The Multilabel Naive Credal Classifier

Alessandro Antonucci and Giorgio Corani
{alessandro,giorgio}@idsia.ch

Istituto "Dalle Molle" di Studi sull'Intelligenza Artificiale - Lugano (Switzerland)
http://ipg.idsia.ch

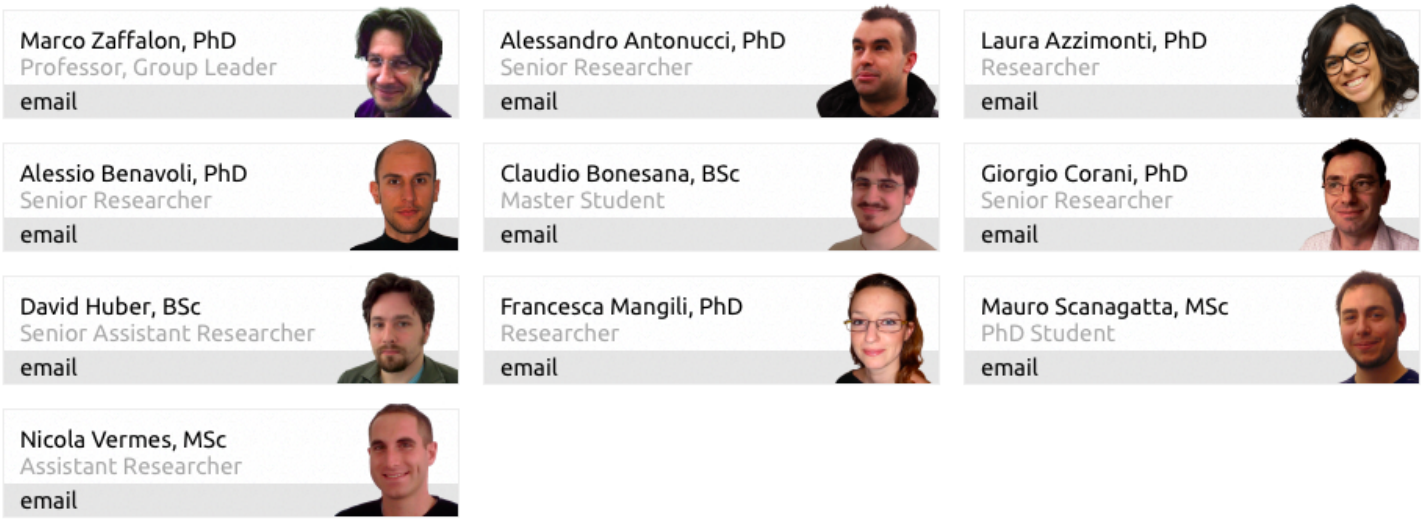
ISIPTA '15, Pescara, July 21st, 2015

IPG $\subset$ IDSIA $\subset$ USI $\cup$ SUPSI $\subset$ LUGANO

IPG {Imprecise Probability Group} @IDSIA

home about people publications

People



IPG $\subset$ **IDSIA** $\subset$ USI $\cup$ SUPSI $\subset$ LUGANO

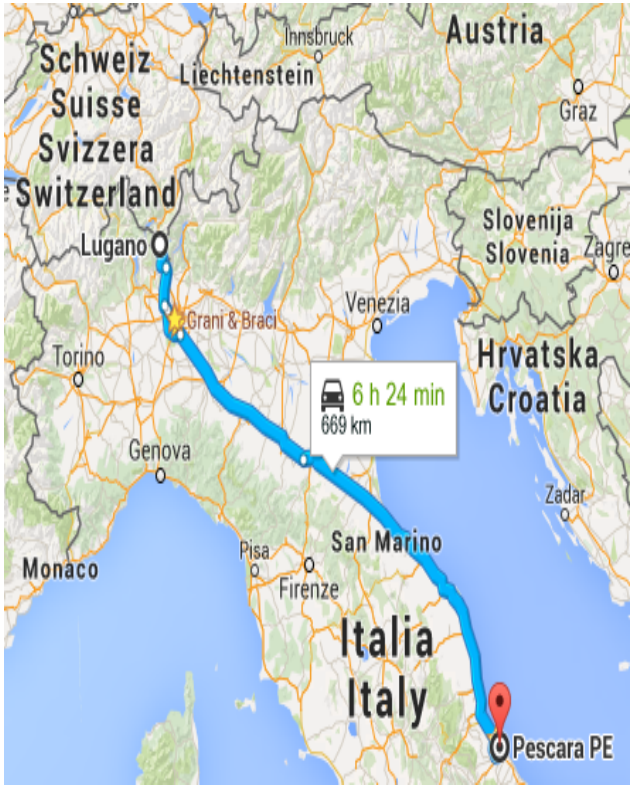
IPG $\subset$ **IDSIA** $\subset$ USI $\cup$ SUPSI $\subset$ LUGANO

IPG $\subset$ IDSIA $\subset$ **USI** $\cup$ **SUPSI** $\subset$ LUGANO

University of Applied Sciences
and Arts of Southern Switzerland
(`supsi.ch`)

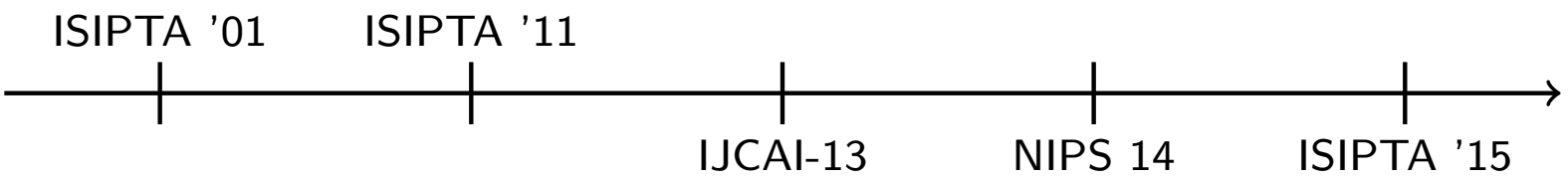
Università della Svizzera Italiana
(`usi.ch`)

IPG $\subset$ IDSIA $\subset$ USI $\cup$ SUPSI $\subset$ **LUGANO**



Chronology (Acknowledgements)

