Complexity Metrics for Hierarchical State Machines

Mathew Hall

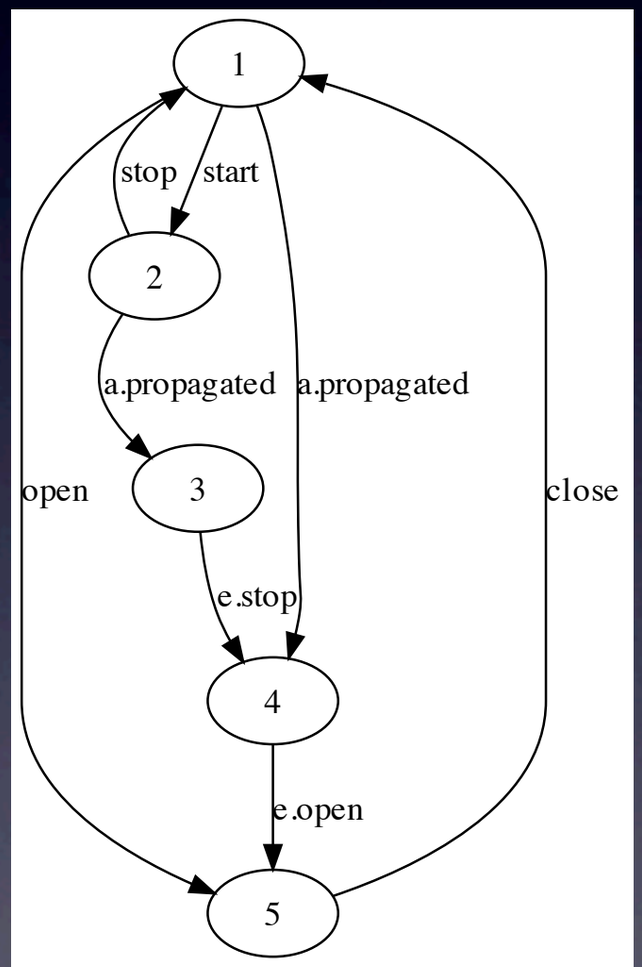
Overview

- PhD Topic
- Hierarchical State Machines
- Metrics
- PhD directions
- Further work

