Metric-based State Space Reduction for MCs

Giovanni Bacci, **Giorgio Bacci**, Kim G. Larsen, Radu Mardare
Aalborg University

QuantLA Seminars, TU Dresden

24th Jan 2017

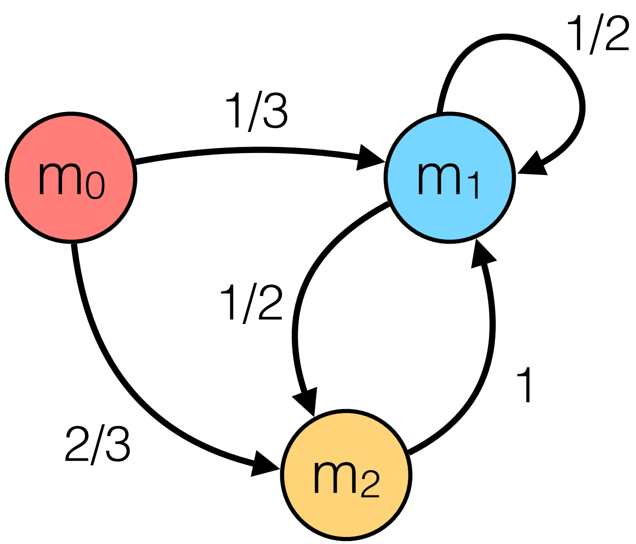
Talk Outline

★ **Labelled Markov Chains**
- probabilistic bisimilarity
- couplings

★ **Behavioral distances on Markov Chains**
- probabilistic bisimilarity distance
- relation with probabilistic model checking

★ **Metric-based state space reduction**
- Closest Bounded Approximant (CBA)
- Minimum Significant Approximant Bound (MSAB)
- Expectation Maximization-like algorithm

Probabilistic Systems



labelled
Markov Chain

$\tau\colon M \to \mathrm{Dist}(M)$

the transitions of a state m are presented by a probability distribution τ(m) on M

$$\tau(m_0)(u) = \begin{cases} 1/3 & \text{if } u = m_1 \\ 2/3 & \text{if } u = m_2 \\ 0 & \text{otherwise} \end{cases}$$

Probabilistic Bisimulation

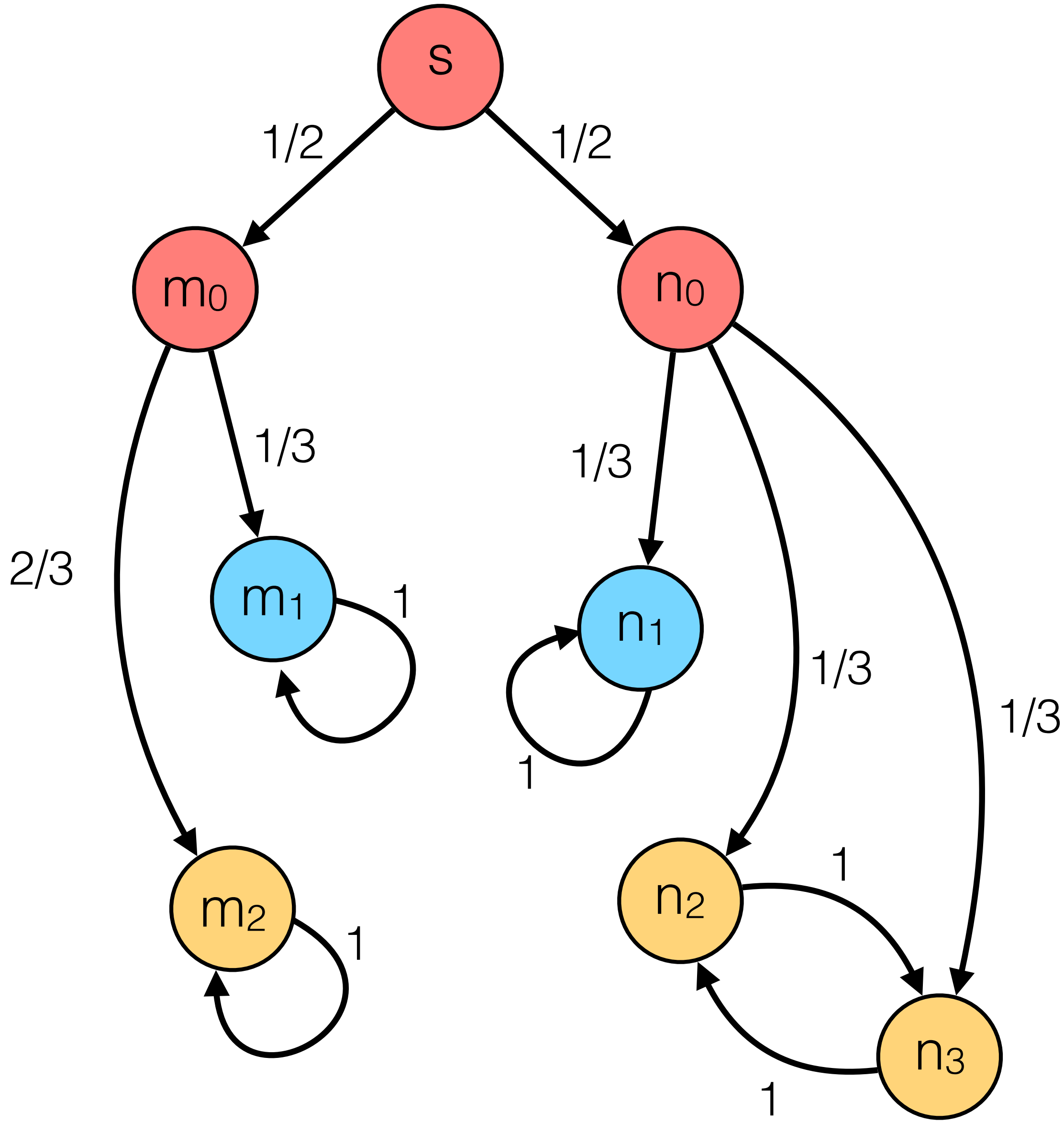
Definition (Larsen & Skou 89)

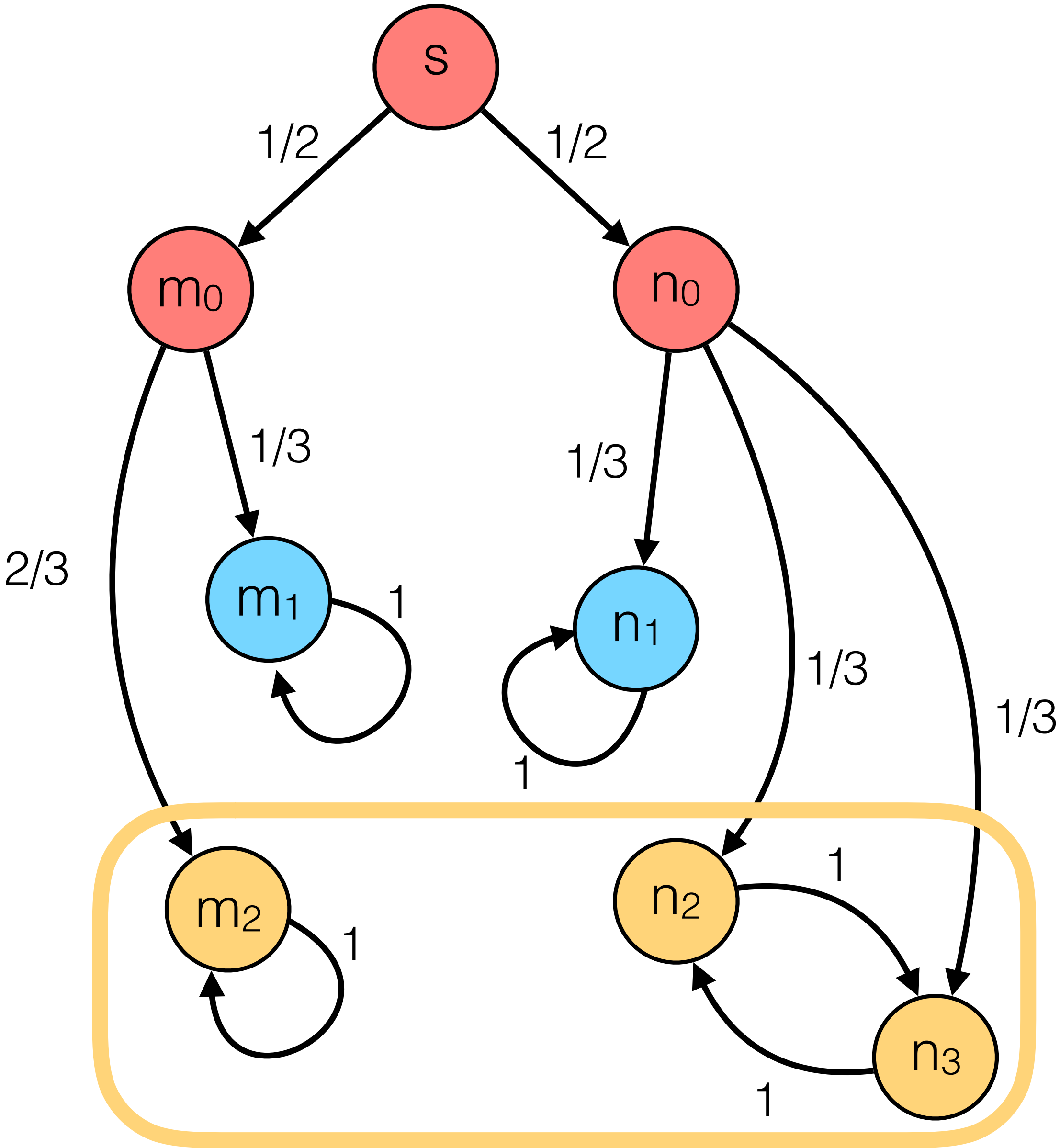
An equivalence relation $R \subseteq M \times M$ is a *probabilistic bisimulation* if for all $(m,n) \in R$