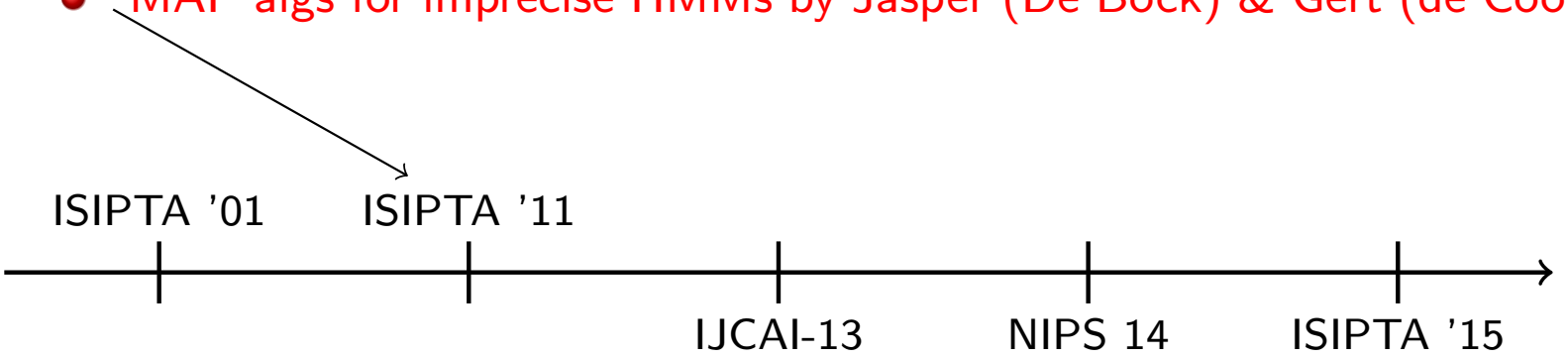
- Credal version of the naive Bayes classifier by Marco (Zaffalon)
- MAP algs for imprecise HMMs by Jasper (De Bock) & Gert (de Cooman)



- Bayes nets as multilabel classifiers by Denis (Mauá) & us
- MAP in generic credal nets by Jasper & Cassio (de Campos) & me
- A credal classifiers based on MAP tasks in credal nets by us

Chronology (Acknowledgements)

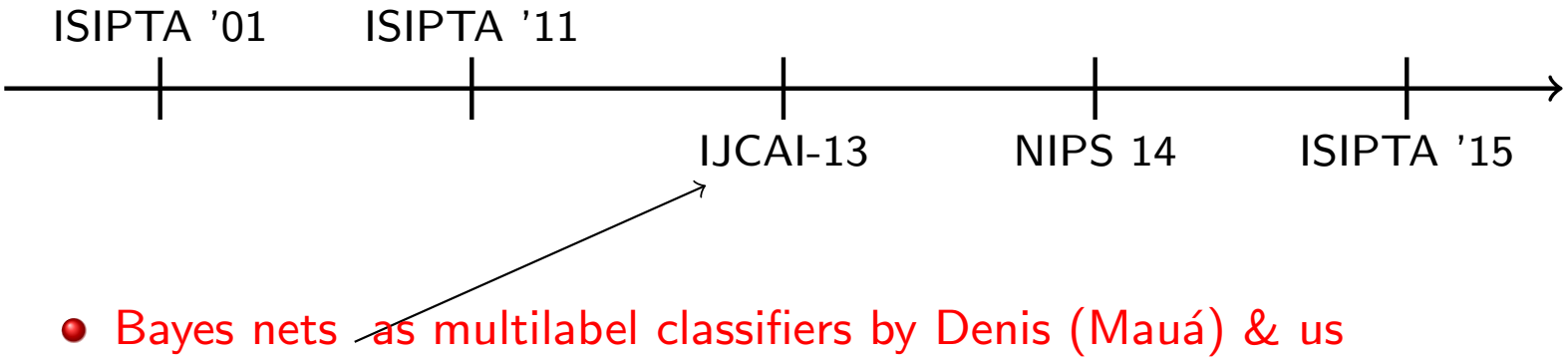
- Credal version of the naive Bayes classifier by Marco (Zaffalon)
- MAP algs for imprecise HMMs by Jasper (De Bock) & Gert (de Cooman)



- Bayes nets as multilabel classifiers by Denis (Mauá) & us
- MAP in generic credal nets by Jasper & Cassio (de Campos) & me
- A credal classifiers based on MAP tasks in credal nets by us

Chronology (Acknowledgements)

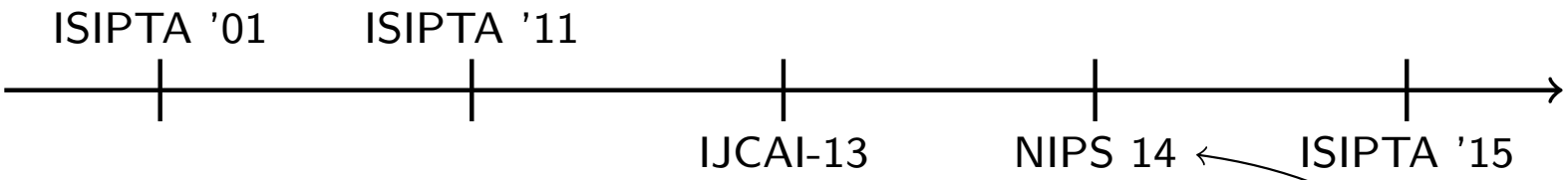
- Credal version of the naive Bayes classifier by Marco (Zaffalon)
- MAP algs for imprecise HMMs by Jasper (De Bock) & Gert (de Cooman)



- Bayes nets as multilabel classifiers by Denis (Mauá) & us
- MAP in generic credal nets by Jasper & Cassio (de Campos) & me
- A credal classifiers based on MAP tasks in credal nets by us

Chronology (Acknowledgements)

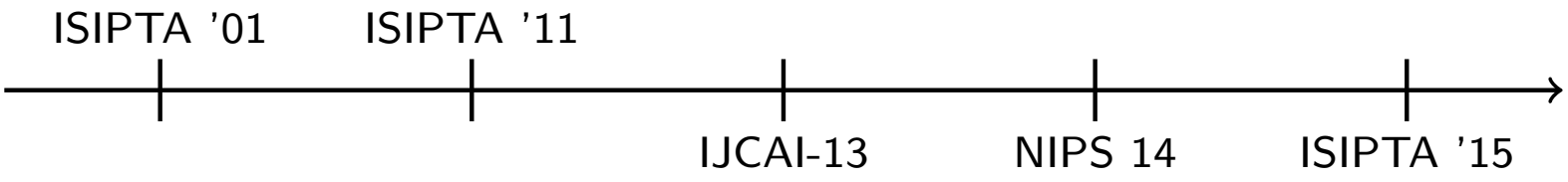
- Credal version of the naive Bayes classifier by Marco (Zaffalon)
- MAP algs for imprecise HMMs by Jasper (De Bock) & Gert (de Cooman)



