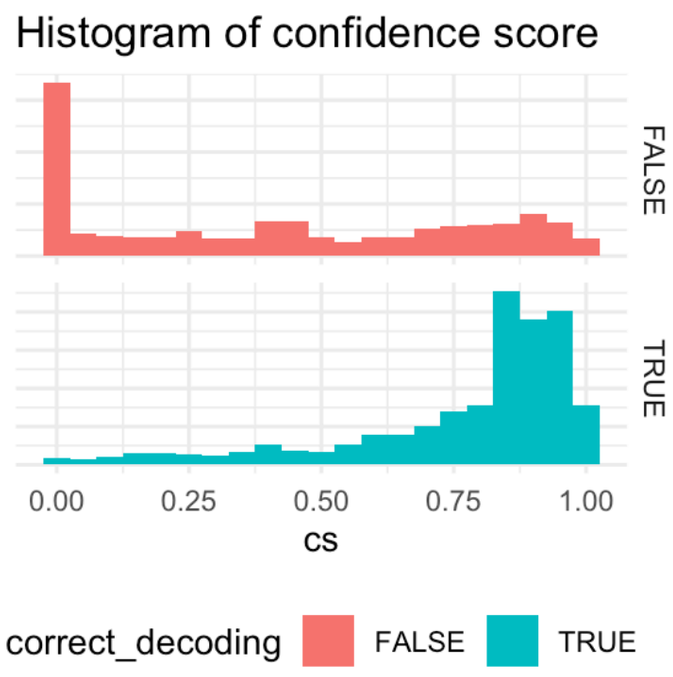
Interactive boosting

- Decode all user's time-series

Epidemiological studies

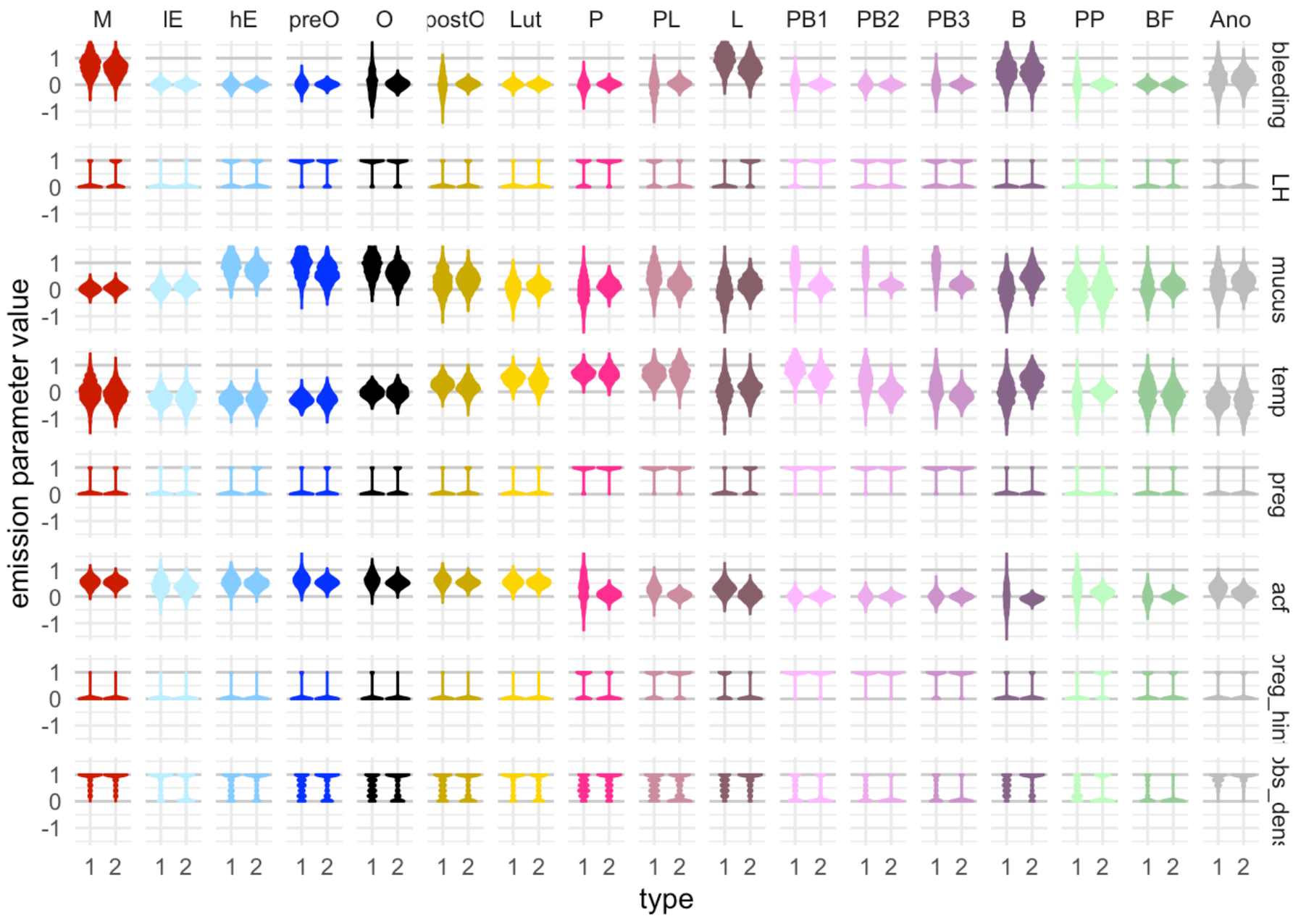
Results

Confidence in the decoded sequences

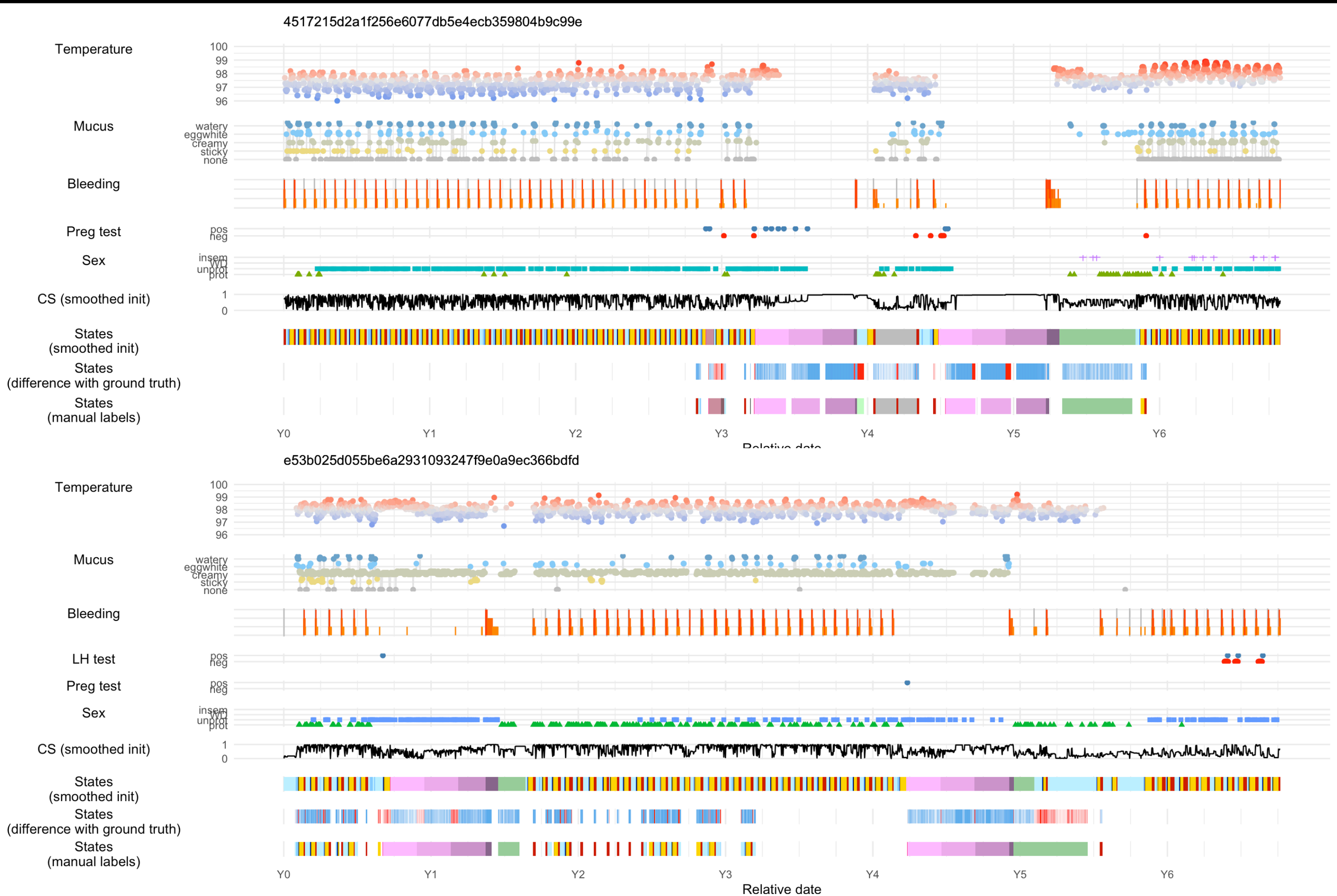