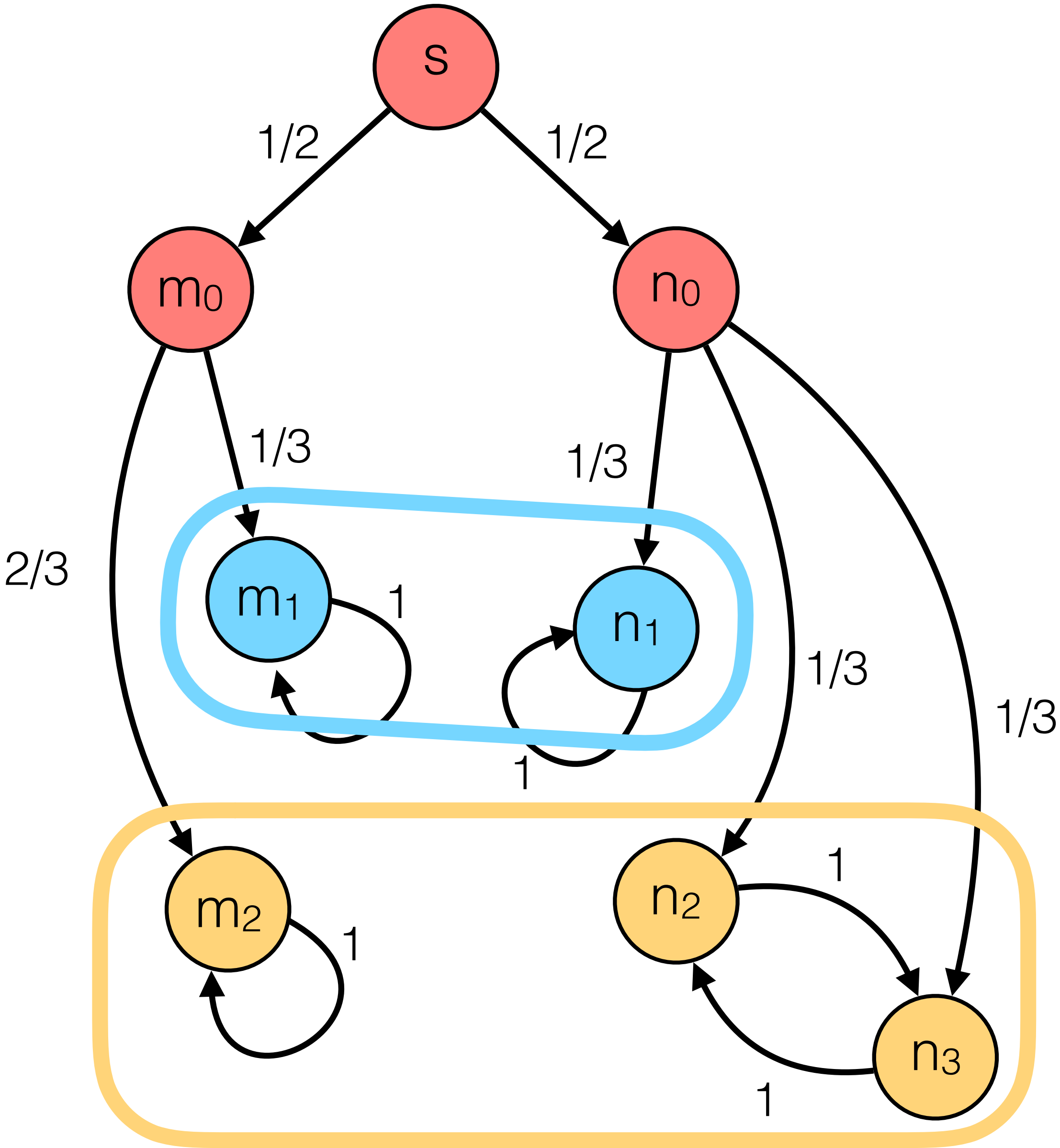
- $\ell(m) = \ell(n)$ and
- for all $C \in M/_R$, $\sum_{u \in C} \tau(m)(u) = \sum_{u \in C} \tau(n)(u)$.

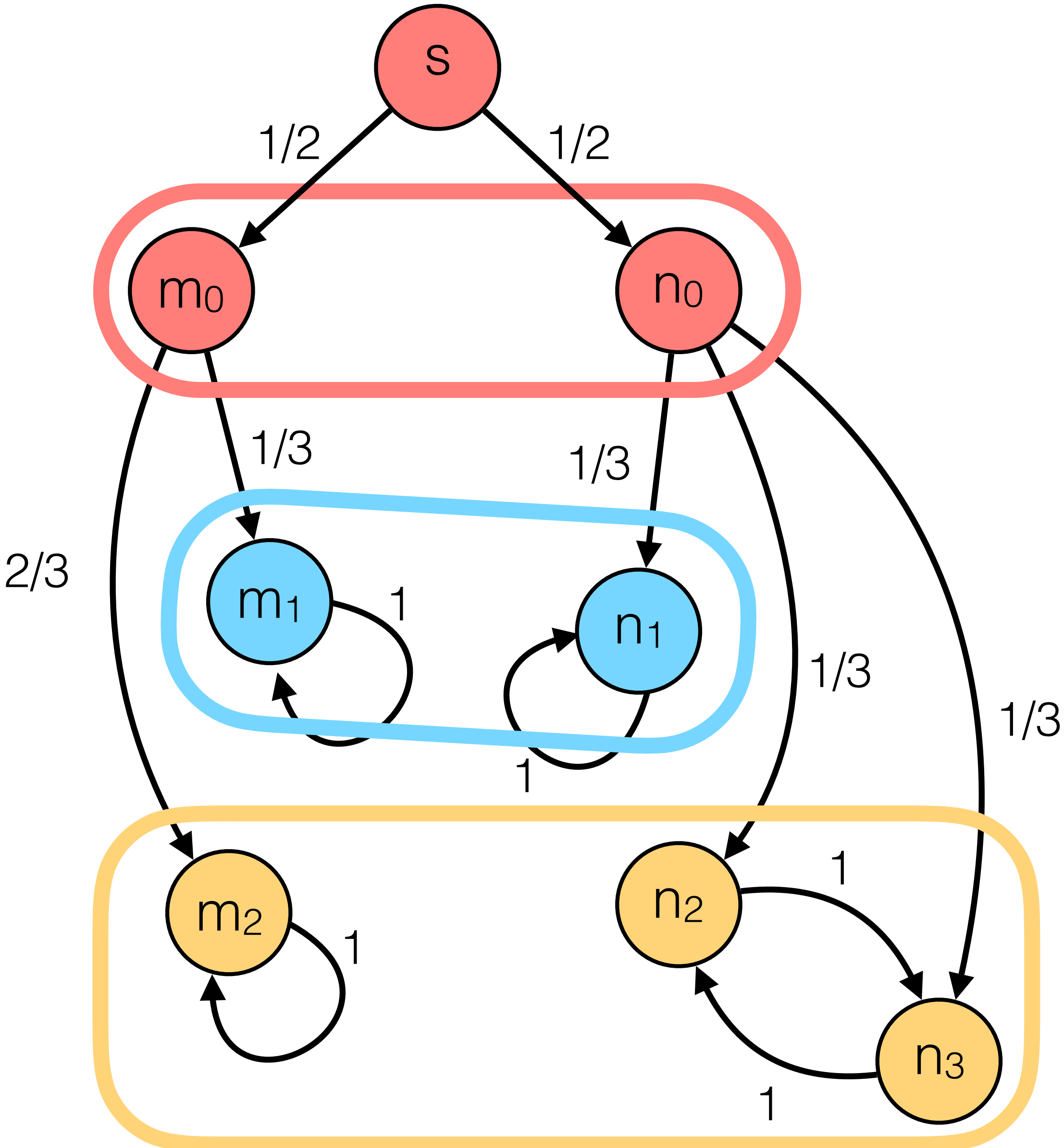
Definition

Probabilistic bisimilarity is the largest probabilistic bisimulation


s
1/2
1/2
m0
n0
1/3
2/3
1/3
m1
1
n1
1/3
1/3
1
m2
1
n2
1
n3
1


s
1/2
1/2
m0
n0
1/3
1/3
2/3
m1
1
n1
1/3
1/3
1
1
m2
1
n2
n3
1


s
1/2
1/2
m0
n0
1/3
1/3
2/3
m1
1
n1
1
1/3
1/3
m2
1
n2
1
n3
1


s
1/2
1/2
m0
n0
1/3
1/3
2/3
m1
1
n1
1
1/3
1/3
m2
1
n2
1
n3
1

Complexity of Bisimulation

Proposition (Jonsson, Larsen 91)

An equivalence relation $R \subseteq M \times M$ is a probabilistic bisimulation if for all $(m,n) \in R$

- $\ell(m) = \ell(n)$ and
- exists $\omega \in \Omega(\tau(m),\tau(n))$ such that $\text{supp}(\omega) \subseteq R$.

set of couplings

Theorem (Baier CAV96)

Probabilistic bisimilarity can be tested in *polynomial time* —specifically $O(h^2e)$

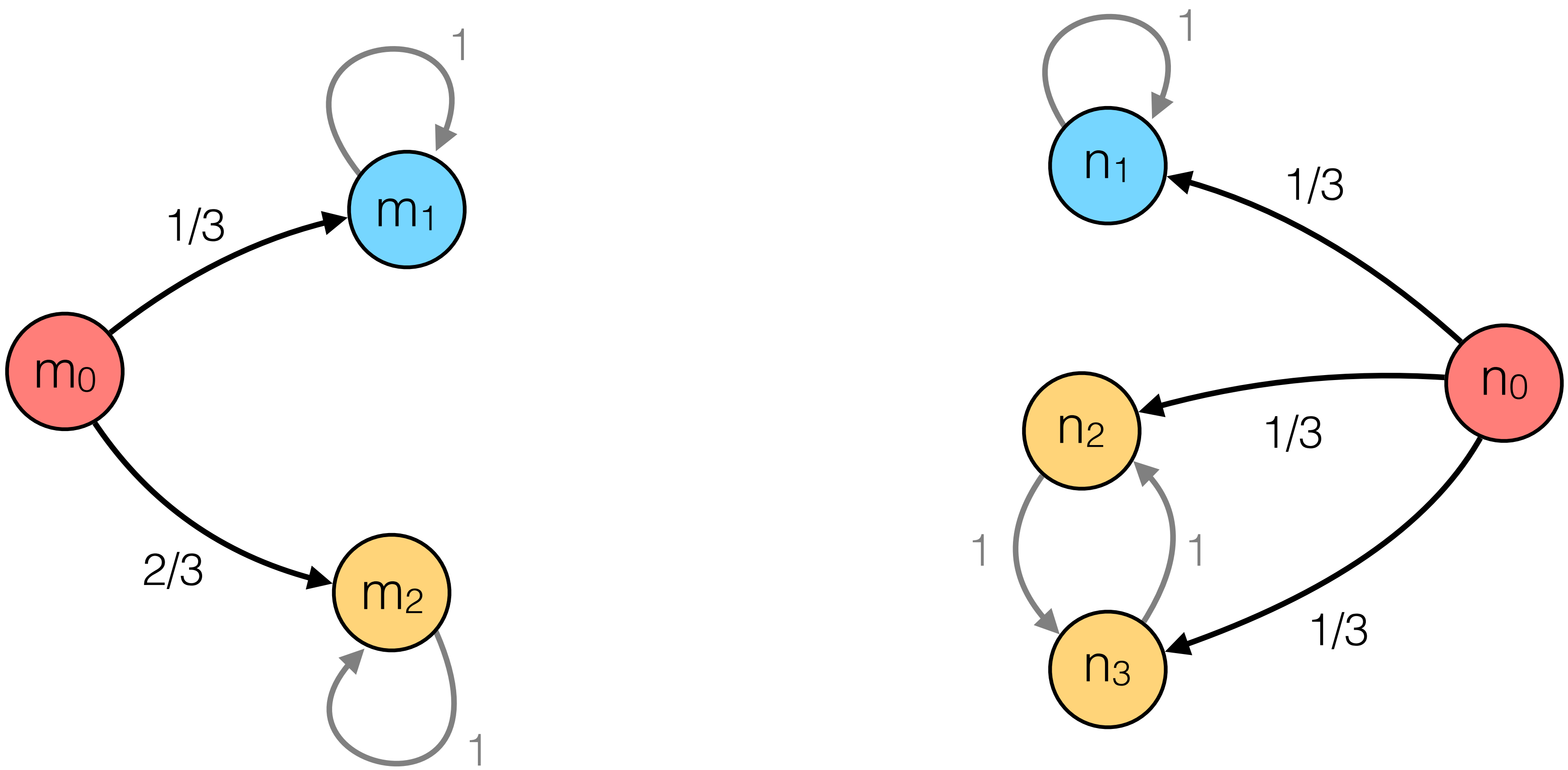
Coupling

Definition (W. Doeblin 36)