PhD Topic

Reverse engineered state machines

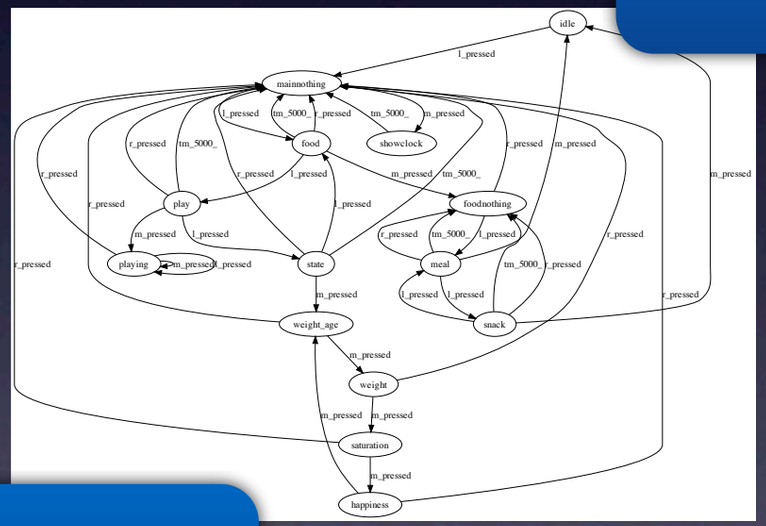
State Machines



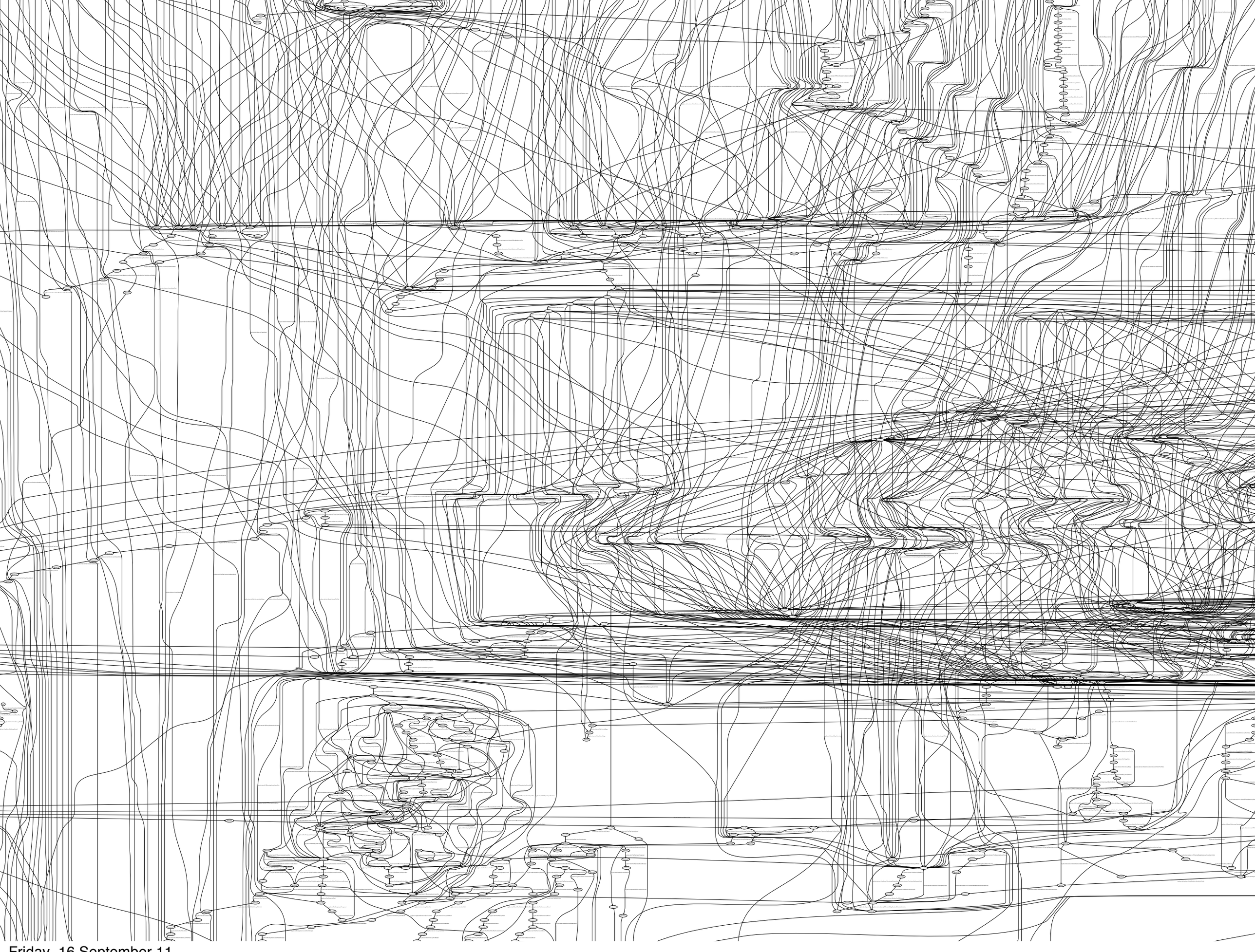
State Machines

Convey Behaviour

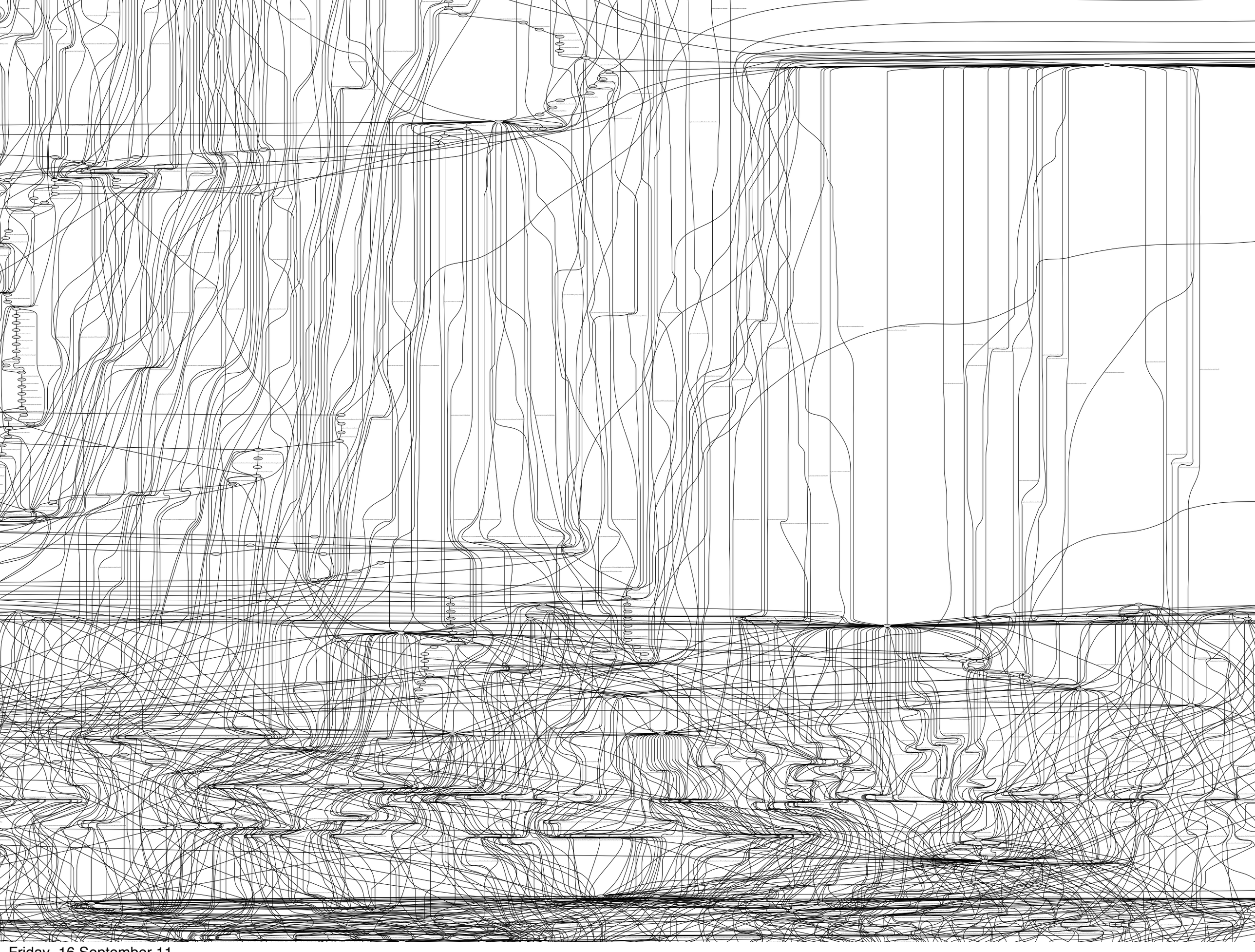
Implementation Independent