Accuracy ~80%

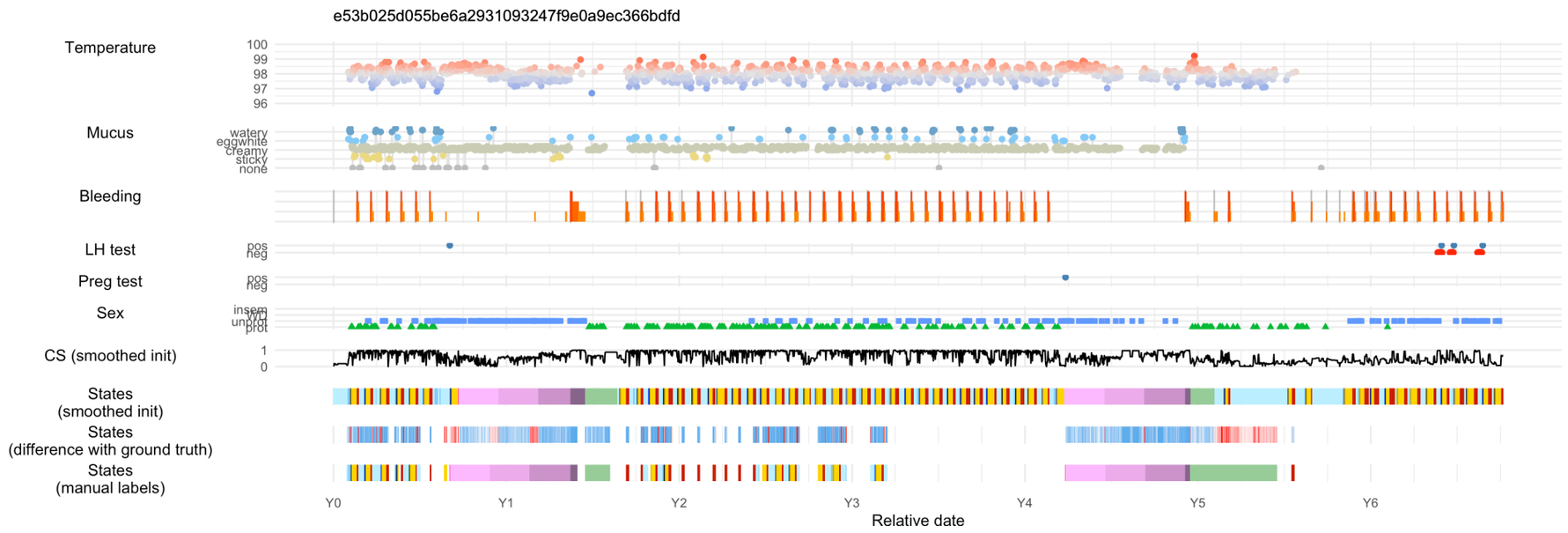
Emission parameters



Results


4517215d2a1f256e6077db5e4ecb359804b9c99e
Temperature
100
99
98
97
96
Mucus
watery
eggwhite
creamy
sticky
none
Bleeding
Preg test
pos
neg
Sex
insem
unprot
prot
CS (smoothed init)
1
0
States (smoothed init)
States (difference with ground truth)
States (manual labels)
Y0
Y1
Y2
Y3
Y4
Y5
Y6
e53b025d055be6a2931093247f9e0a9ec366bdfd
LH test
Relative date