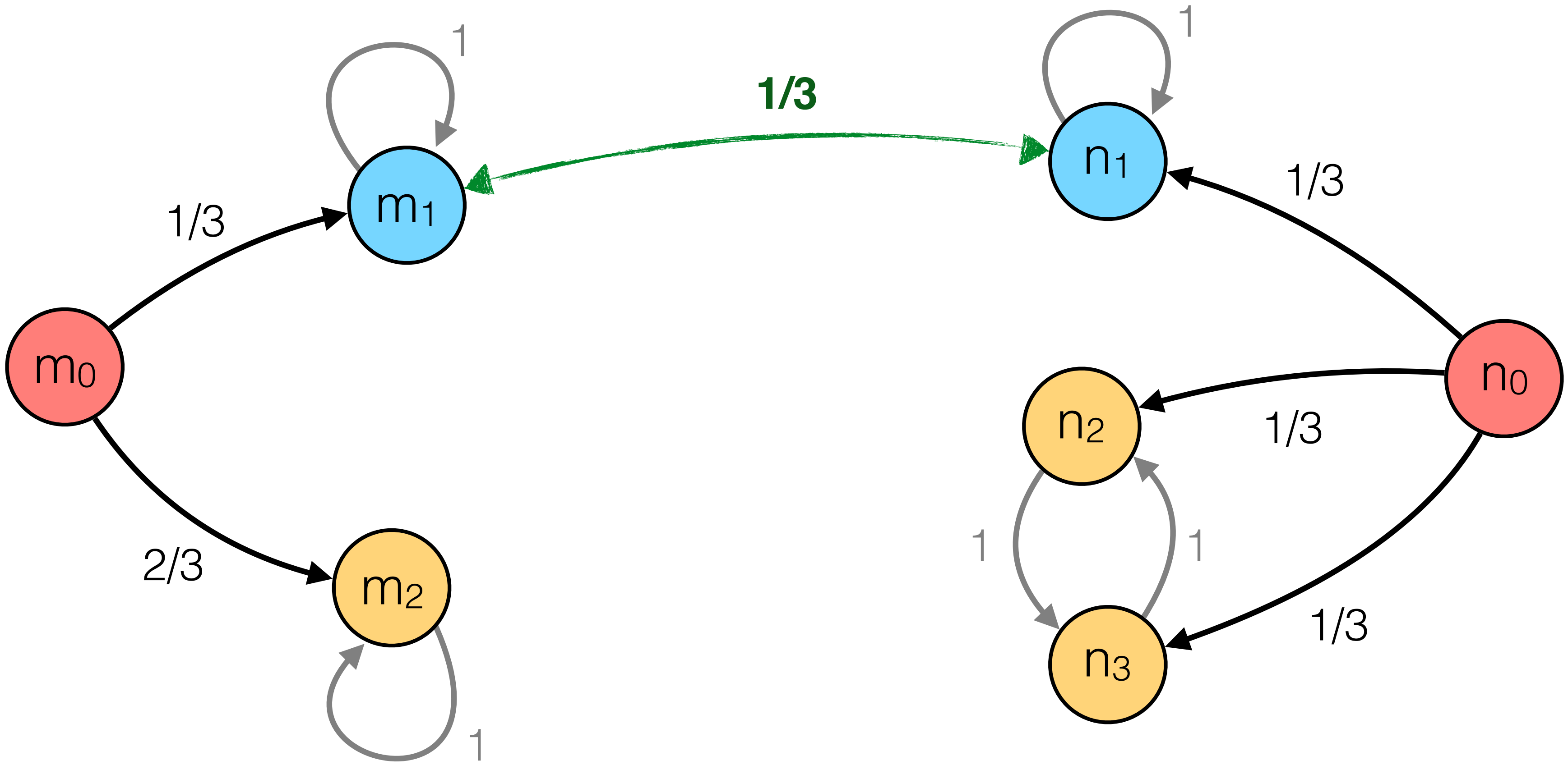
A *coupling* of a pair (μ,ν) of probability distributions on M is a distribution ω on M×M such that

- $\sum_{n \in M} \omega(m,n) = \mu(m)$ *(left marginal)*
- $\sum_{m \in M} \omega(m,n) = \nu(n)$ *(right marginal)*.

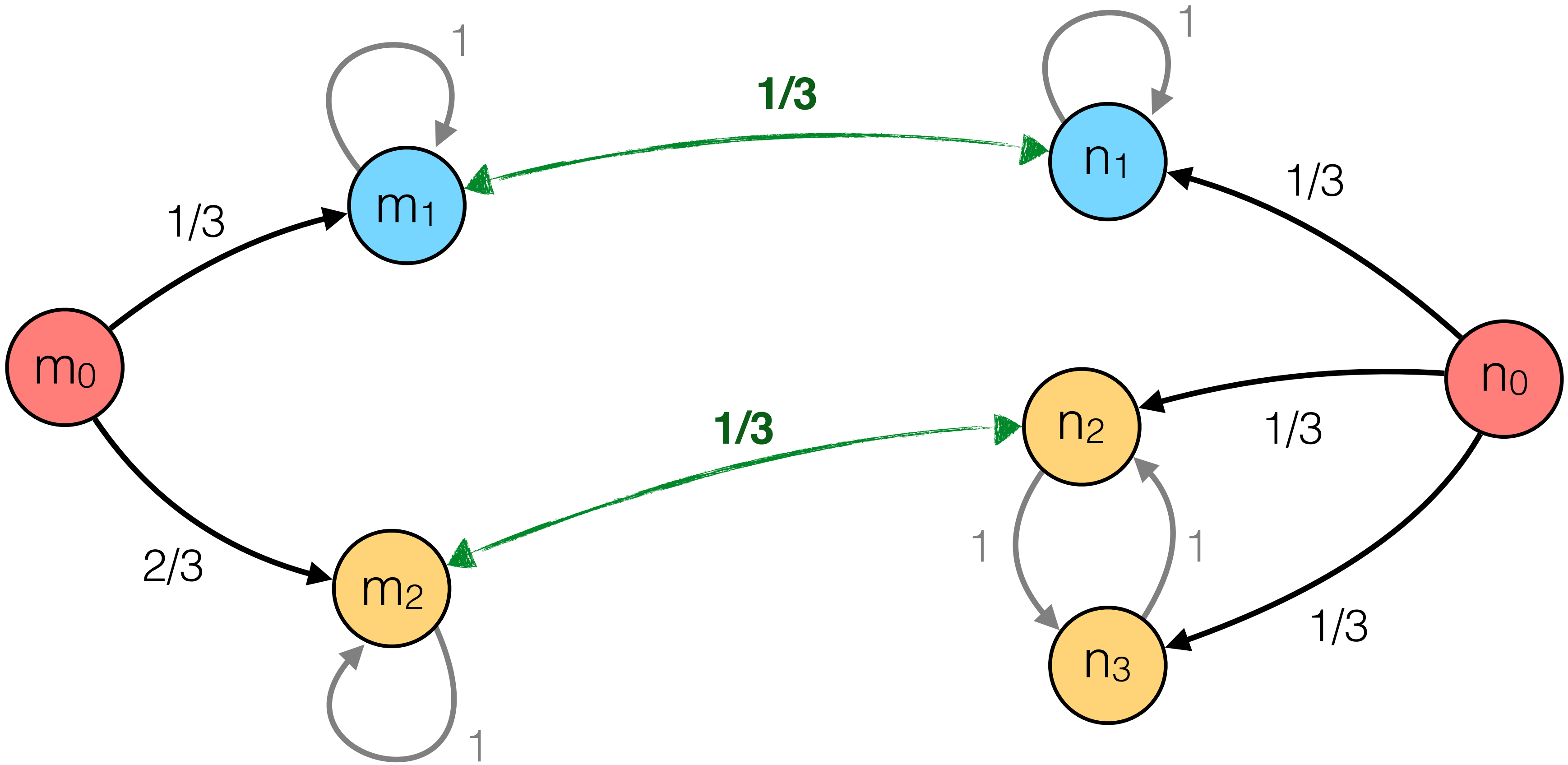
One can think of a coupling as a measure-theoretic relation between probability distribution



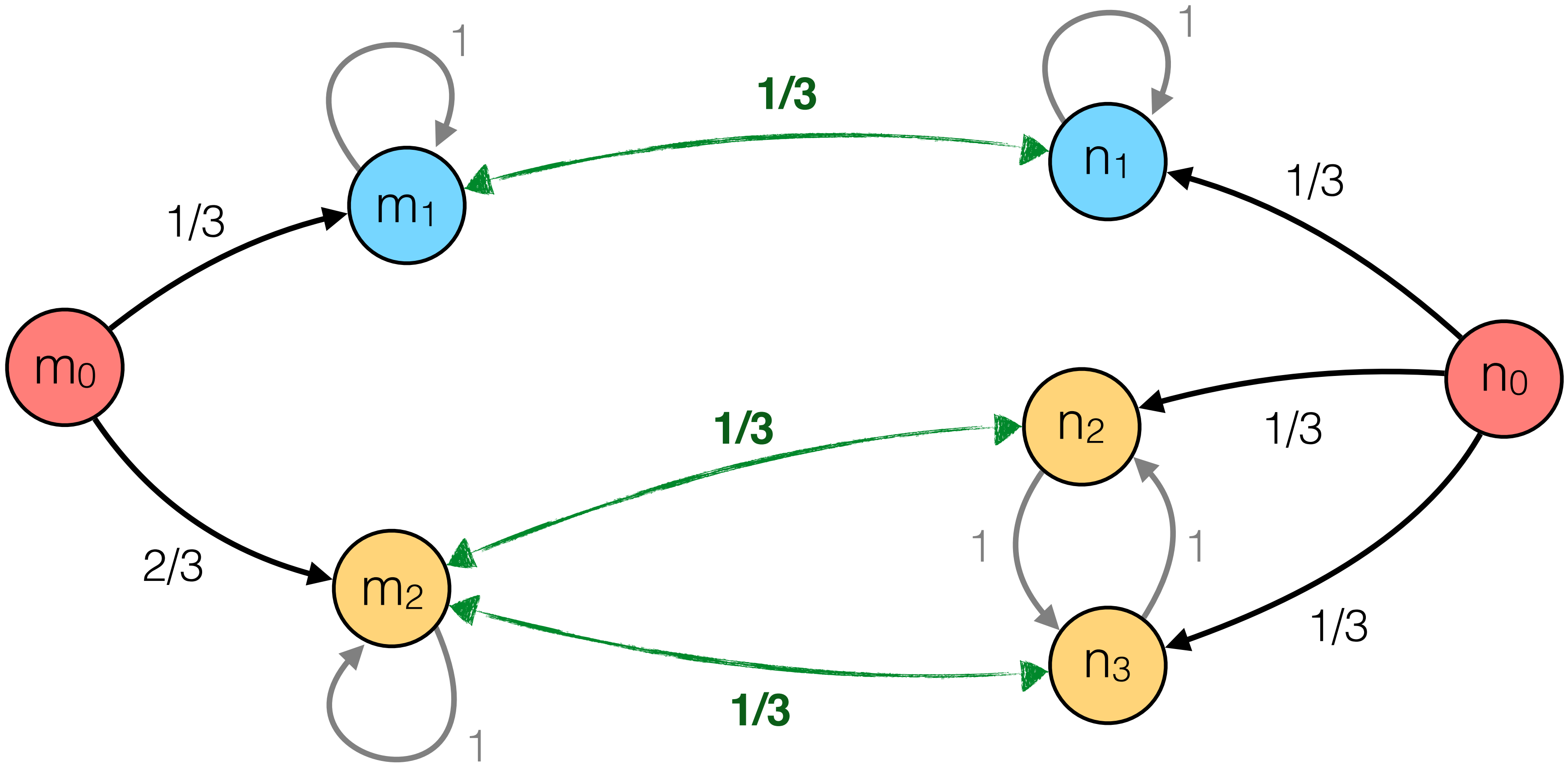
$$\sum_{u,v \in M} \omega(u,v)\ \mathbb{1}_R(u,v) \overset{?}{=} 1$$