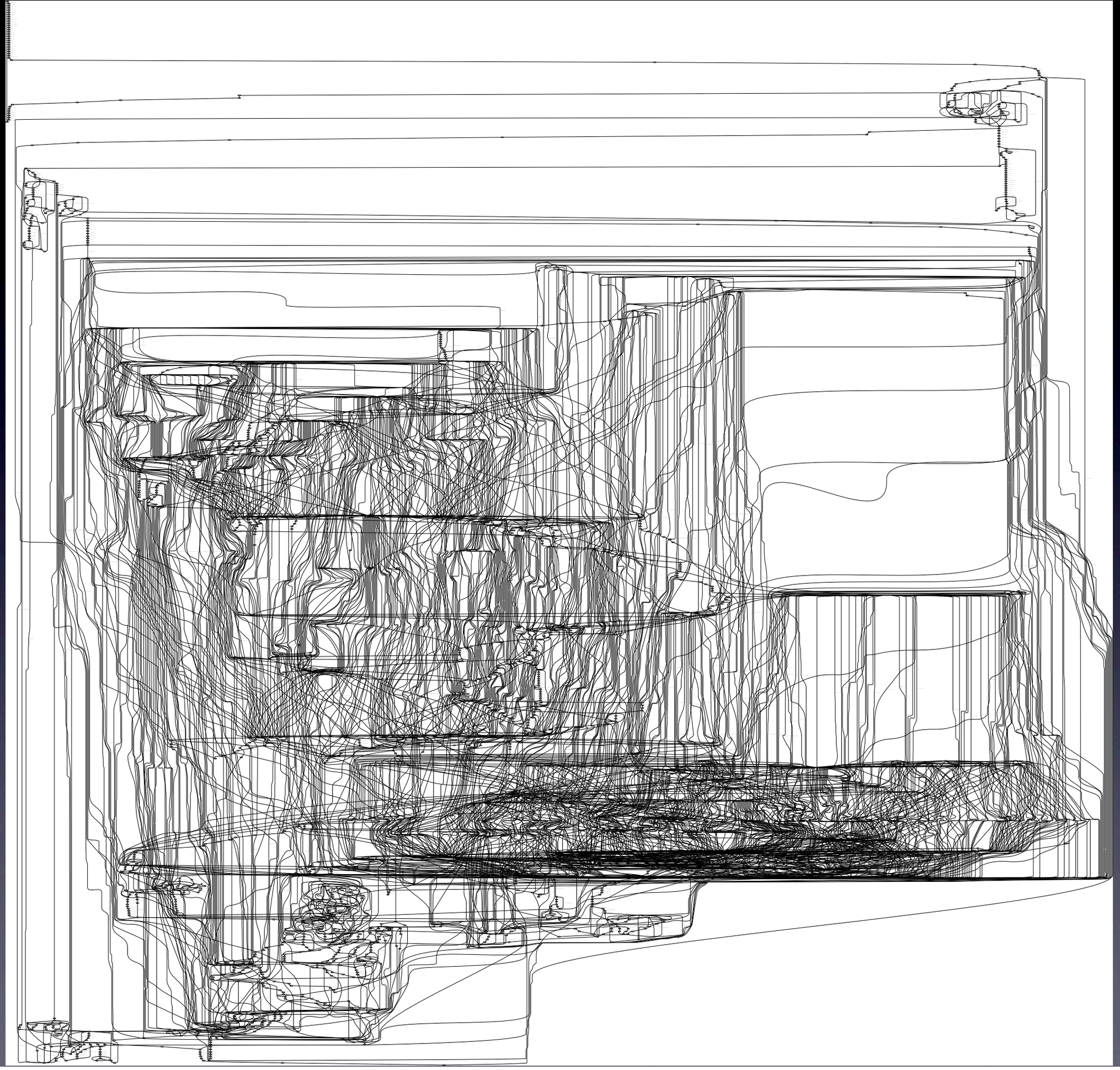
Can be Reverse Engineered

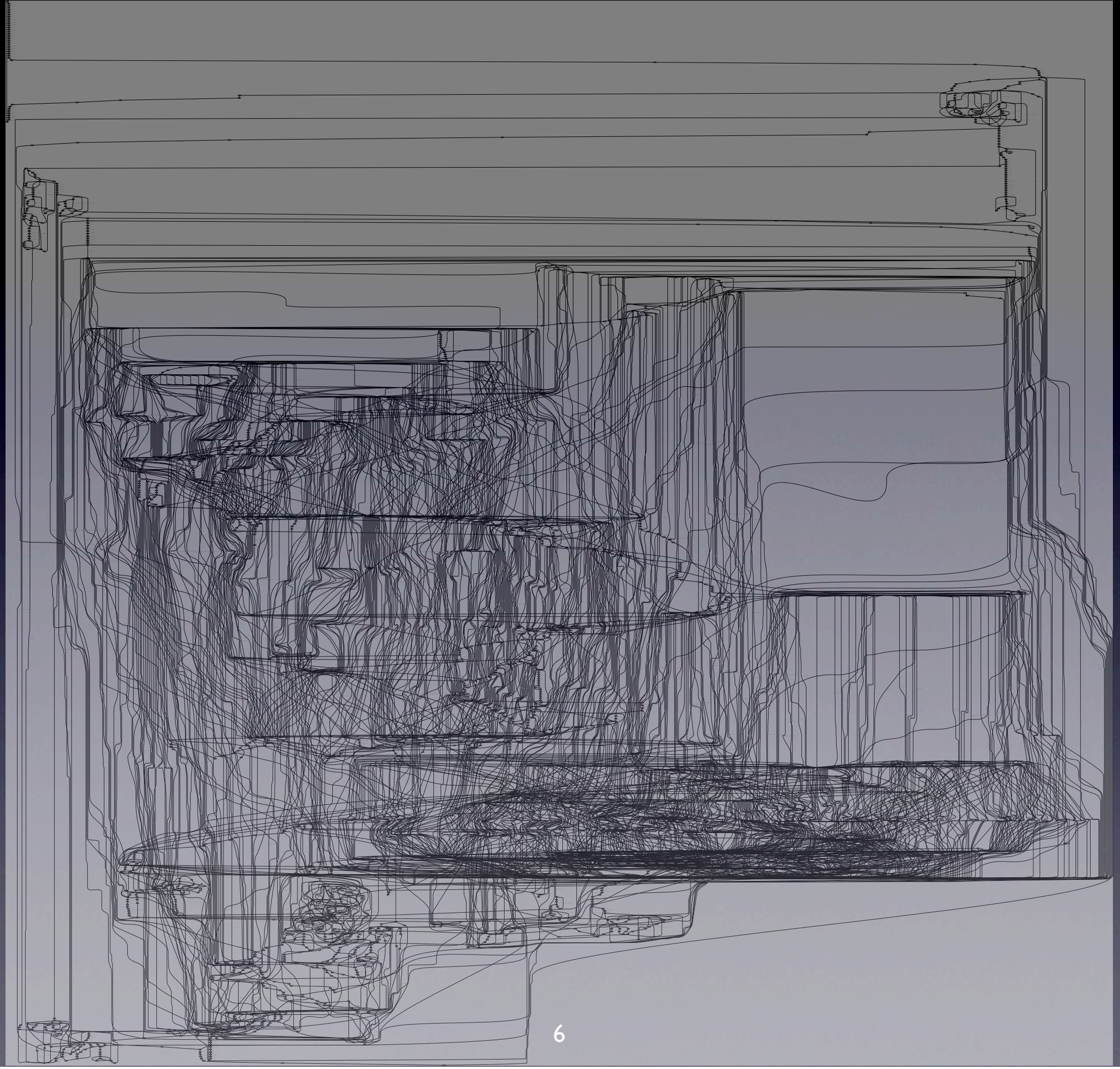


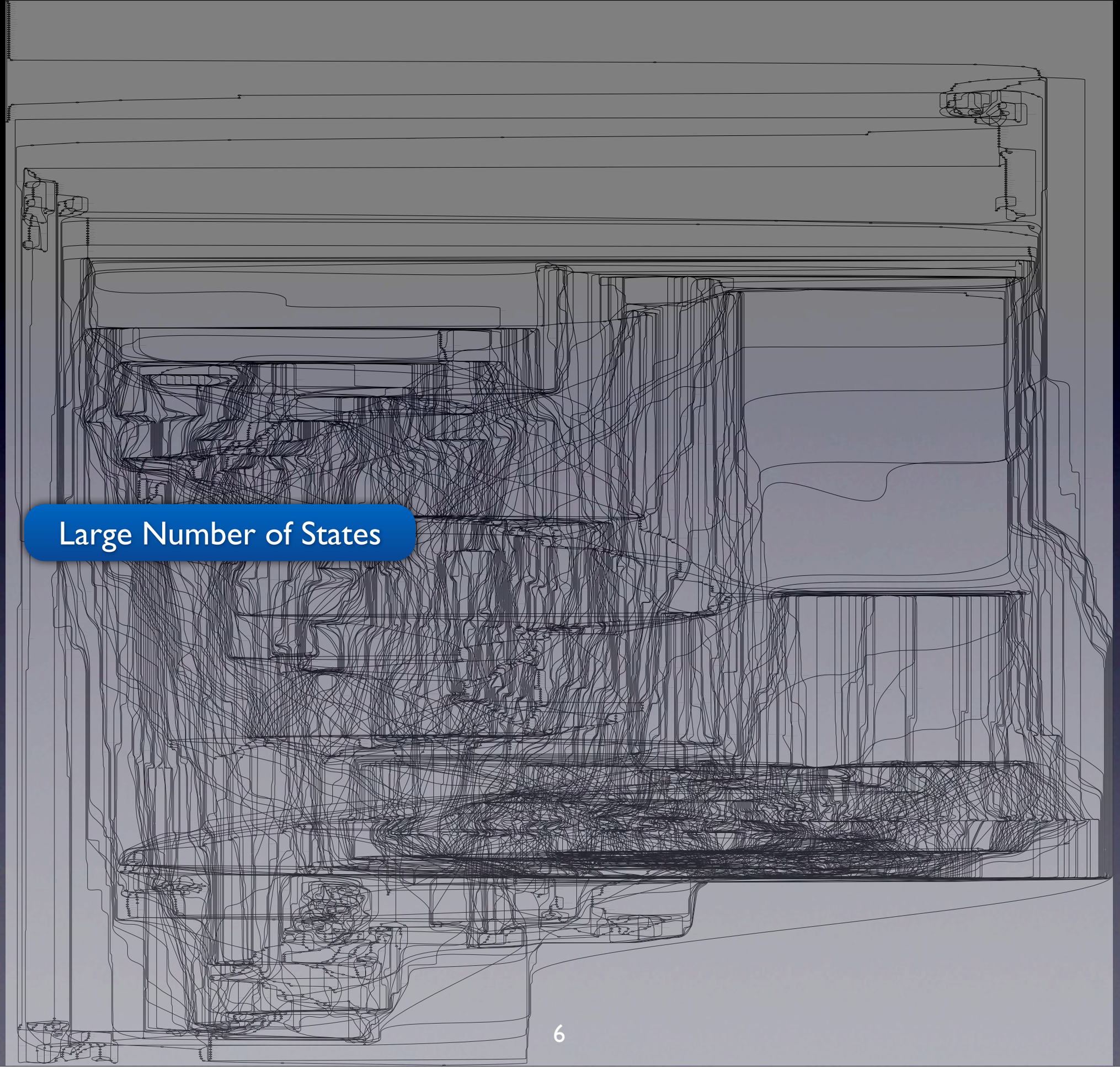
Friday, 16 September 11

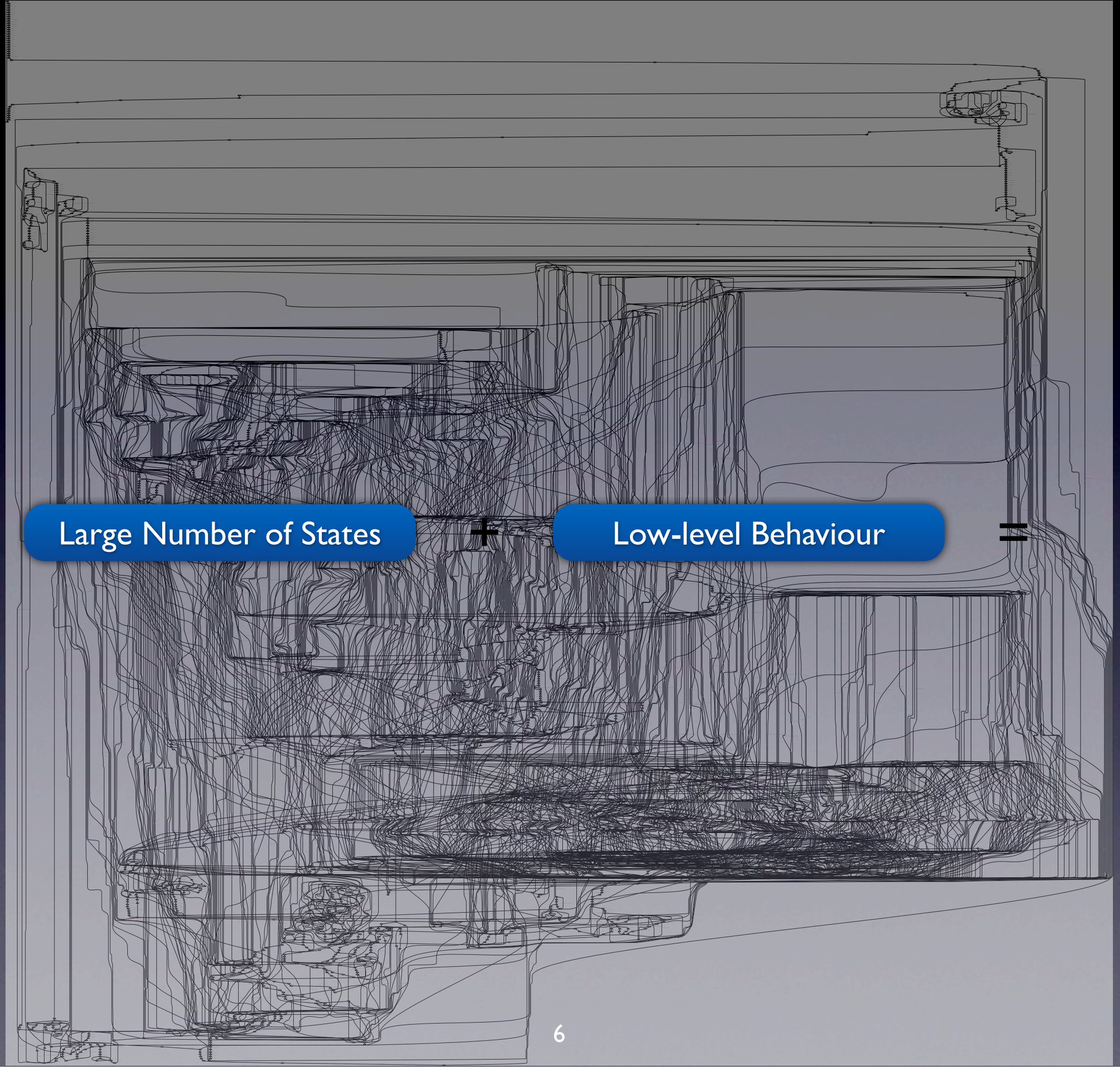


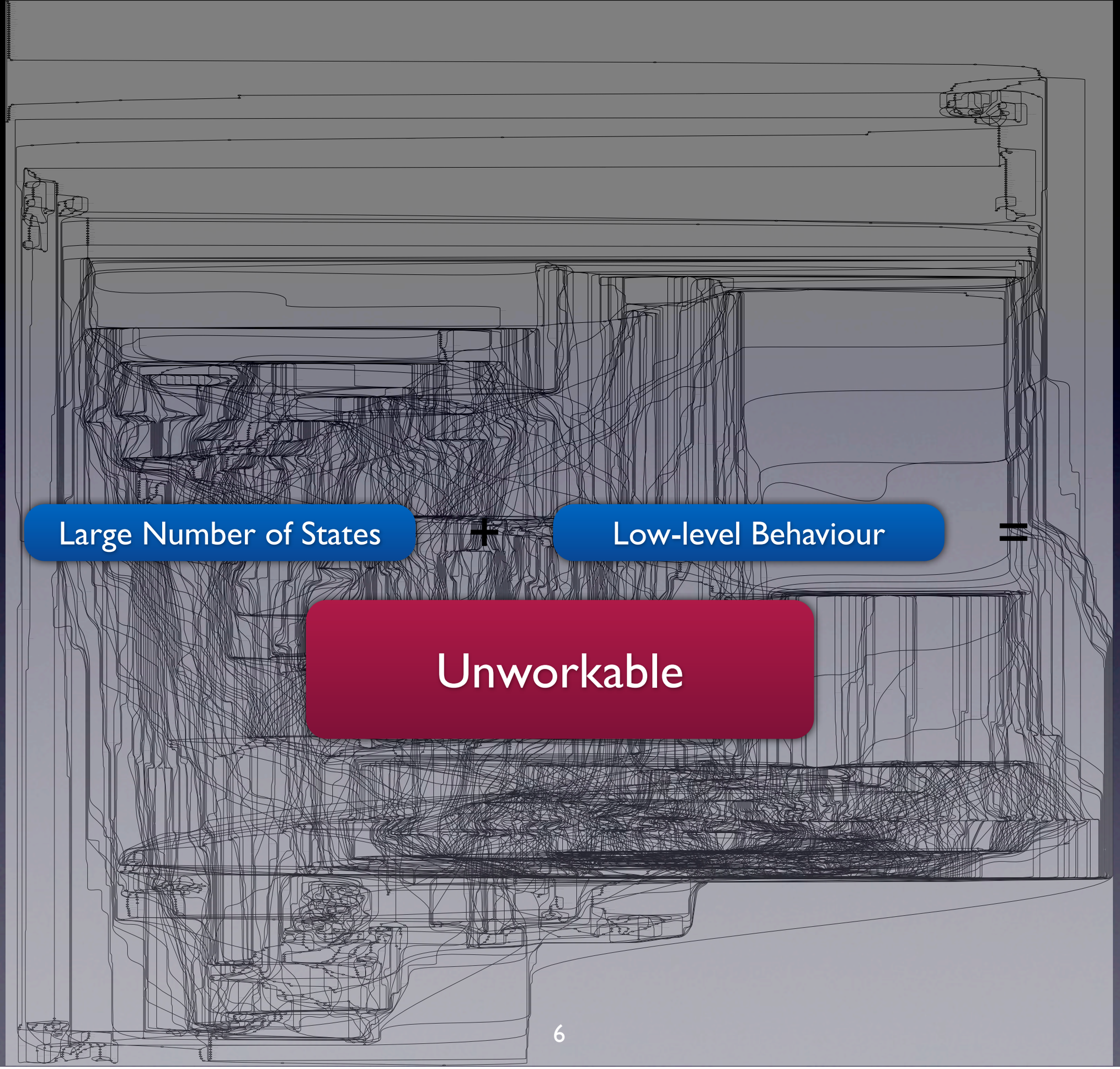