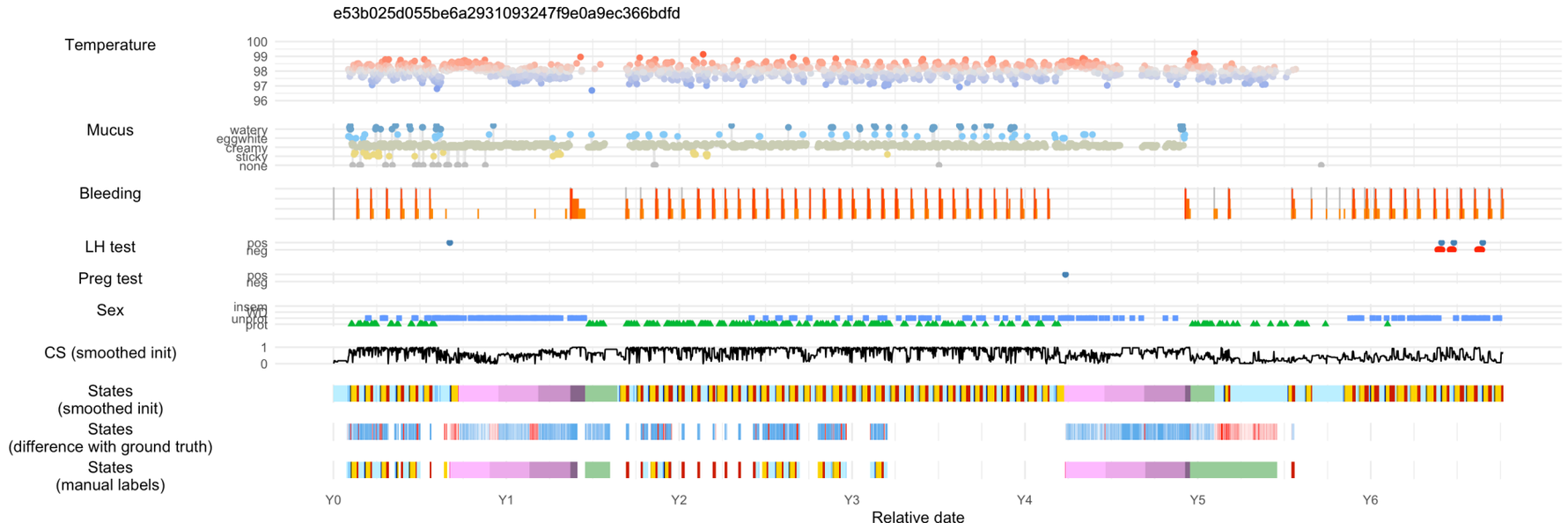
Conclusions



Upgraded `'hsmm'` R package that

- **can deal with high level of data missingness** (as found in any digital health project)
- **allows for any parametric or non-parametric distribution** for the **observed variables**
  - which allow for **discrete**, **categorical** or **continuous** variables
- can **quantify uncertainty** by providing a **confidence score** for the decoded sequence, based on **parametric bootstrap**, **independent of whether labelled data are available**
- **Available on my GitHub: @lasy**