- Bayes nets as multilabel classifiers by Denis (Mauá) & us
- MAP in generic credal nets by Jasper & Cassio (de Campos) & me
- A credal classifiers based on MAP tasks in credal nets by us

Chronology (Acknowledgements)

- Credal version of the naive Bayes classifier by Marco (Zaffalon)
- MAP algs for imprecise HMMs by Jasper (De Bock) & Gert (de Cooman)



- Bayes nets as multilabel classifiers by Denis (Mauá) & us
- MAP in generic credal nets by Jasper & Cassio (de Campos) & me
- A credal classifiers based on MAP tasks in credal nets by us

Single- vs. multi-label classification

- A (fictious) classifier to detect eyes color
- Possible classes $\mathcal{C} := \{\text{brown,green,blue}\}$
- *Heterochromia iridum*: two (or more) colors
- Possible values in $2^{\mathcal{C}}$, a multilabel task!
- Trivial approaches
  - Standard classification over the power set
    Exponential in the number of labels!
  - Each label as a separate Boolean variable
    a (standard) classifier for each label
    Ignored relations among classes !
- Graphical models (GMs) to depict relations among class labels (and features)
- Classification as (standard) inference in GMs

SINGLE-LABEL

C = green

MULTI-LABEL

C = {blue,brown}

Credal classifiers are not (yet) multilabel classifiers

- Class variable C and (discrete) features $\boldsymbol{F}$, a test instance $\tilde{\boldsymbol{f}}$
- Standard (single-label) classifier are maps: $\boldsymbol{\mathcal{F}} \to \mathcal{C}$

  learn $P(C, \boldsymbol{F})$ from data and return $c^* := \arg\max_{c \in \mathcal{C}} P(c, \tilde{\boldsymbol{f}})$
- Multi-label classifiers: $\boldsymbol{\mathcal{F}} \to 2^{\mathcal{C}}$

  $\boldsymbol{C} = (C_1, \ldots, C_n)$ as an array of Boolean vars, one for each label

  learn $P(\boldsymbol{C}, \boldsymbol{F})$ and solve the MAP task $\boldsymbol{c}^* := \arg\max_{\boldsymbol{c} \in \{0,1\}^n} P(\boldsymbol{c}, \tilde{\boldsymbol{f}})$
- Credal (single-label) classifiers: $\boldsymbol{\mathcal{F}} \to 2^{\mathcal{C}}$

  learn credal set $K(C, \boldsymbol{F})$ and return all $c'' \in \mathcal{C}$ s.t.

  $\nexists c' : P(c', \tilde{\boldsymbol{f}}) > P(c'', \tilde{\boldsymbol{f}}) \quad \forall P(C, \boldsymbol{F}) \in K(C, \boldsymbol{F})$
- Multilabel credal classifier (MCC): $\boldsymbol{F} \to 2^{2^{\mathcal{C}}}$

  learn credal set $K(\boldsymbol{C}, \boldsymbol{F})$ and return all sequences $\boldsymbol{c}''$ s.t.

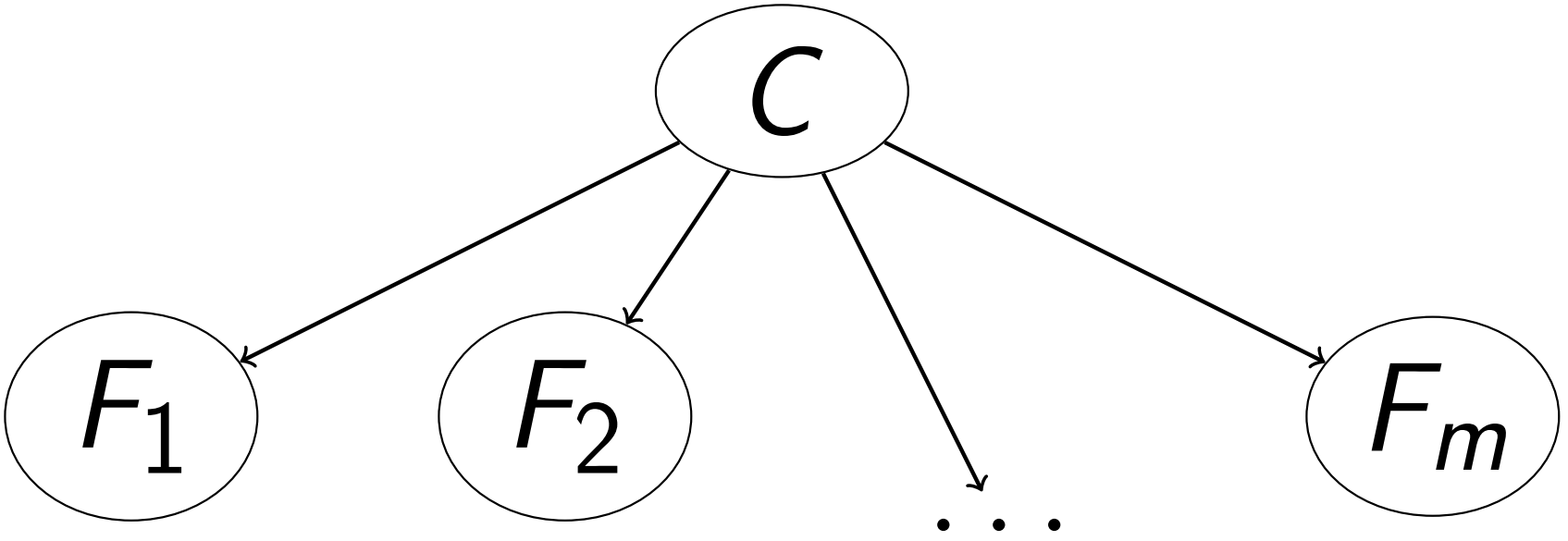
  $\nexists \boldsymbol{c}' : P(\boldsymbol{c}', \tilde{\boldsymbol{f}}) > P(\boldsymbol{c}'', \tilde{\boldsymbol{f}}) \quad \forall P(\boldsymbol{C}, \boldsymbol{F}) \in K(\boldsymbol{C}, \boldsymbol{F})$