$$\sum_{u,v \in M} \omega(u,v)\ \mathbb{1}_R(u,v) \stackrel{?}{=} 1$$



$$\sum_{u,v \in M} \omega(u,v)\ \mathbb{1}_R(u,v) \stackrel{?}{=} 1$$

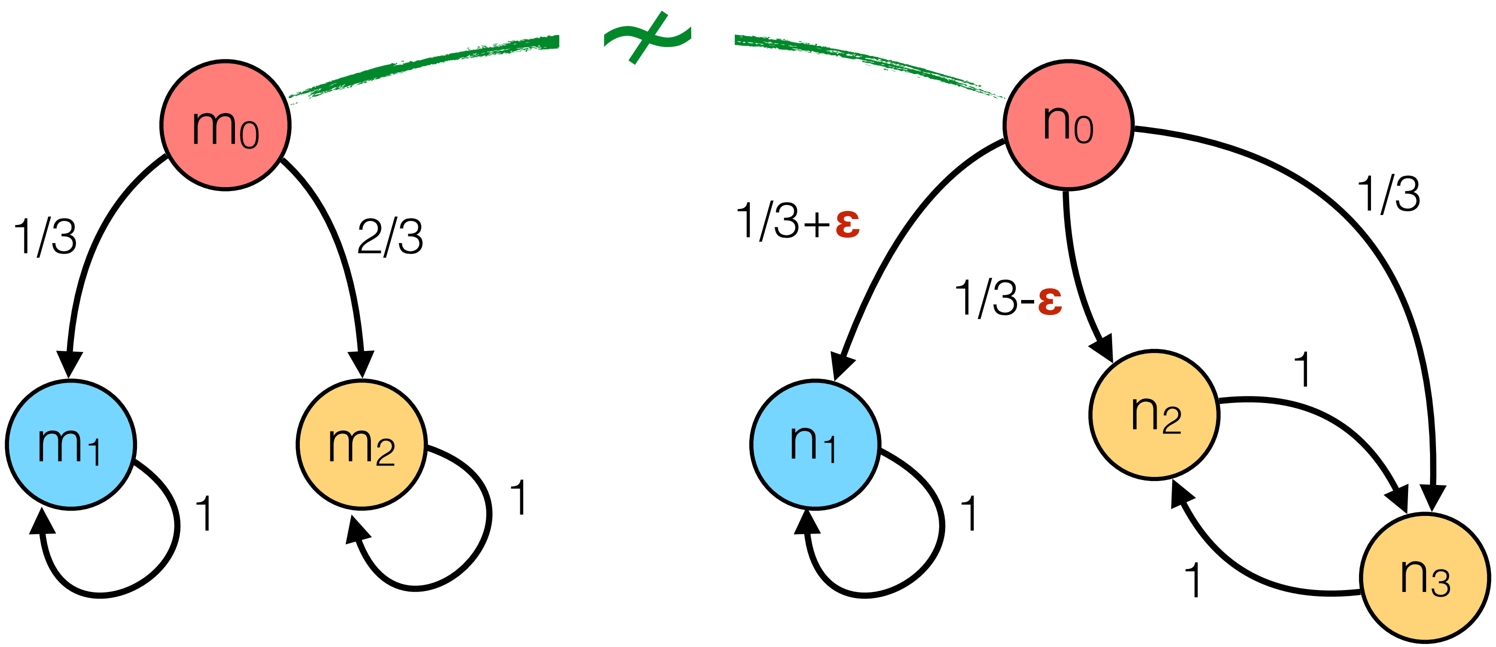


$$\sum_{u,v \in M} \omega(u,v)\ \mathbb{1}_R(u,v) \stackrel{?}{=} 1$$

Bisimilarity is not robust

Fundamental problem

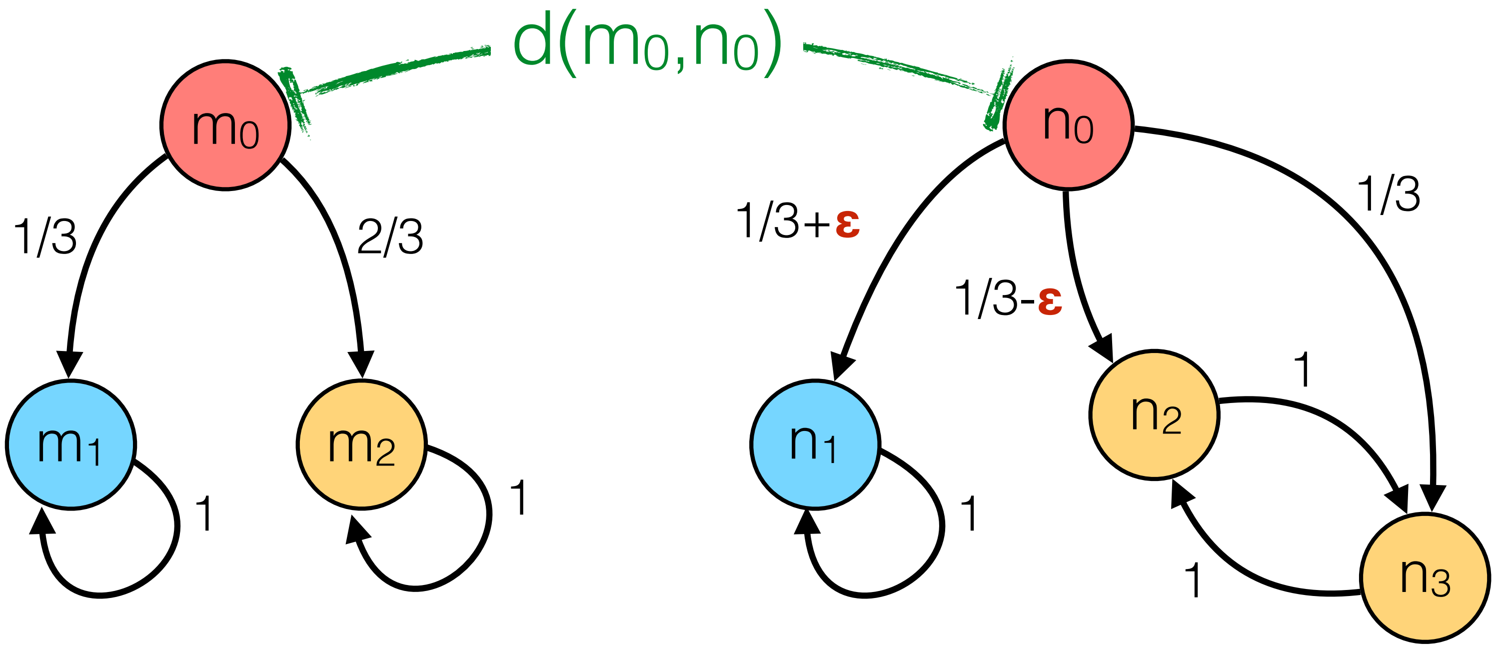
Smolka (1990) observed that behavioral equivalences are not robust for systems with real-valued data