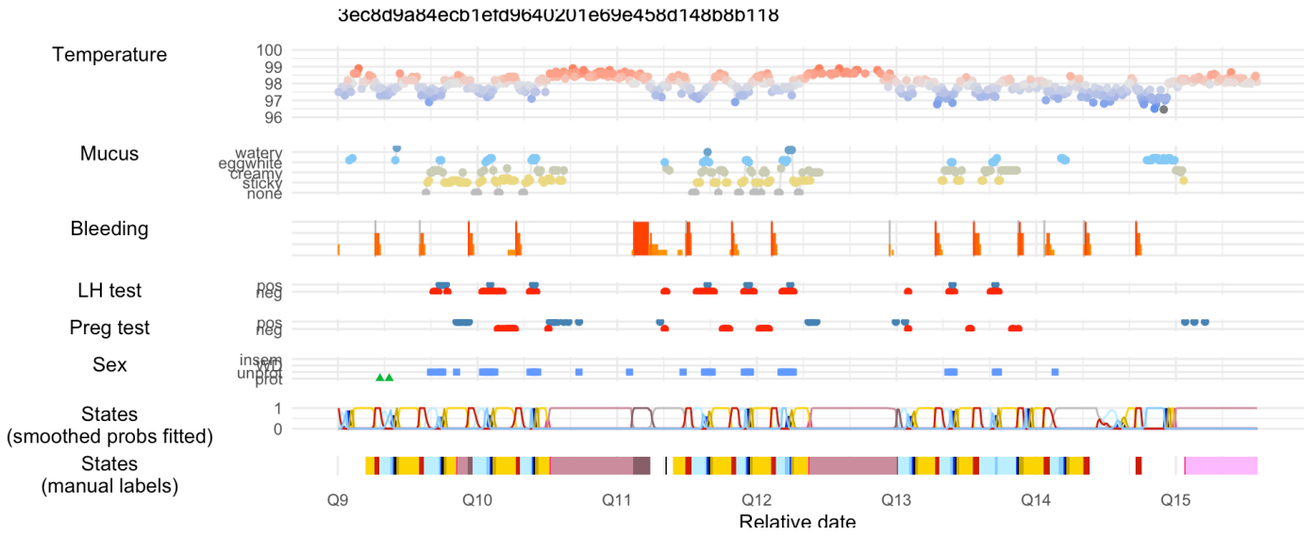
Conclusions



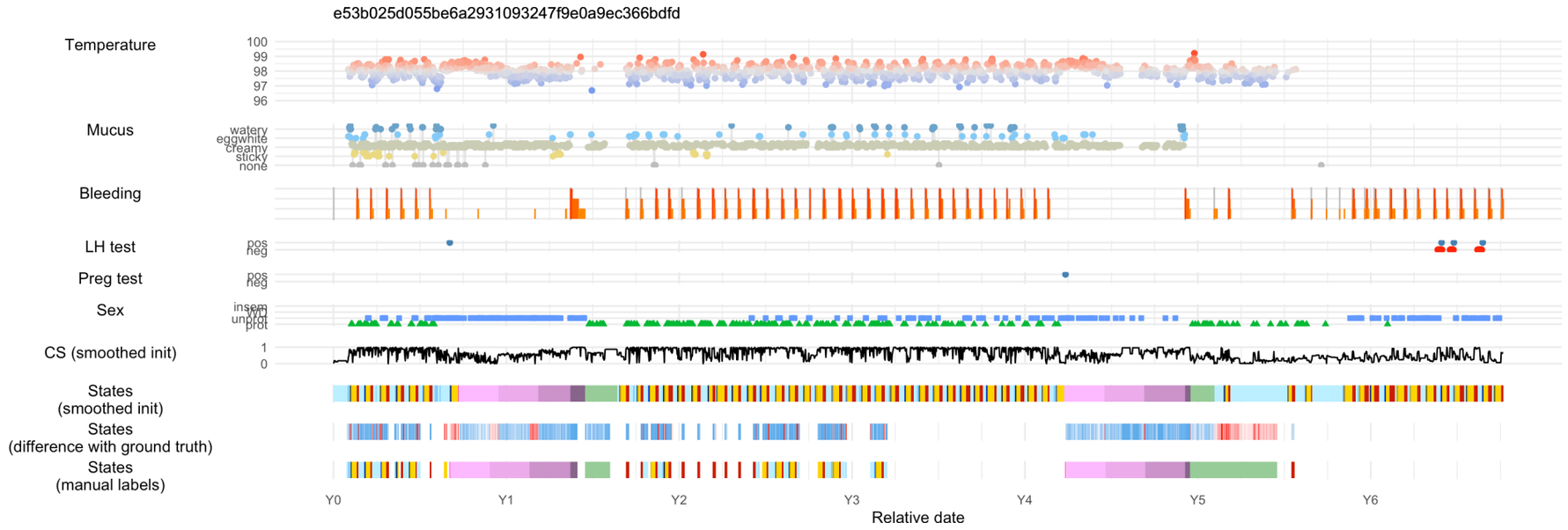
Upgraded `'hsmm'` R package that

- **can deal with high level of data missingness** (as found in any digital health project)
- **allows for any parametric or non-parametric distribution** for the **observed variables**
  - which allow for **discrete**, **categorical** or **continuous** variables
- can **quantify uncertainty** by providing a **confidence score** for the decoded sequence, based on **parametric bootstrap**, **independent of whether labelled data are available**
- **Available on my GitHub: @lasy**