Compact Representation of the Output

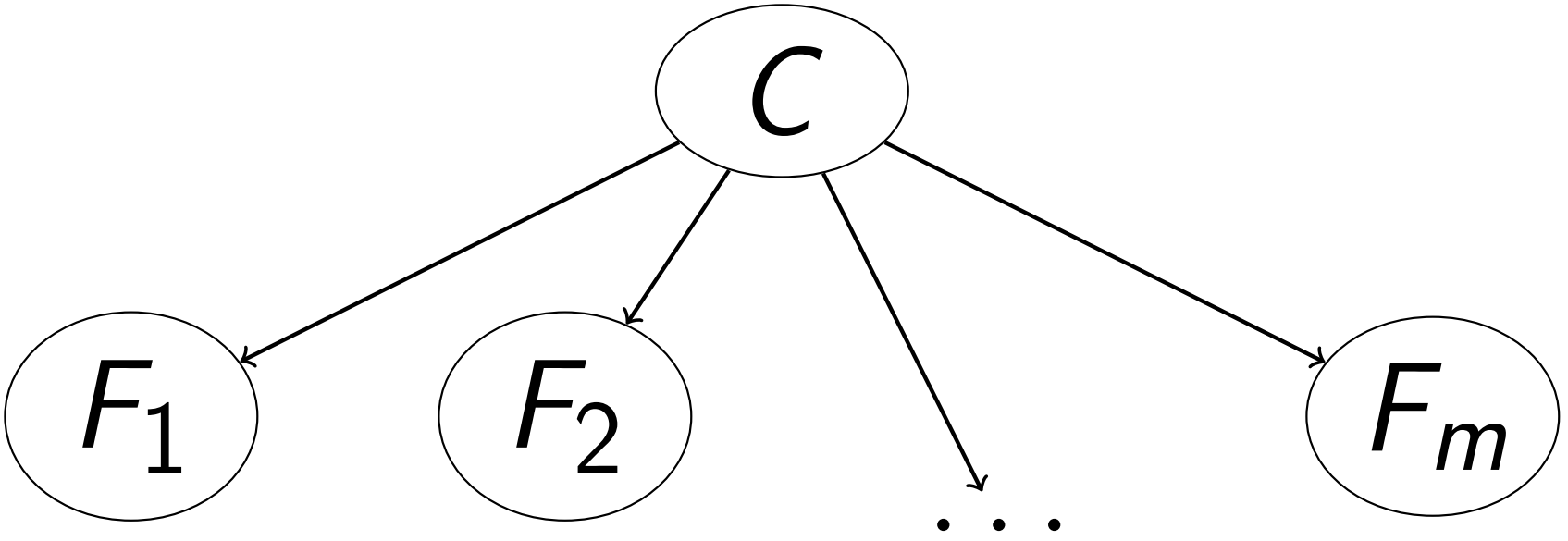
- Output of a MCC might be exponentially large
- Jasper & Gert's idea to fix this with imprecise HMMs (Viterbi): decide whether or not there is at least an optimal sequence sucht that a variable is in a particular state (for each variable and state)
- With MCCs, for each class label, we can decide whether:
  - the class is active for all the optimal sequences
  - the class is inactive fro all the optimal sequences
  - there are optimal sequences with the label active, and others with the label inactive
- Optimization task

$$\min_{\boldsymbol{c}'':c_l''=0/1} \max_{\boldsymbol{c}'} \inf_{P(\boldsymbol{C},\boldsymbol{F})\in K(\boldsymbol{C},\boldsymbol{F})} \frac{P(\boldsymbol{c}',\boldsymbol{f})}{P(\boldsymbol{c}'',\boldsymbol{f})} \leq 1$$

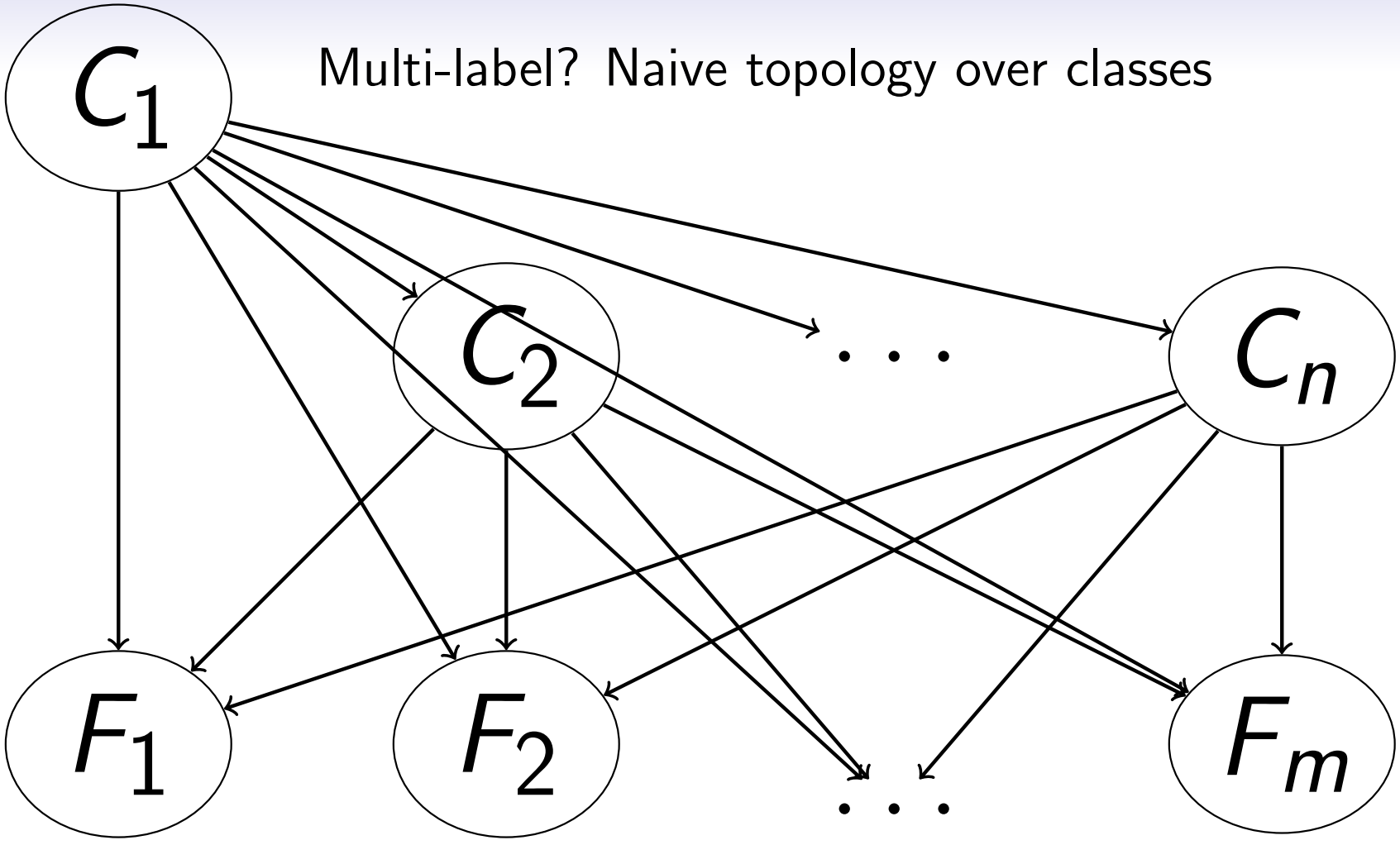
- $O(2^{\text{treewidth}})$ for separately specified credal nets (e.g., local IDM)
- More complex with non-separate specifications