Friday, 16 September 11






Large Number of States


Large Number of States
+
Low-level Behaviour
=


Large Number of States
+
Low-level Behaviour
=
Unworkable

Sizing up the Problem

- Machine built from traces
- Traces are method calls
- Very low-level behaviour

Statecharts

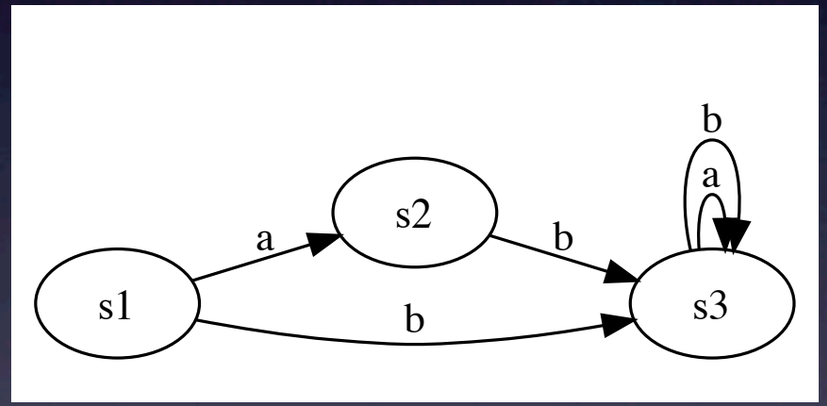
- Extension to state machines
- Composite and concurrent states
- Guards
- Events

Statecharts

- Extension to state machines
- Composite and concurrent states
- Guards
- Events