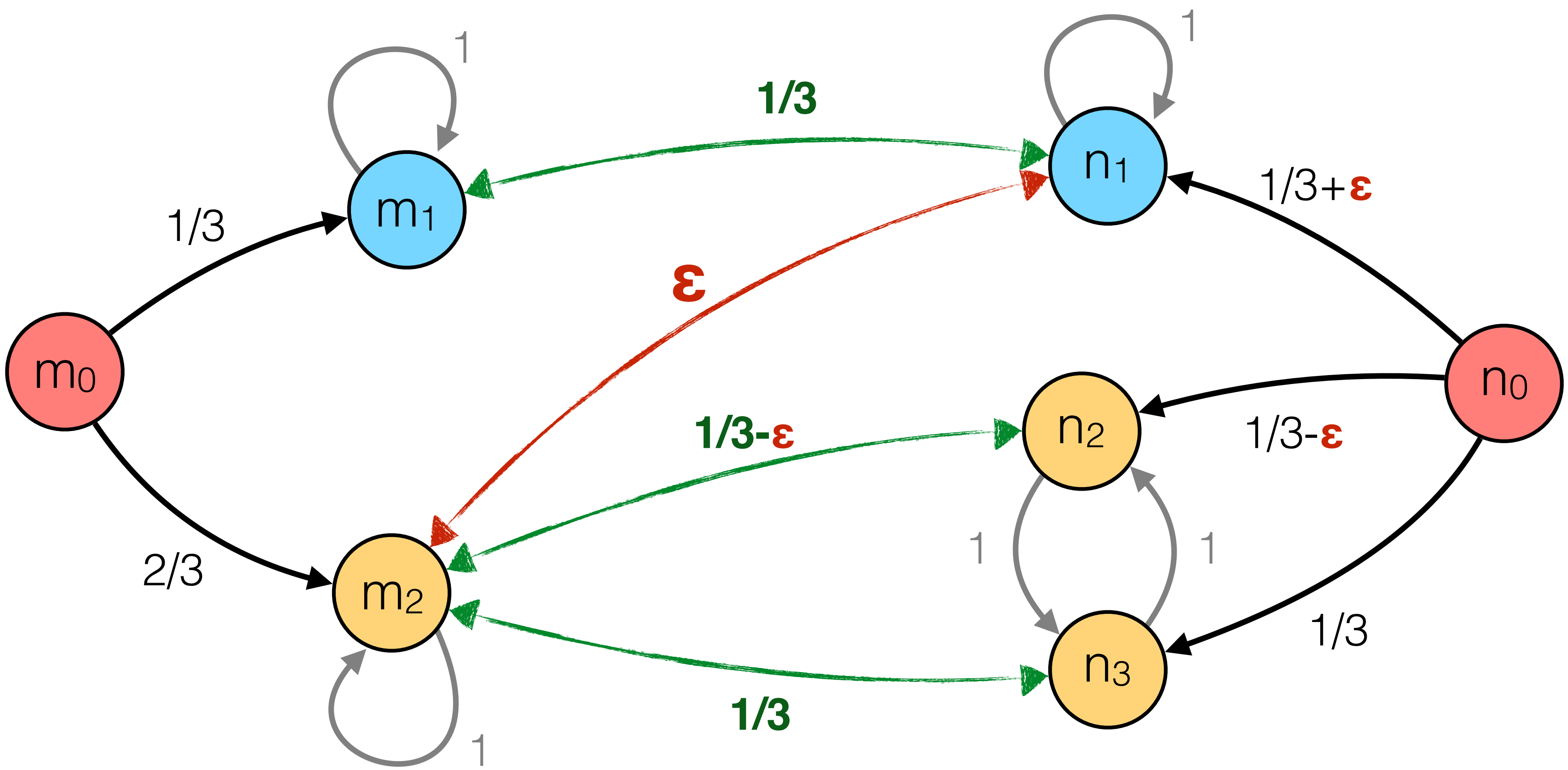


Behavioral Pseudometric

Robust Alternative

Equivalence Relation		Pseudometric
R: $M \times M \rightarrow \{true, false\}$	$\Rightarrow$	d: $M \times M \rightarrow [0,1]$





$$\text{minimize} \sum_{u,v \in M} \omega(u,v)\, d(u,v)$$

A quantitative generalization of probabilistic bisimilarity

The **λ**-discounted *probabilistic bisimilarity pseudometric* is the smallest $d_\lambda : M \times M \to [0,1]$ such that

$$d_\lambda(m,n) = \begin{cases} 1 & \text{if } \ell(m) \neq \ell(n) \\ \min\limits_{\omega \in \Omega(\tau(m),\tau(n))} \lambda \sum\limits_{u,v \in M} \omega(u,v)\, d_\lambda(u,v) & \text{otherwise} \end{cases}$$

A quantitative generalization of probabilistic bisimilarity

The λ-discounted *probabilistic bisimilarity pseudometric* is the smallest $d_\lambda: M \times M \rightarrow [0,1]$ such that

$$d_\lambda(m,n) = \begin{cases} 1 & \text{if } \ell(m) \neq \ell(n) \\ \min_{\omega \in \Omega(\tau(m),\tau(n))} \lambda \sum_{u,v \in M} \omega(u,v)\, d_\lambda(u,v) & \text{otherwise} \end{cases}$$

Kantorovich distance

$$K(d)(\mu,\nu) = \min_{\omega \in \Omega(\mu,\nu)} \sum_{u,v \in M} \omega(u,v)\, d(u,v)$$

Remarkable properties

Theorem (Desharnais et. al 99)

$$m \sim n \quad \text{iff} \quad d_\lambda(m,n) = 0$$

Theorem (Chen, van Breugel, Worrell 12)

The probabilistic bisimilarity distance
can be computed in polynomial time

Relation with Model Checking

Theorem (Chen, van Breugel, Worrell 12)

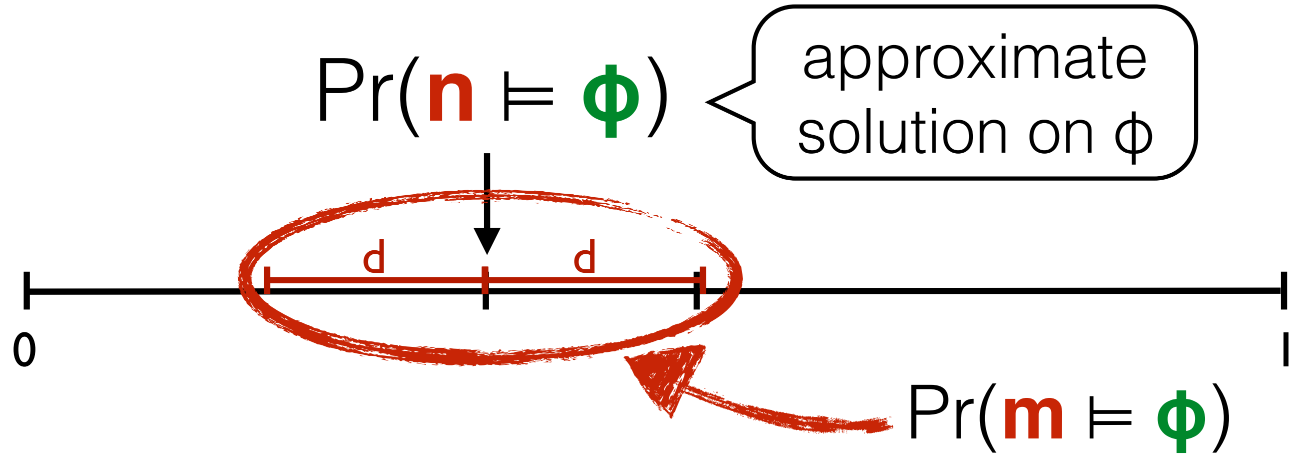
For all $\phi \in$ LTL $\quad | \Pr(m \vDash \phi) - \Pr(n \vDash \phi) | \leq d_1(m,n)$

Relation with Model Checking

Theorem (Chen, van Breugel, Worrell 12)

For all $\phi \in$ LTL $\quad | \Pr(m \models \phi) - \Pr(n \models \phi) | \leq d_1(m,n)$

...imagine that $|M| \gg |N|$, we can use N in place of M



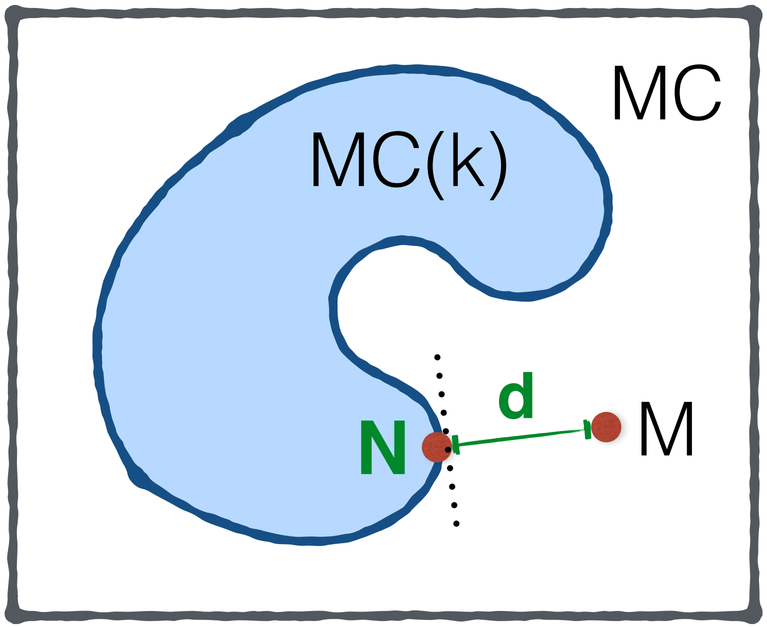
Metric-based State Space Reduction

Closest Bounded Approximant (CBA)

Minimum Significant Approximant Bound (MSAB)

Metric-based State Space Reduction

Closest Bounded Approximant (CBA)

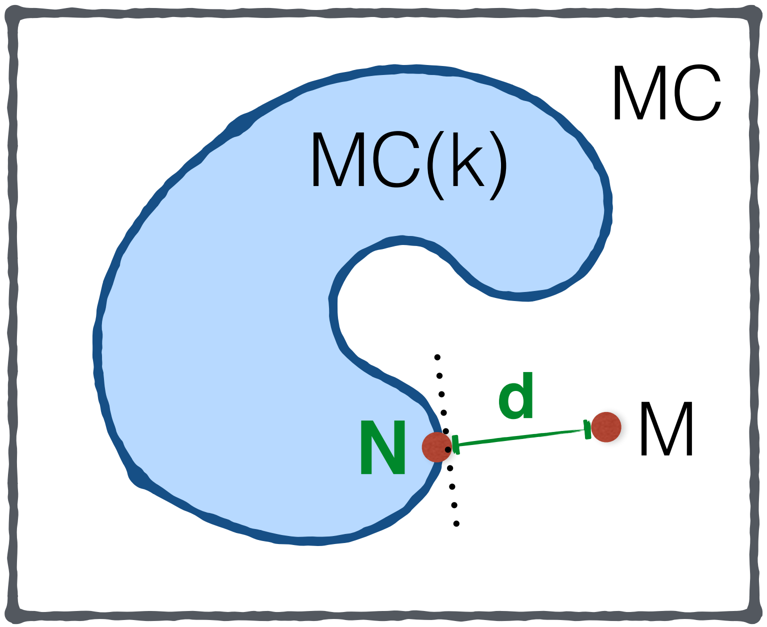


Minimum Significant Approximant Bound (MSAB)

Metric-based State Space Reduction

Closest Bounded Approximant (CBA)