NBC

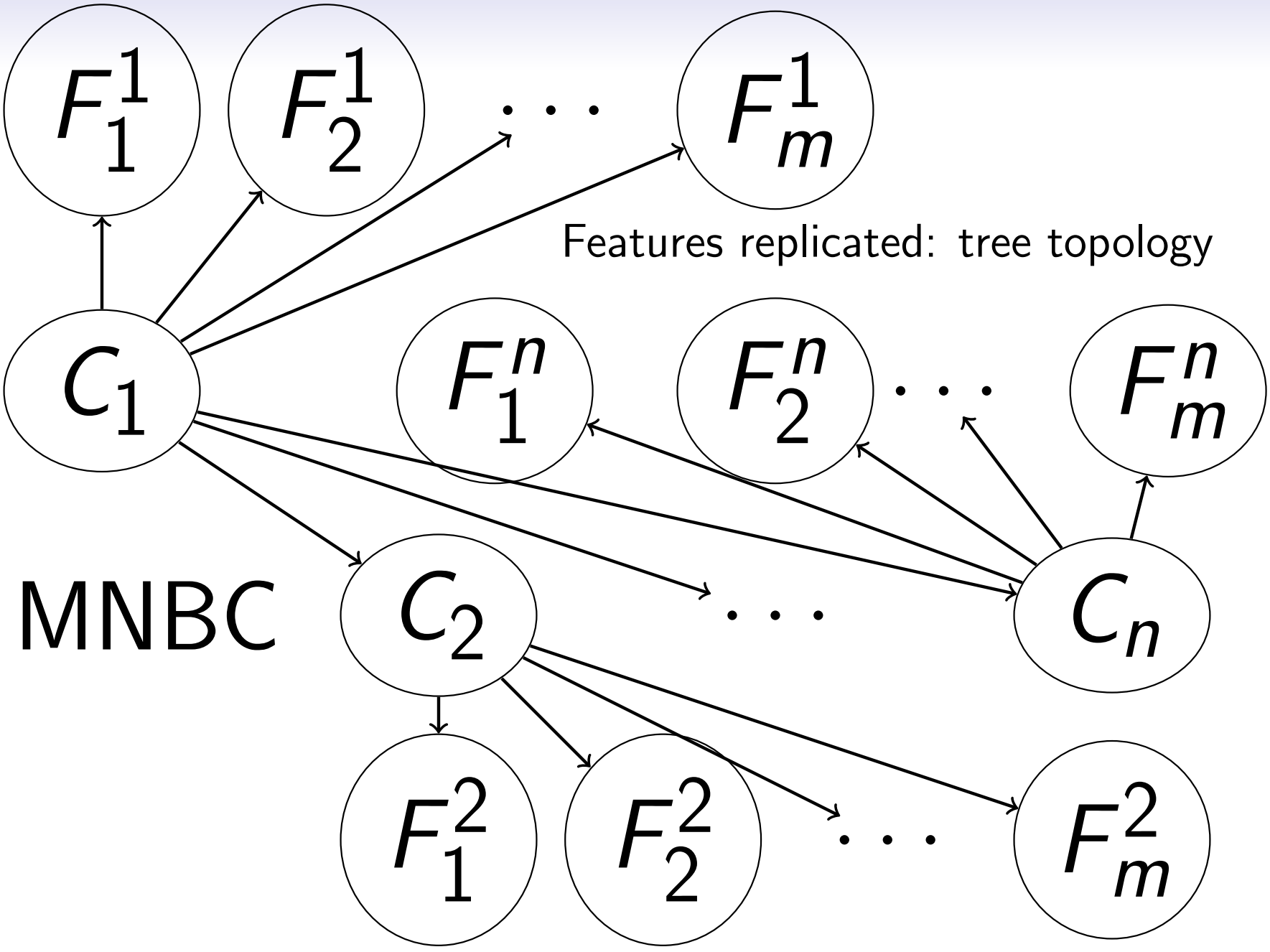


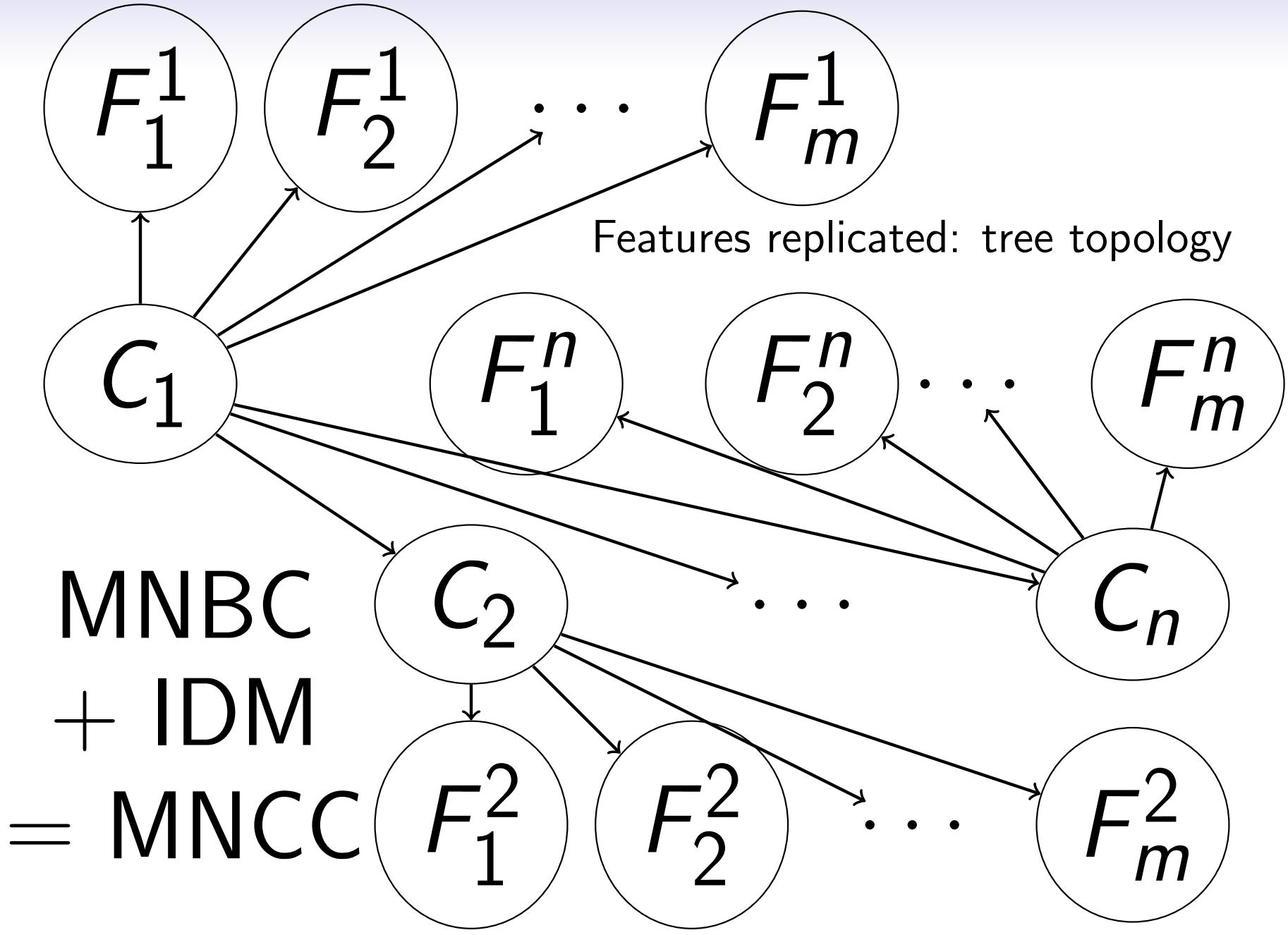
NCC=NBC+IDM



Structural learning to bound # of parents of the features

and to select the super-class C_1


F_1^1
F_2^1
$\cdots$
F_m^1
Features replicated: tree topology
C_1
F_1^n
F_2^n
$\cdots$
F_m^n
MNBC
C_2
$\cdots$
C_n
F_1^2
F_2^2
$\cdots$
F_m^2


F_1^1
F_2^1
$\cdots$
F_m^1
Features replicated: tree topology
C_1
F_1^n
F_2^n
$\cdots$
F_m^n
MNBC
C_2
$\cdots$
C_n
+ IDM
= MNCC
F_1^2
F_2^2
$\cdots$
F_m^2

During the poster session I can

- Explain some detail about the learning of the structure
- Explain the *feature replication* trick (tis makes inference simpler)
- Explain the non-separate IDM-based quantification of the model
- Explain the detail of the (convex) optimization
- ...

MNCC: the algorithm

Input: test instance $\boldsymbol{f}$ (+ dataset $\mathcal{D}$) / Output initialized:

	C_1	C_2	$\ldots$	C_n
active	0	0	$\ldots$	0
inactive	0	0	$\ldots$	0

for $l = 1, \ldots, n$ **do**
  for $c_l = 0, 1$ **do**
    if $\min_{\boldsymbol{c}'' : c_l'' = c_l} \max_{\boldsymbol{c}'} \inf_{\boldsymbol{t}} \frac{P_{\boldsymbol{t}}(\boldsymbol{c}', \boldsymbol{f})}{P_{\boldsymbol{t}}(\boldsymbol{c}'', \boldsymbol{f})} \leq 1$ **then**
      Output(l, c_l)=1
    end if
  end for
end for