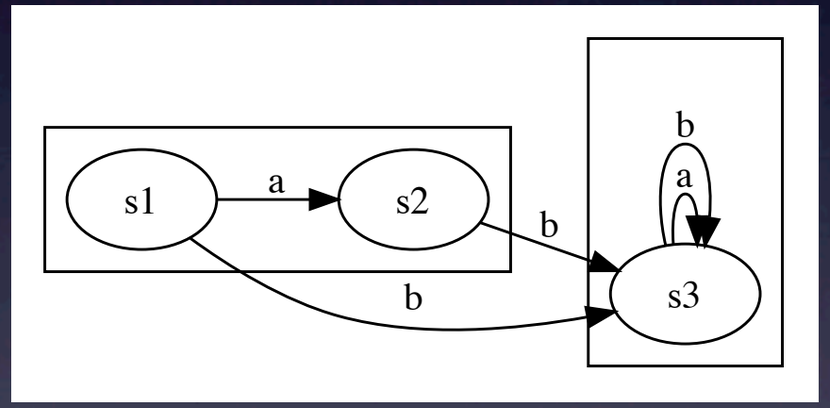
Hierarchical State Machines

- Superstate groups common behaviour



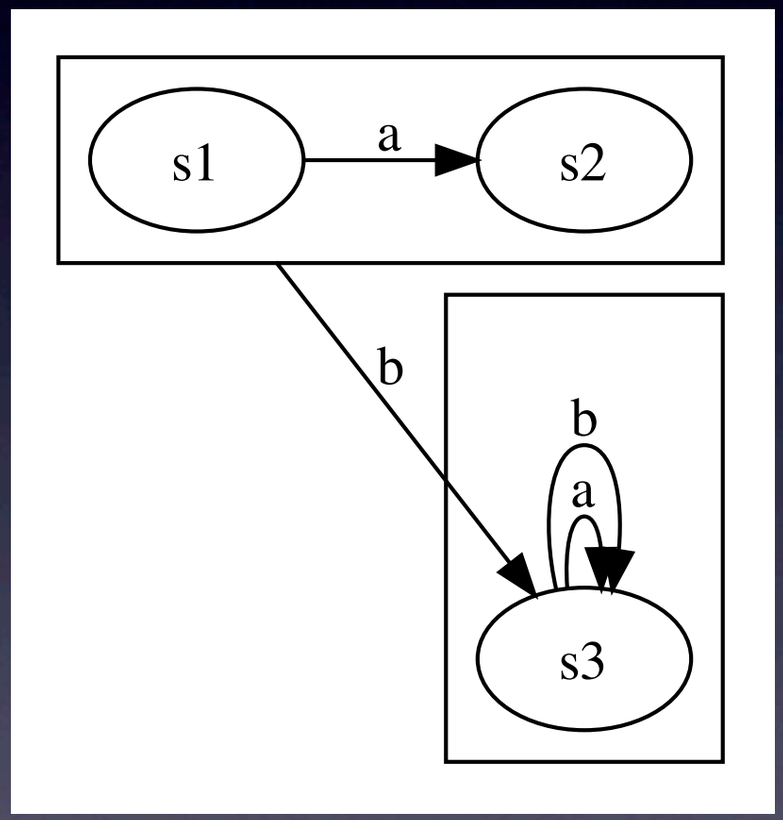
Hierarchical State Machines

- Superstate groups common behaviour



Hierarchical State Machines

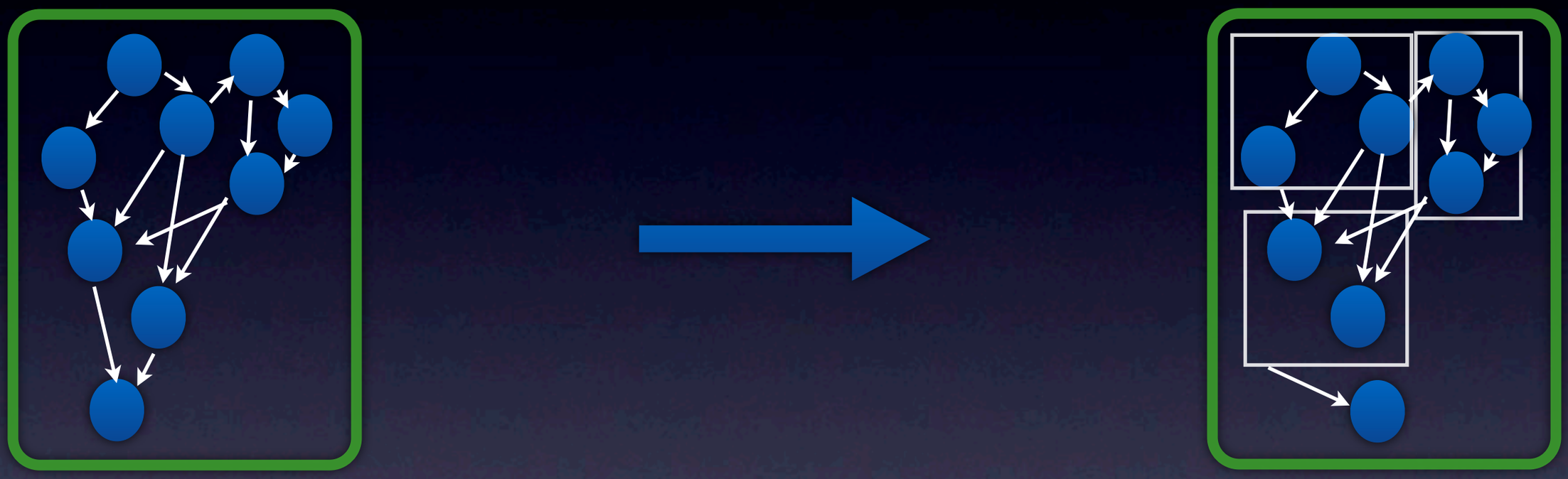
- Superstate groups common behaviour
- Transitions from all states in a superstate rewritten as one



FSM Size Problem

- HSM a potential solution
- Abstract low-level behaviour away
- Easier comprehension
- Allow variable level of detail

Goal



Existing Technique

- Rule based
- Identify groupings where edges could be reduced
- Kumar IACC09
- Need to check all pairs
- Rigid constraint

Related Technique

- Bunch - Mitchell & Mancoridis [TSE06]
- Search-based hierarchy generation for module dependency graphs
- Key contribution: MQ
- Lower number of edges between superstates is better

Related Technique

$$MQ = \sum_{i=1}^{k} CF_i$$

$$CF_i = \begin{cases} 0 & \mu_i = 0 \\ \dfrac{2\mu_i}{2\mu_i + \sum\limits_{\substack{j=1 \\ j \neq i}}^{k} (\epsilon_{i,j} + \epsilon_{j,i})} & \text{otherwise} \end{cases}$$

- μ: Number of edges within a module
- ε: Number of edges that leave a module

FSM to HSM

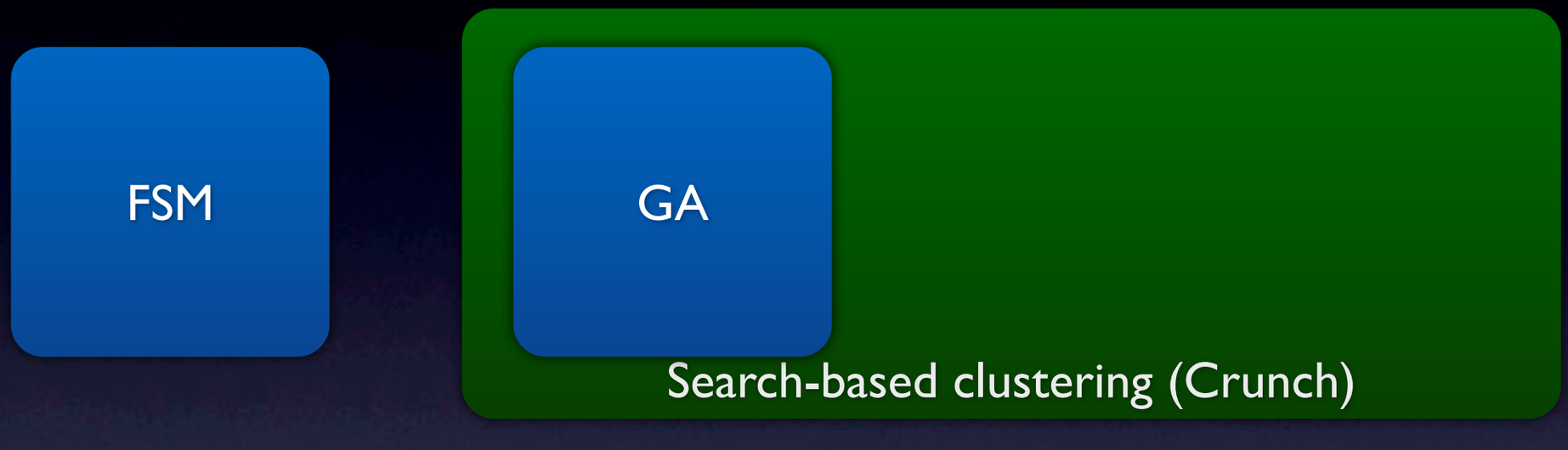
FSM to HSM

FSM to HSM

FSM

Search-based clustering (Crunch)

FSM to HSM



FSM to HSM

FSM to HSM


FSM
GA
Fitness Function
Individual: Hierarchy
Search-based clustering (Crunch)