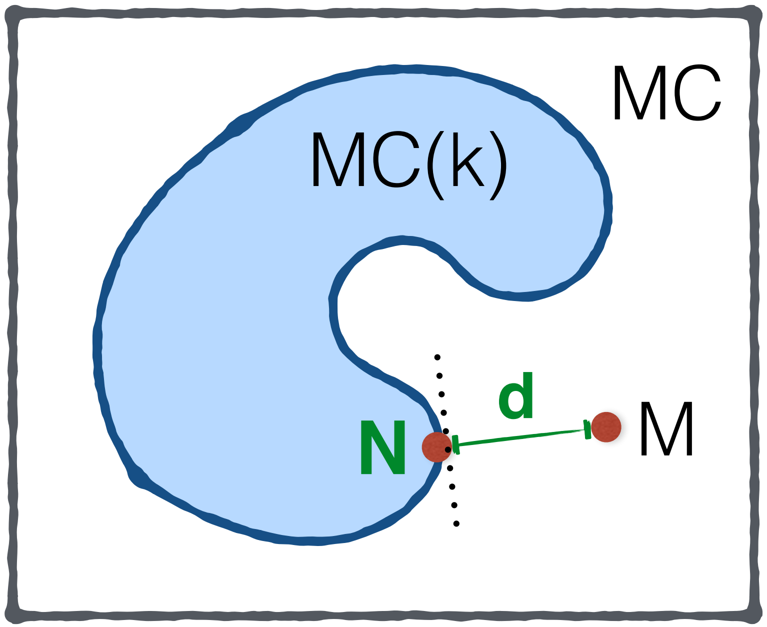
Minimum Significant Approximant Bound (MSAB)



minimize d

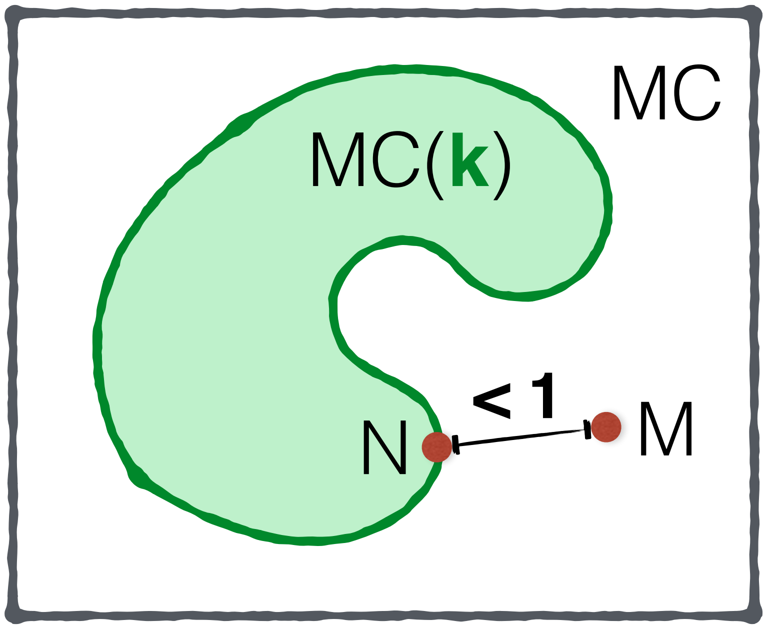
Metric-based State Space Reduction

Closest Bounded Approximant (CBA)



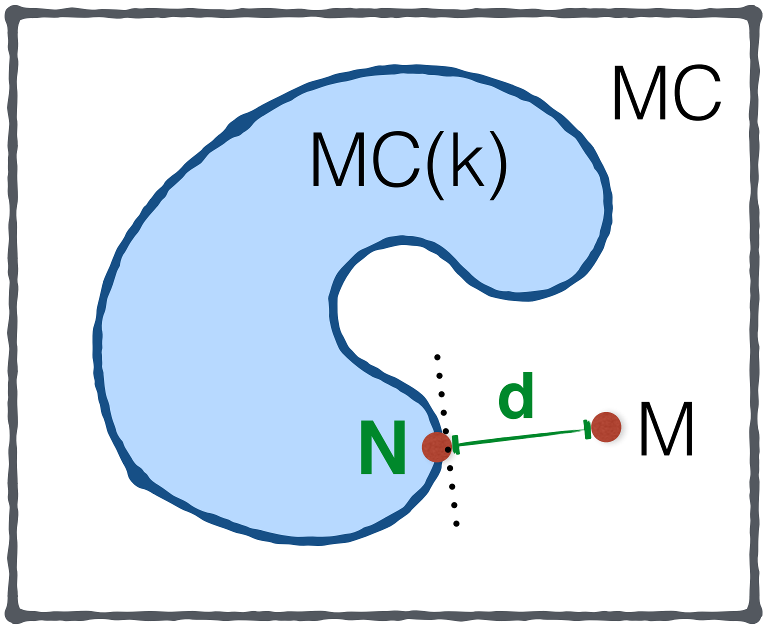
minimize d

Minimum Significant Approximant Bound (MSAB)



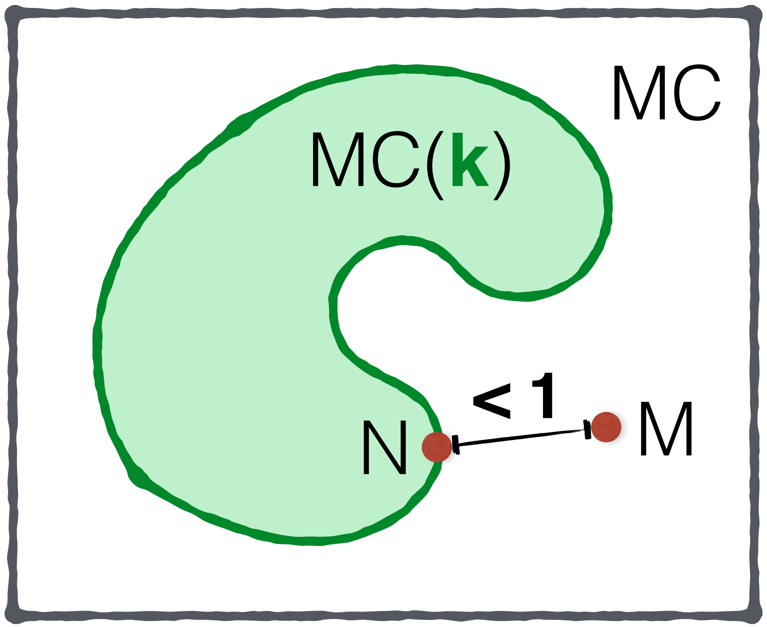
Metric-based State Space Reduction

Closest Bounded Approximant (CBA)



minimize d

Minimum Significant Approximant Bound (MSAB)



minimize k

List of our Results

- CBA as bilinear program
- The CBA's threshold problem is
  - NP-hard (complexity lower bound)
  - PSPACE (complexity upper bound)
- The MSAB's threshold problem is NP-complete
- Expectation Maximization heuristic for CBA

The CBA-λ problem

The Closest Bounded Approximant wrt d_λ

Instance: An MC M, and a positive integer k

Ouput: An MC Ñ, with at most k states minimizing $d_\lambda(m_0,\tilde{n}_0)$

$$d_\lambda(m_0,\tilde{n}_0) = \inf \{ d_\lambda(m_0,n_0) \mid N \in MC(k) \}$$

we get a solution iff the infimum is a minimum

The CBA-λ problem

The Closest Bounded Approximant wrt d_λ

Instance: An MC M, and a positive integer k

Ouput: An MC Ñ, with at most k states minimizing $d_\lambda(m_0,\tilde{n}_0)$

$$d_\lambda(m_0,\tilde{n}_0) = \inf \{ d_\lambda(m_0,n_0) \mid N \in MC(k) \}$$

we get a solution iff the infimum is a minimum

generalization of bisimilarity quotient

CBA-λ as a Bilinear Program

$$d_\lambda(m_0,\tilde{n}_0) = \inf \{ d_\lambda(m_0,n_0) \mid N \in MC(k) \}$$
$$= \inf \{ d(m_0,n_0) \mid \Gamma_\lambda(d) \leq d,\ N \in MC(k) \}$$

CBA-λ as a Bilinear Program

$d_\lambda(m_0,\tilde{n}_0) = \inf \{ d_\lambda(m_0,n_0) \mid N \in MC(k) \}$
$= \inf \{ d(m_0,n_0) \mid \Gamma_\lambda(d) \leq d,\ N \in MC(k) \}$

mimimize	d_{m_0,n_0}	
such that	$d_{m,n} = 1$	$\ell(m) \neq \alpha(n)$
	$\lambda \sum_{(u,v)\in M\times N} c^{m,n}_{u,v} \cdot d_{u,v} \leq d_{m,n}$	$\ell(m) = \alpha(n)$
	$\sum_{v\in N} c^{m,n}_{u,v} = \tau(m)(u)$	$m, u \in M,\ n \in N$
	$\sum_{u\in M} c^{m,n}_{u,v} = \theta_{n,v}$	$m \in M,\ n, v \in N$
	$c^{m,n}_{u,v} \geq 0$	$m, u \in M,\ n, v \in N$

CBA-λ as a Bilinear Program

$d_\lambda(m_0,\tilde{n}_0) = \inf \{ d_\lambda(m_0,n_0) \mid N \in MC(k) \}$

$= \inf \{ d(m_0,n_0) \mid \Gamma_\lambda(d) \leq d, N \in MC(k) \}$

mimimize	d_{m_0,n_0}	
such that	$d_{m,n} = 1$	$\ell(m) \neq \alpha(n)$
	$\lambda \sum_{(u,v) \in M \times N} c_{u,v}^{m,n} \cdot d_{u,v} \leq d_{m,n}$	$\ell(m) = \alpha(n)$
	$\sum_{v \in N} c_{u,v}^{m,n} = \tau(m)(u)$	$m, u \in M,\ n \in N$
	$\sum_{u \in M} c_{u,v}^{m,n} = \theta_{n,v}$	$m \in M,\ n, v \in N$
	$c_{u,v}^{m,n} \geq 0$	$m, u \in M,\ n, v \in N$

CBA-λ as a Bilinear Program

$$d_\lambda(m_0,\tilde{n}_0) = \inf \{\, d_\lambda(m_0,n_0) \mid N \in MC(k) \,\}$$
$$= \inf \{\, d(m_0,n_0) \mid \Gamma_\lambda(d) \leq d,\ N \in MC(k) \}$$

mimimize d_{m_0,n_0}

such that

$d_{m,n} = 1$	$\ell(m) \neq \alpha(n)$
$\lambda \sum_{(u,v)\in M\times N} c^{m,n}_{u,v} \cdot d_{u,v} \leq d_{m,n}$	$\ell(m) = \alpha(n)$
$\sum_{v \in N} c^{m,n}_{u,v} = \tau(m)(u)$	
$\sum_{u \in M} c^{m,n}_{u,v} = \theta_{n,v}$	
$c^{m,n}_{u,v} \geq 0$	$m, u \in M,\ n, v \in N$

what labels should the MC N have?