A statistical model that can accurately decode menstrual/fertility app data at different time-scale:



Conclusions

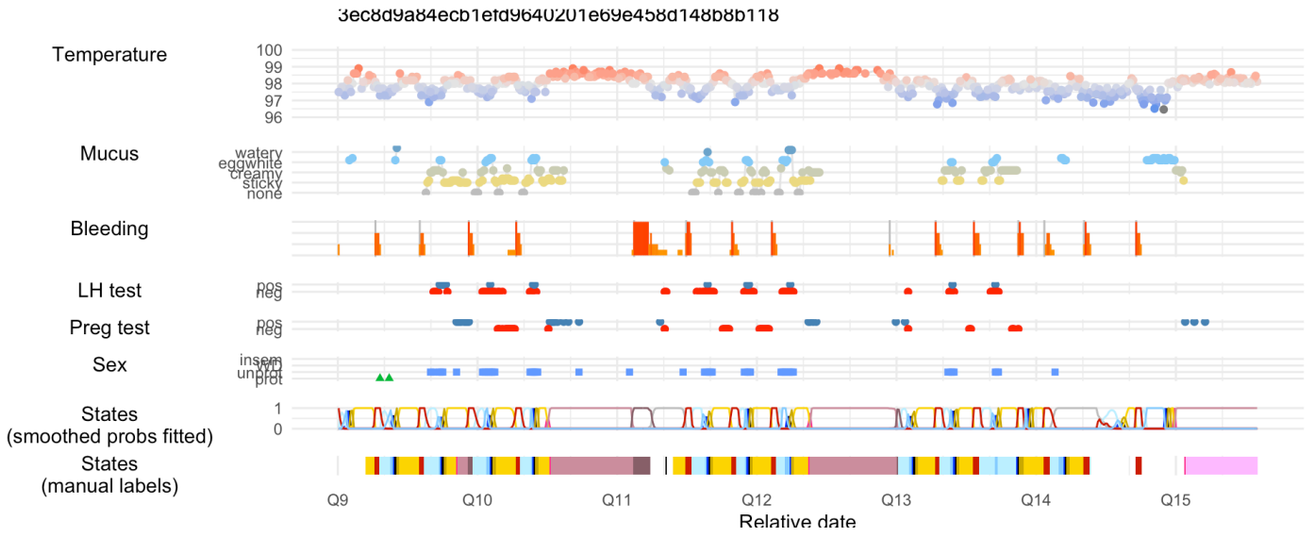


Upgraded `'hsmm'` R package that

- **can deal with high level of data missingness** (as found in any digital health project)
- **allows for any parametric or non-parametric distribution** for the **observed variables**
  - which allow for **discrete**, **categorical** or **continuous** variables
- can **quantify uncertainty** by providing a **confidence score** for the decoded sequence, based on **parametric bootstrap**, **independent of whether labelled data are available**
- **Available on my GitHub: @lasy**

A statistical model that can accurately decode menstrual/fertility app data at different time-scale:

> **About 200k users time-series labelled.**
Now we can interrogate these data.



Acknowledgments

Stanford University

Prof. Holmes

Stanford University

Prof. Hillard

Columbia University

Prof. Martinez

Acknowledgments

Stanford University
Prof. Holmes

Stanford University
Prof. Hillard

Columbia University
Prof. Martinez

Acknowledgments

Stanford University
Prof. Holmes

Stanford University
Prof. Hillard

Columbia University
Prof. Martinez

Funding

Acknowledgments

Stanford University
Prof. Holmes

Stanford University
Prof. Hillard

Columbia University
Prof. Martinez

KINDARA

Clue

sympto.org

Funding

FNSNF Postdoc Mobility

FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

Stanford University

Clinical Data Science fellowship