FSM to HSM

FSM

GA

Fitness Function

Individual: Hierarchy

Search-based clustering (Crunch)

HSM

Output

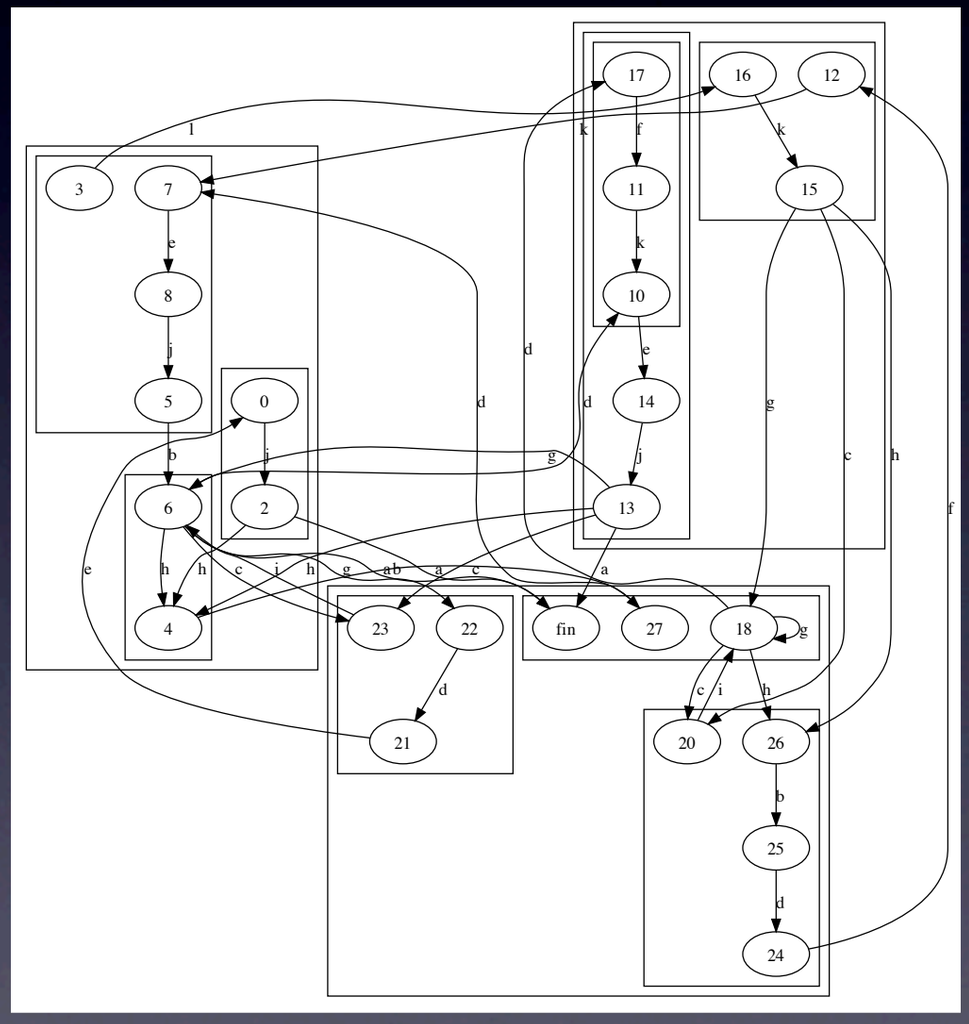
How can you tell how good a solution is?

Evaluating an HSM

- Existing metric for FSMs is CC
- E - N + 2
- Ignores hierarchy
- Need a better alternative...

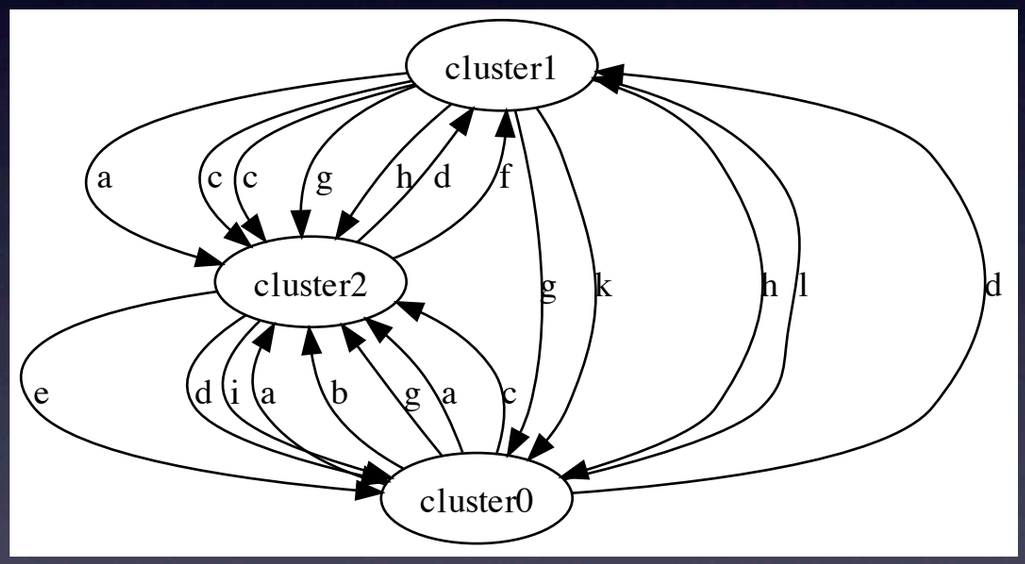
TLCC

- Treat each superstate as a simple state
- Ignore all edges within superstates
- Calculate CC