CBA-λ as a Bilinear Program
(continued)

Lemma (Meaningful labels)

For any N∈MC(k), there exists N'∈MC(k) with labels taken from M, such that $d_\lambda(M,N) \geq d_\lambda(M,N')$

CBA-λ as a Bilinear Program (continued)

Lemma (Meaningful labels)

For any N∈MC(k), there exists N'∈MC(k) with labels taken from M, such that $d_\lambda(M,N) \geq d_\lambda(M,N')$

$$
\begin{array}{lll}
\text{mimimize} & d_{m_0,n_0} & \\
\text{such that} & l_{m,n} \leq d_{m,n} \leq 1 & m \in M,\ n \in N \\
& \lambda \sum_{(u,v)\in M\times N} c^{m,n}_{u,v} \cdot d_{u,v} \leq d_{m,n} & m \in M,\ n \in N \\
& l_{m,n} \cdot l_{u,n} = 0 & n \in N,\ \ell(m) \neq \ell(u) \\
& l_{m,n} + l_{u,n} = 1 & n \in N,\ \ell(m) \neq \ell(u) \\
& l_{m,n} = l_{u,n} & n \in N,\ \ell(m) = \ell(u) \\
& \sum_{m\in M} l_{m,n} \leq |M| - 1 & n \in N \\
& \sum_{v\in N} c^{m,n}_{u,v} = \tau(m)(u) & m,u \in M,\ n \in N \\
& \sum_{u\in M} c^{m,n}_{u,v} = \theta_{n,v} & m \in M,\ n,v \in N \\
& c^{m,n}_{u,v} \geq 0 & m,u \in M,\ n,v \in N
\end{array}
$$

CBA-λ as a Bilinear Program
(continued)

Lemma (Meaningful labels)

For any N∈MC(k), there exists N'∈MC(k) with labels taken from M, such that $d_\lambda(M,N) \geq d_\lambda(M,N')$

$$
\begin{array}{lll}
\text{mimimize} & d_{m_0,n_0} & \\
\text{such that} & l_{m,n} \leq d_{m,n} \leq 1 & m \in M,\ n \in N \\
& \lambda \sum_{(u,v) \in M \times N} c^{m,n}_{u,v} \cdot d_{u,v} \leq d_{m,n} & m \in M,\ n \in N \\
& l_{m,n} \cdot l_{u,n} = 0 & n \in N,\ \ell(m) \neq \ell(u) \\
& l_{m,n} + l_{u,n} = 1 & n \in N,\ \ell(m) \neq \ell(u) \\
& l_{m,n} = l_{u,n} & n \in N,\ \ell(m) = \ell(u) \\
& \sum_{m \in M} l_{m,n} \leq |M| - 1 & n \in N \\
& \sum_{v \in N} c^{m,n}_{u,v} = \tau(m)(u) & m,u \in M,\ n \in N \\
& \sum_{u \in M} c^{m,n}_{u,v} = \theta_{n,v} & m \in M,\ n,v \in N \\
& c^{m,n}_{u,v} \geq 0 & m,u \in M,\ n,v \in N
\end{array}
$$

CBA-λ as a Bilinear Program (continued)

this characterization has two main consequences…

1. CBA-λ admits always a solution
2. CBA-λ is decidable

Complexity of CBA-λ

"To study the complexity of an optimization problem
one has to look at its decision variant"

(C. Papadimitriou)

Complexity of CBA-λ

"To study the complexity of an optimization problem one has to look at its decision variant"

(C. Papadimitriou)

Bounded Approximant threshold wrt d_λ

Instance: An MC M, a positive integer k, and a **rational $\varepsilon > 0$**

Ouput: **yes** iff there exists N with at most k states such that $d_\lambda(m_0,n_0)$ **$\leq \varepsilon$**

Complexity upper bound

Theorem

BA-λ is in **PSPACE**

Proof sketch: *we can encode the question $\langle M,k,\varepsilon\rangle \in$ BA-λ to that of checking the feasibility of a set of bilinear inequalities. This can be encoded as a decision problem for the existential theory of the reals, thus it can be solved in PSPACE [Canny—STOC88].*

Complexity lower bound

Theorem

BA-λ is **NP-hard**

Complexity lower bound

Theorem

BA-λ is **NP-hard**

unlikely to solve CBA as simple linear program

The MSAB-λ problem

The Minimum Significant Approximant Bound wrt d_λ

Instance: An MC M

Ouput: The smallest k such that $d_\lambda(m_0,n_0)<1$, for some $N \in MC(k)$

The MSAB-λ problem

The Minimum Significant Approximant Bound wrt d_λ

Instance: An MC M

Ouput: The smallest k such that $d_\lambda(m_0,n_0)<1$, for some N∈MC(k)

For **λ<1**, the MSAB-λ problem is trivial, because the solution is always k=1

The MSAB-λ problem

The Minimum Significant Approximant Bound wrt d_λ

Instance: An MC M

Ouput: The smallest k such that $d_\lambda(m_0,n_0)<1$, for some $N \in MC(k)$

For **λ<1**, the MSAB-λ problem is trivial, because the solution is always k=1

For **λ=1**, the same problem is surprisingly difficult…

Complexity of MSAB-1

…as before we should look at its decision variant

Complexity of MSAB-1

…as before we should look at its decision variant

Significant Bounded Approximant wrt d_1

Instance: An MC M and a **positive k**

Ouput: yes iff there exists N with **at most k** states such that $d_1(m_0,n_0)<1$.

Complexity of MSAB-1

…as before we should look at its decision variant

Significant Bounded Approximant wrt d_1

Instance: An MC M and a **positive k**

Ouput: **yes** iff there exists N with **at most k** states such that $d_1(m_0,n_0)<1$.

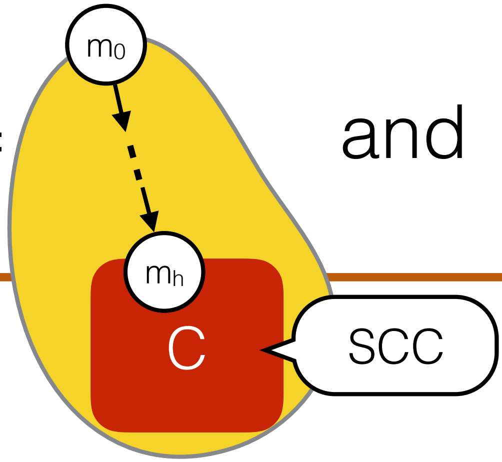
Theorem

SBA-1 is **NP-complete**

SBA-1 $\subseteq$ **NP**

Lemma

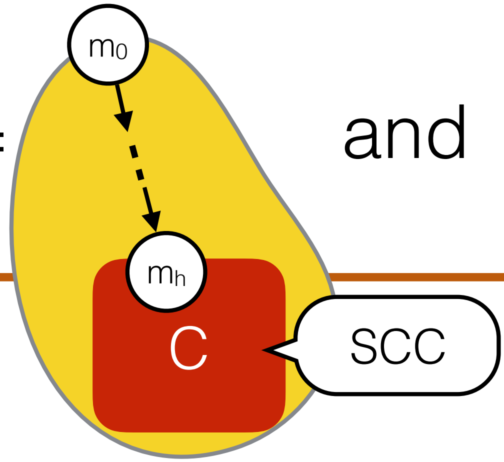
Assume M be maximally collapsed. Then,

$\langle M,k\rangle \in$ SBA-1 iff $\mathcal{G}(M)$ =  and $h+|C| \leq k$

SBA-1 ⊆ **NP**

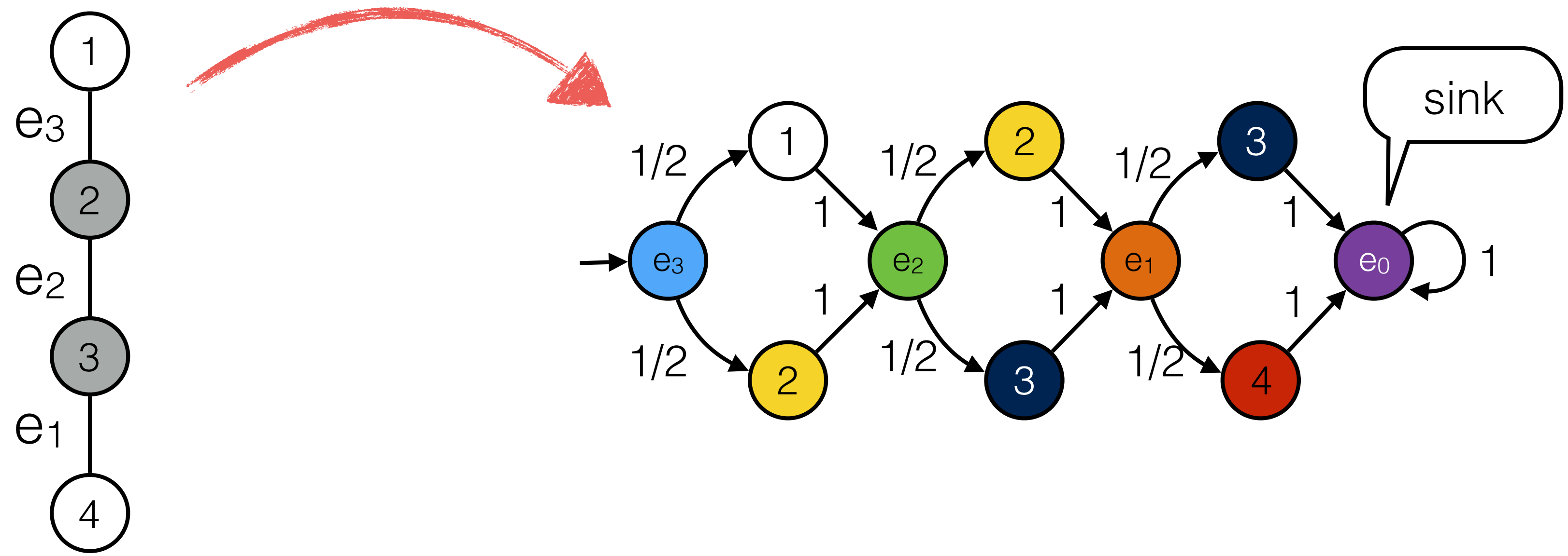
Lemma

Assume M be maximally collapsed. Then,

$\langle M,k\rangle \in$ SBA-1 iff $\mathcal{G}(M)$ =  and $h+|C| \leq k$

Proof sketch: *compute with Tarjan all the SCCs of $\mathcal{G}(M)$. Then non deterministically choose an SCC and a path to it. In poly-time we can check the size of the path and of the SCC.*

SBA-1 is **NP-hard**



***Proof sketch:** by reduction to VERTEX COVER:*

⟨G,h⟩∈VERTEX COVER iff ⟨M_G, h+m+1⟩∈SBA-1

Towards an Algorithm…

Towards an Algorithm…

- The CBA can be solved as a bilinear program. Theoretically nice, but practically unfeasible!
  (our implementation in PENBMI can handle MCs with at most 5 states…)