linear representation of a (exponential) number of maximal seqs

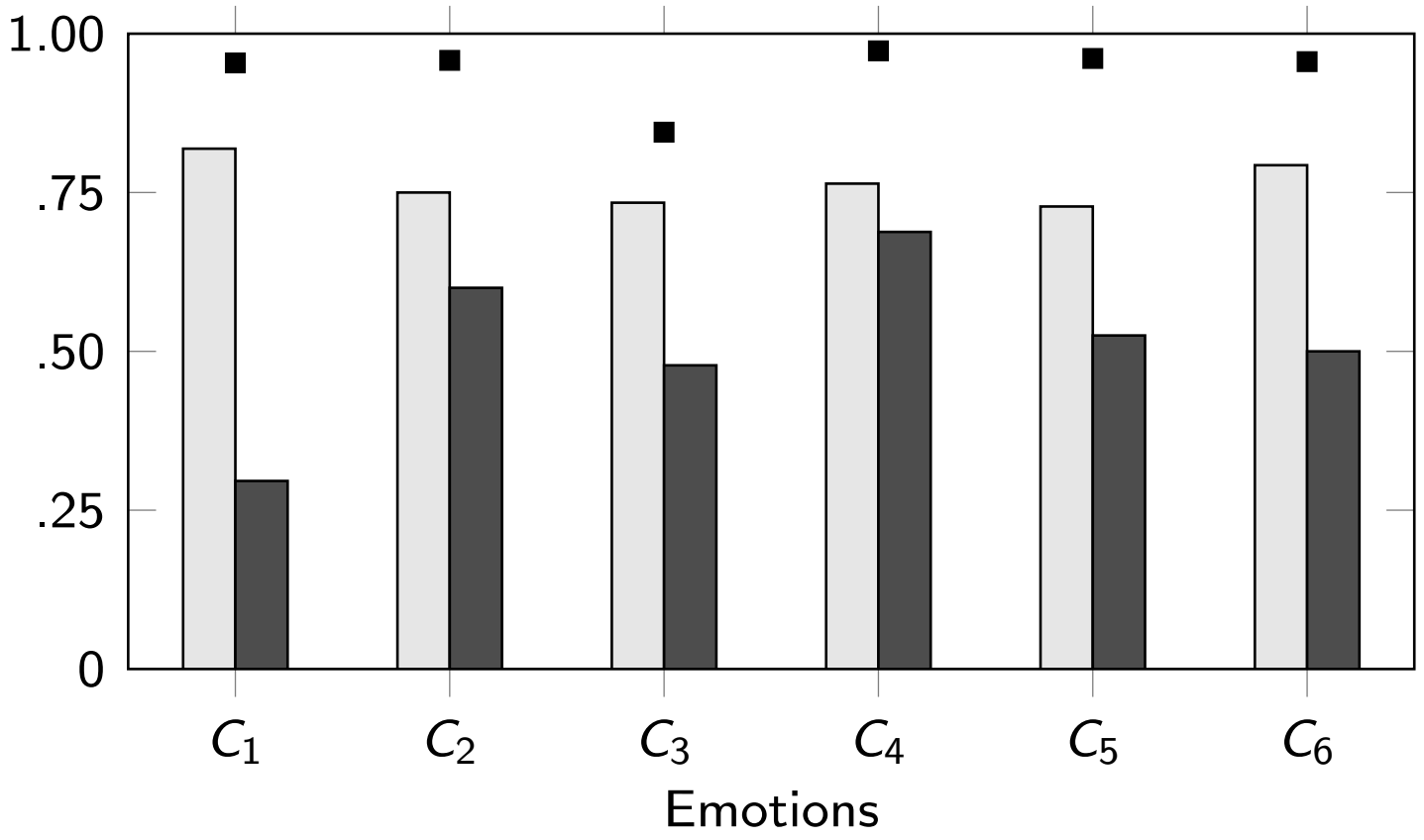
1	1	1	0
0	1	0	1

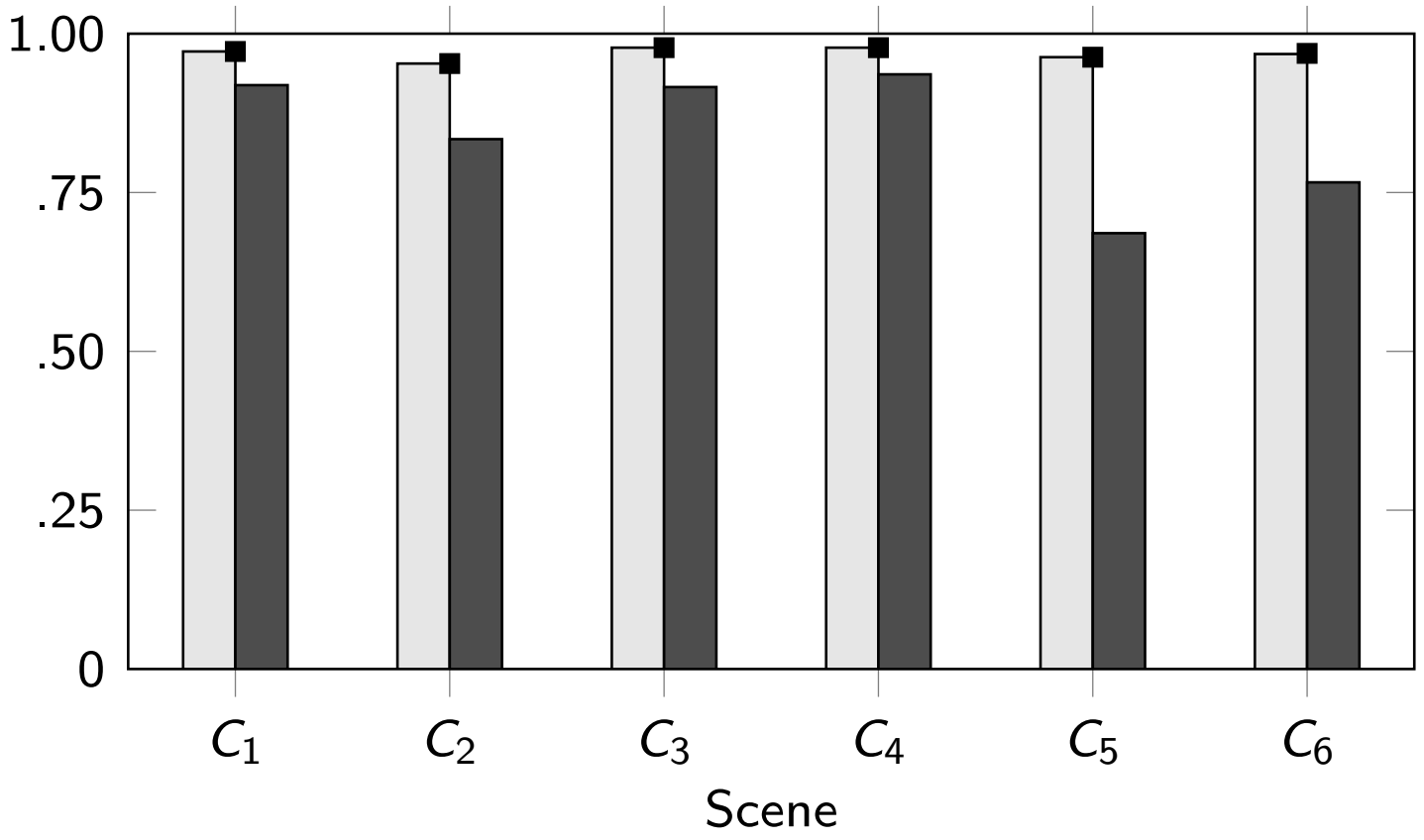
Testing MNCC

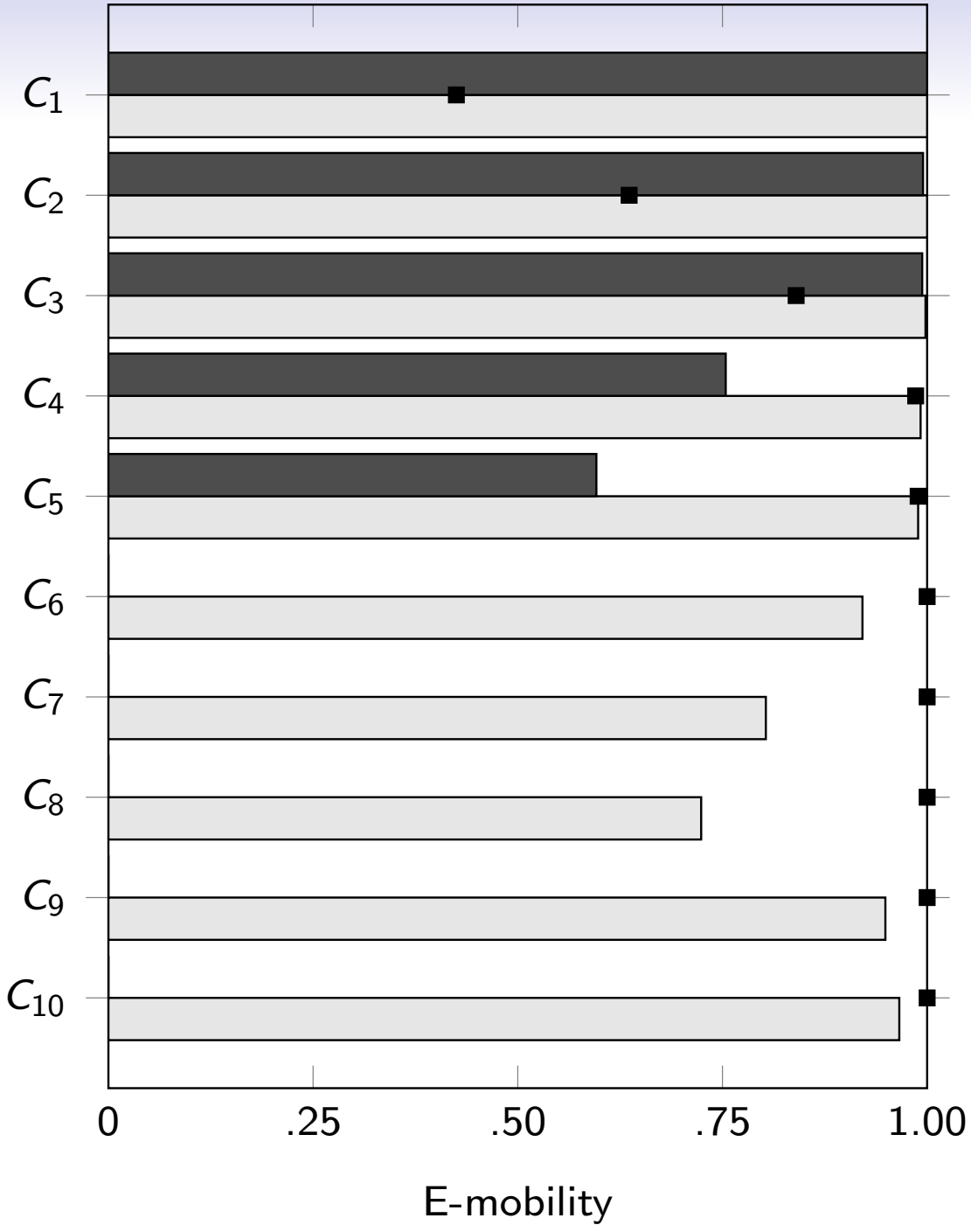
- Preliminary tests on real-world datasets

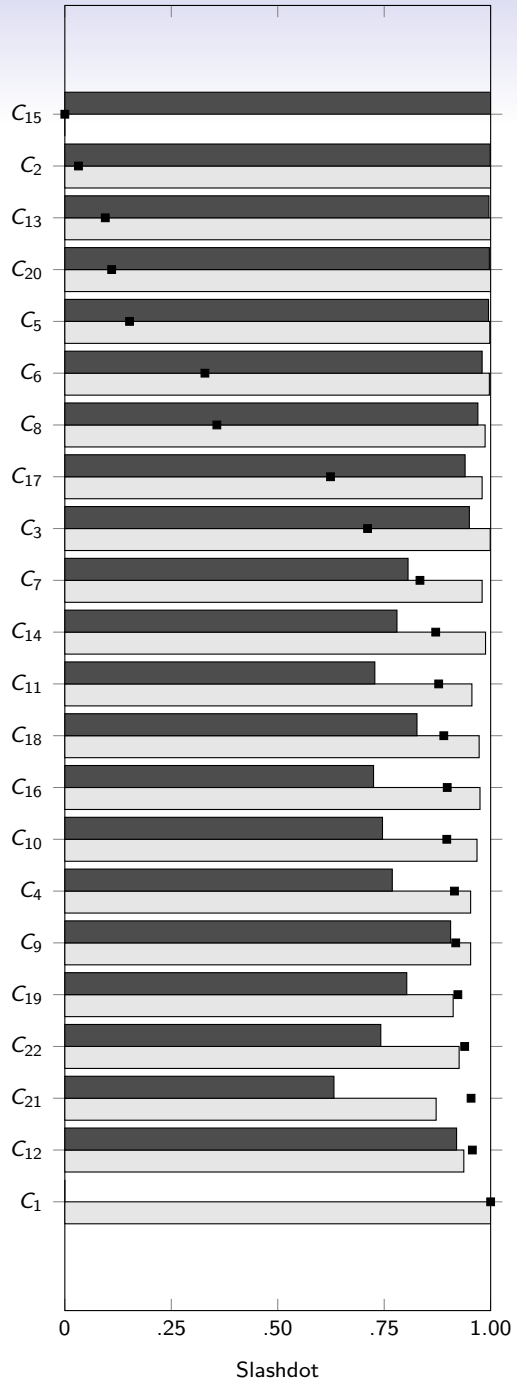
Data set	Classes	Features	Instances
Emotions	6	44/72	593
Scene	6	224/294	2407
E-mobility	10	14/18	4226
Slashdot	22	496/1079	3782

- Perfomance described by:
  - % of instance s.t. all maximal seqs all in the same state
  - Accuracy of the precise model when MNCC is determinate
  - Accuracy of the precise model when MNCC is indeterminate


1.00
.75
.50
.25
0
C_1
C_2
C_3
C_4
C_5
C_6
Emotions


1.00
.75
.50
.25
0
C_1
C_2
C_3
C_4
C_5
C_6
Scene


C_1
C_2
C_3
C_4
C_5
C_6
C_7
C_8
C_9
C_{10}
0
.25
.50
.75
1.00
E-mobility


C_{15}
C_2
C_{13}
C_{20}
C_5
C_6
C_8
C_{17}
C_3
C_7
C_{14}
C_{11}
C_{18}
C_{16}
C_{10}
C_4
C_9
C_{19}
C_{22}
C_{21}
C_{12}
C_1
0
.25
.50
.75
1.00
Slashdot

Conclusions, Outlooks and Acks

- Among the first tools for robust multilabel classification

Still lots of things to do:

- Extension to multidimensional/hierarchical case
- Extension to continuous variables (features)
- Extension to continuous class (multi-target interval-valued regression)
- More complex topologies (ETAN, de Campos, 2014)
- Variational approach to features replication
- Not only 0/1 losses (imprecise losses?)