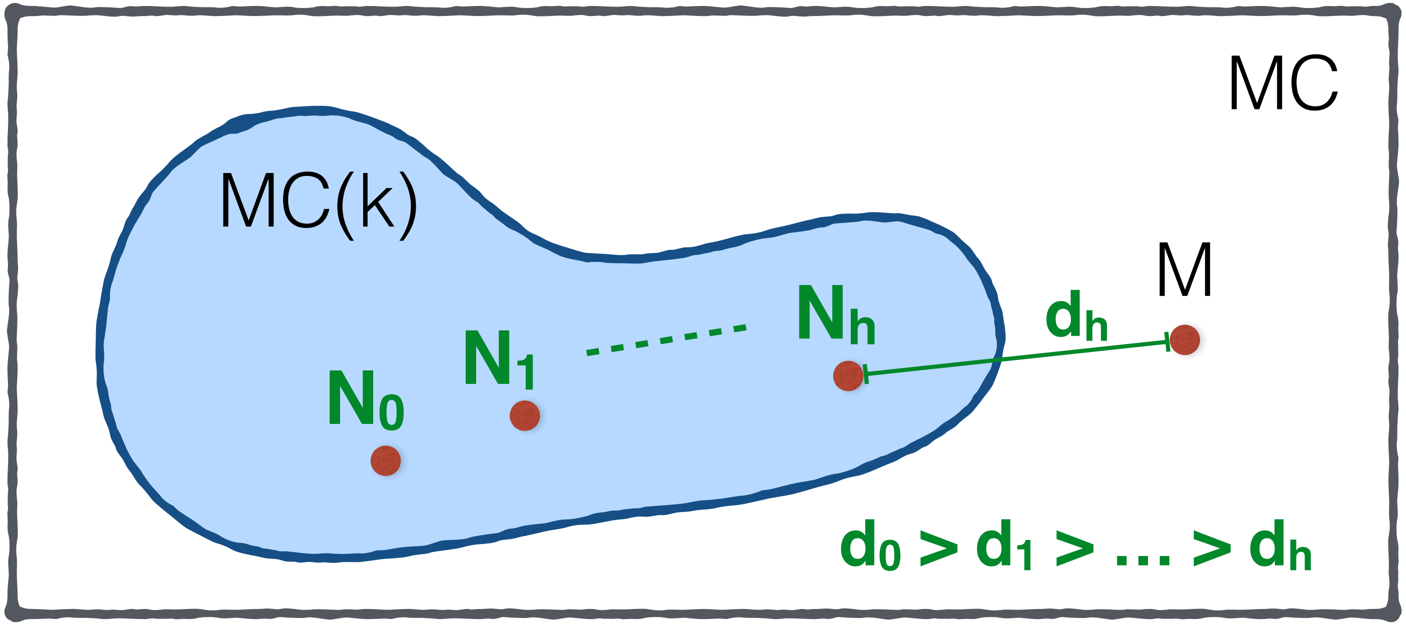
Towards an Algorithm…

- The CBA can be solved as a bilinear program. Theoretically nice, but practically unfeasible!
  (our implementation in PENBMI can handle MCs with at most 5 states…)

- We are happy with **sub-optimal solutions** if they can be obtained by a practical algorithm.

EM-like Algorithm

- Given the MC M and an initial approximant N_0
- it produces a sequence N_0, …, N_h of approximants having strictly decreasing distance from M
- N_h may be a sub-optimal solution of CBA-λ



EM-like Algorithm

Algorithm 1

Input: $\mathcal{M} = (M, \tau, \ell)$, $\mathcal{N}_0 = (N, \theta_0, \alpha)$, and $h \in \mathbb{N}$.

1. $i \leftarrow 0$
2. **repeat**
3. $\quad i \leftarrow i + 1$
4. $\quad$ compute $\mathcal{C} \in \Omega(\mathcal{M}, \mathcal{N}_{i-1})$ such that $\delta_\lambda(\mathcal{M}, \mathcal{N}_{i-1}) = \gamma_\lambda^{\mathcal{C}}(\mathcal{M}, \mathcal{N}_{i-1})$
5. $\quad \theta_i \leftarrow \textsc{UpdateTransition}(\theta_{i-1}, \mathcal{C})$
6. $\quad \mathcal{N}_i \leftarrow (N, \theta_i, \alpha)$
7. **until** $\delta_\lambda(\mathcal{M}, \mathcal{N}_i) > \delta_\lambda(\mathcal{M}, \mathcal{N}_{i-1})$ or $i \geq h$
8. **return** $\mathcal{N}_{i-1}$

EM-like Algorithm

Algorithm 1

Input: $\mathcal{M} = (M, \tau, \ell)$, $\mathcal{N}_0 = (N, \theta_0, \alpha)$, and $h \in \mathbb{N}$.

1. $i \leftarrow 0$
2. **repeat**
3. $\quad i \leftarrow i + 1$
4. $\quad$ compute $\mathcal{C} \in \Omega(\mathcal{M}, \mathcal{N}_{i-1})$ such that $\delta_\lambda(\mathcal{M}, \mathcal{N}_{i-1}) = \gamma_\lambda^{\mathcal{C}}(\mathcal{M}, \mathcal{N}_{i-1})$
5. $\quad \theta_i \leftarrow \textsc{UpdateTransition}(\theta_{i-1}, \mathcal{C})$
6. $\quad \mathcal{N}_i \leftarrow (N, \theta_i, \alpha)$
7. **until** $\delta_\lambda(\mathcal{M}, \mathcal{N}_i) > \delta_\lambda(\mathcal{M}, \mathcal{N}_{i-1})$ or $i \geq h$
8. **return** $\mathcal{N}_{i-1}$

Intuitive Idea

UpdateTransition assigns greater probability to transitions that are most representative of the behavior of M

Two update heuristics

- Averaged Marginal (AM): given N_k we construct N_{k+1} by averaging the marginal of certain "coupling variables" obtained by optimizing the number of occurrences of the edges that are most likely to be seen in M.

- Averaged Expectations (AE): similar to the above, but now the N_{k+1} looks only the expectation of the number of occurrences of the edges likely to be found in M.

Case	$\|M\|$	k	$\lambda = 1$				$\lambda = 0.8$			
			δ_λ-init	δ_λ-final	#	time	δ_λ-init	δ_λ-final	#	time
IPv4 (AM)	23	5	0.775	0.054	3	4.8	0.576	0.025	3	4.8
	53	5	0.856	0.062	3	25.7	0.667	0.029	3	25.9
	103	5	0.923	0.067	3	116.3	0.734	0.035	3	116.5
	53	6	0.757	0.030	3	39.4	0.544	0.011	3	39.4
	103	6	0.837	0.032	3	183.7	0.624	0.017	3	182.7
	203	6	–	–	–	TO	–	–	–	TO
IPv4 (AE)	23	5	0.775	0.109	2	2.7	0.576	0.049	3	4.2
	53	5	0.856	0.110	2	14.2	0.667	0.049	3	21.8
	103	5	0.923	0.110	2	67.1	0.734	0.049	3	100.4
	53	6	0.757	0.072	2	21.8	0.544	0.019	3	33.0
	103	6	0.837	0.072	2	105.9	0.624	0.019	3	159.5
	203	6	–	–	–	TO	–	–	–	TO
DrkW (AM)	39	7	0.565	0.466	14	259.3	0.432	0.323	14	252.8
	49	7	0.568	0.460	14	453.7	0.433	0.322	14	420.5
	59	8	0.646	–	–	TO	0.423	–	–	TO
DrkW (AE)	39	7	0.565	0.435	11	156.6	0.432	0.321	2	28.6
	49	7	0.568	0.434	10	247.7	0.433	0.316	2	46.2
	59	8	0.646	0.435	10	588.9	0.423	0.309	2	115.7

Table 1. Comparison of the performance of EM algorithm on the IPv4 zeroconf protocol and the classic Drunkard's Walk w.r.t. the heuristics AM and AE.

What we have seen

Theoretical

Metric-based state space reduction for MCs

1. Closest Bounded Approximant (CBA)
   encoded as a bilinear program
2. Bounded Approximant (BA)
   PSPACE & NP-hard for all $\lambda \in (0,1]$
3. Significant Bounded Approximant (SBA)
   NP-complete for $\lambda = 1$

Practical

We proposed an EM-like method to obtain a sub-optimal approximants

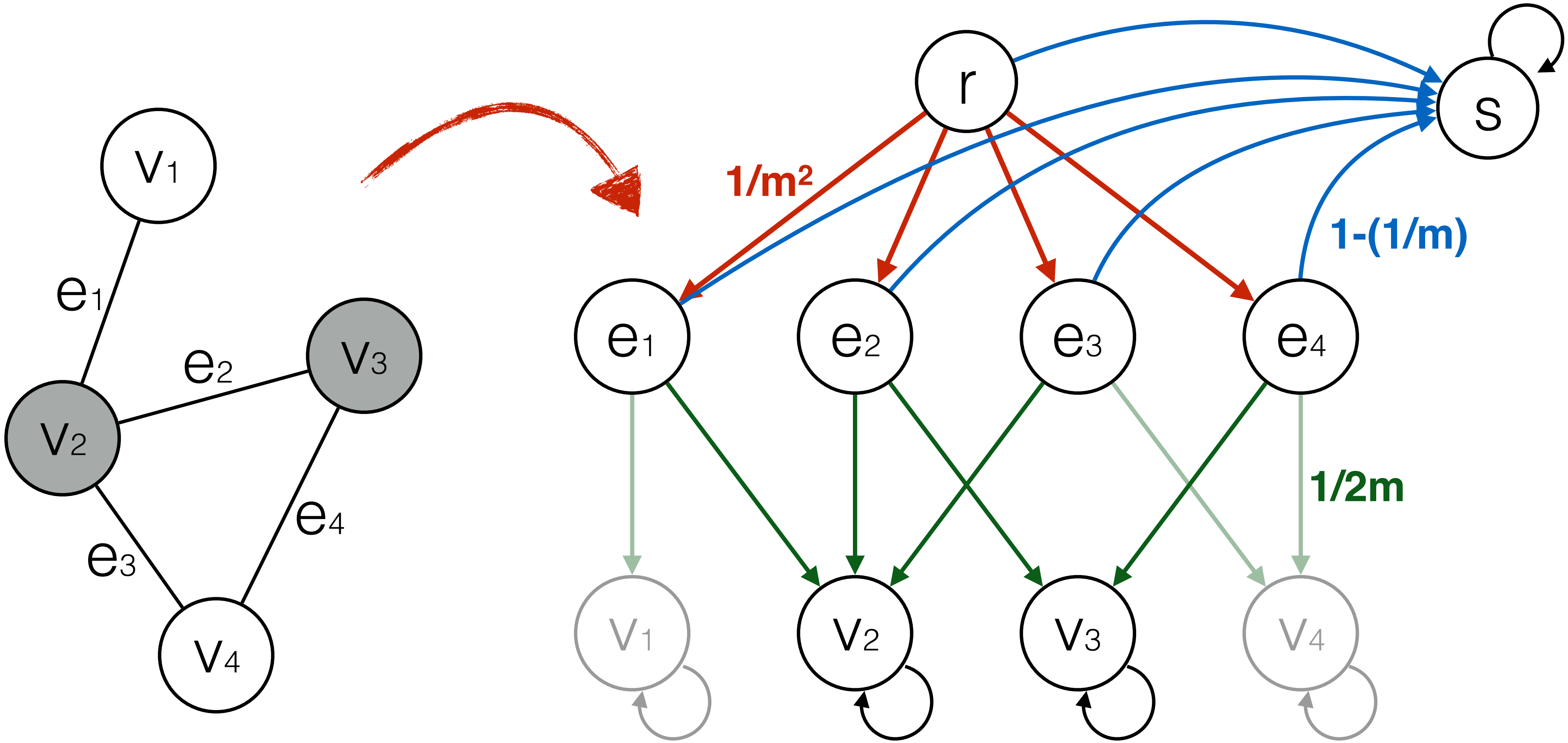
Future Work

- Is BA-λ SUM-OF-SQUARE-ROOTS-hard?
- Can we obtain a real/better EM-heuristics?
- What about different models/distances?

Thank you
for your attention

Appendix

BA-λ is NP-hard



$\langle G,h \rangle \in$ *VERTEX COVER* *iff* $\langle M_G, m+h+2, \lambda^2/2m^2 \rangle \in BA\text{-}\lambda$