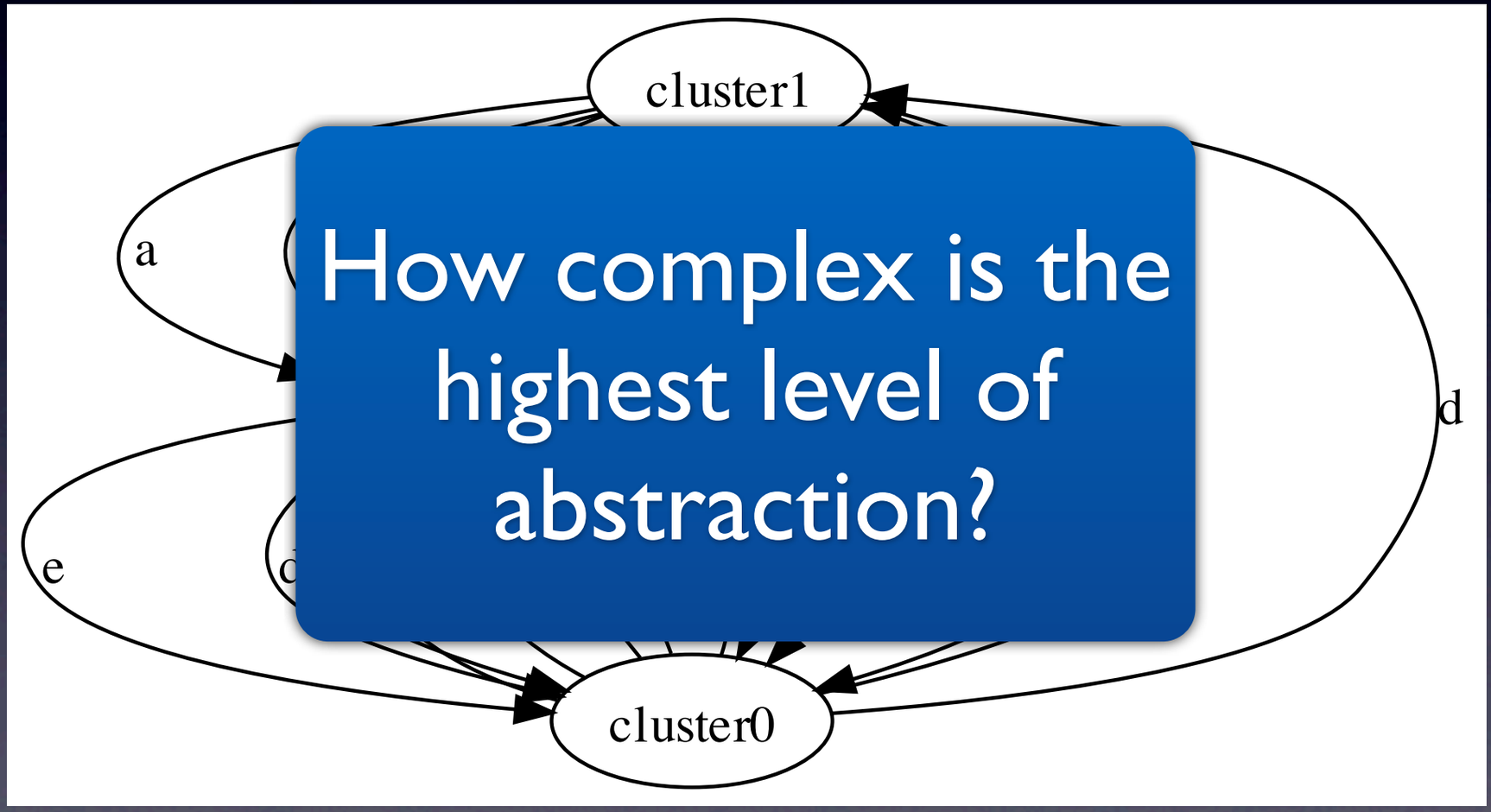


TLCC

- Treat each superstate as a simple state
- Ignore all edges within superstates
- Calculate CC



TLCC = 20 - 3 + 2 = 19



How complex is the highest level of abstraction?

Problems with TLCC

- Doesn't measure all the hierarchy
- Invalid solutions can get high results
  - When top level contains all states TLCC = 1

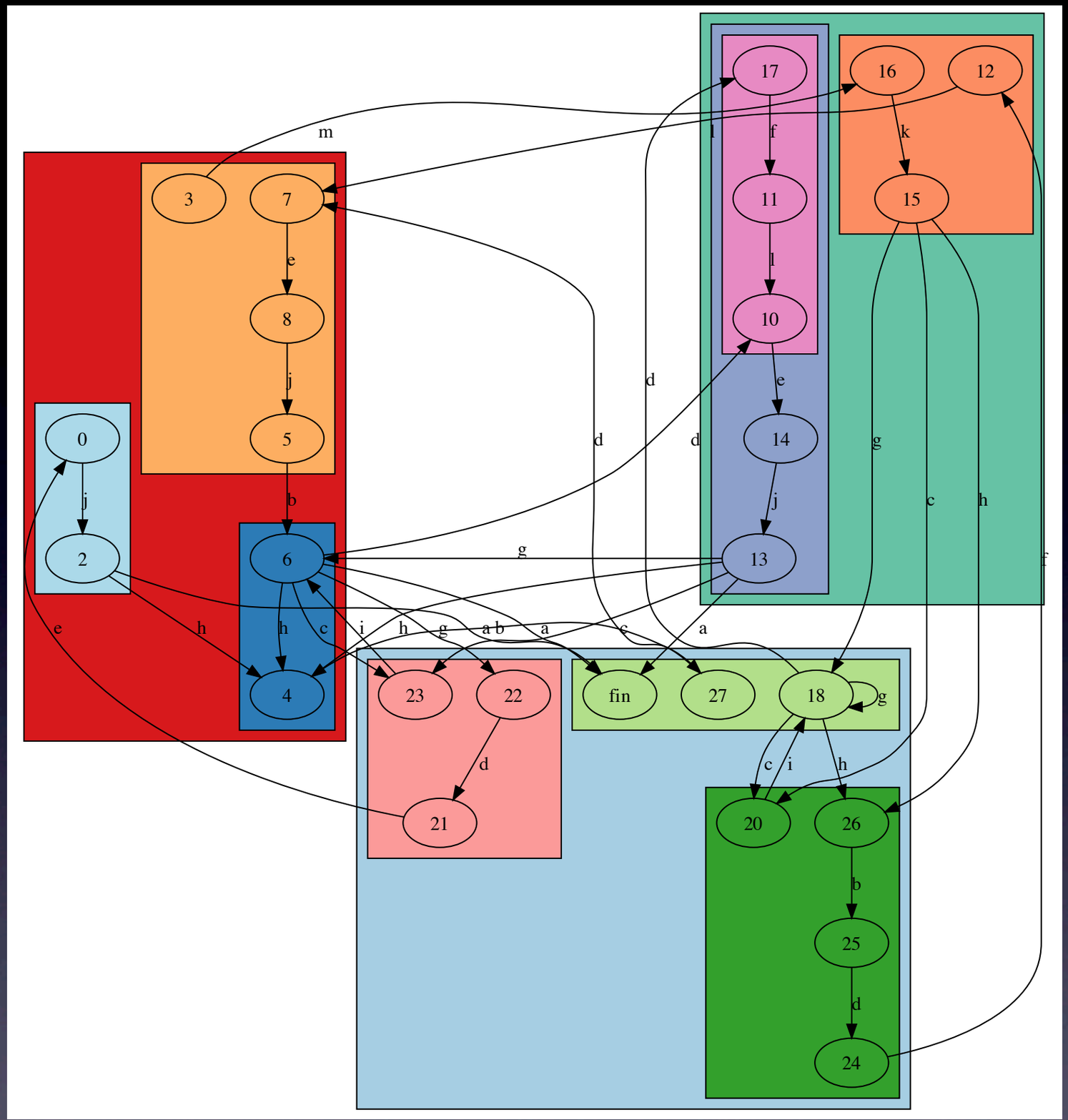
HCC

$$HCC(S,T,d) = \frac{TLCC(H,T)}{d} + \sum_{c \in H} HCC(c,T,d+1)$$

- Favours solutions that place lower level behaviour further down the hierarchy

HCC

2-4+2
1-2+2
+ 1-2+2)
/3



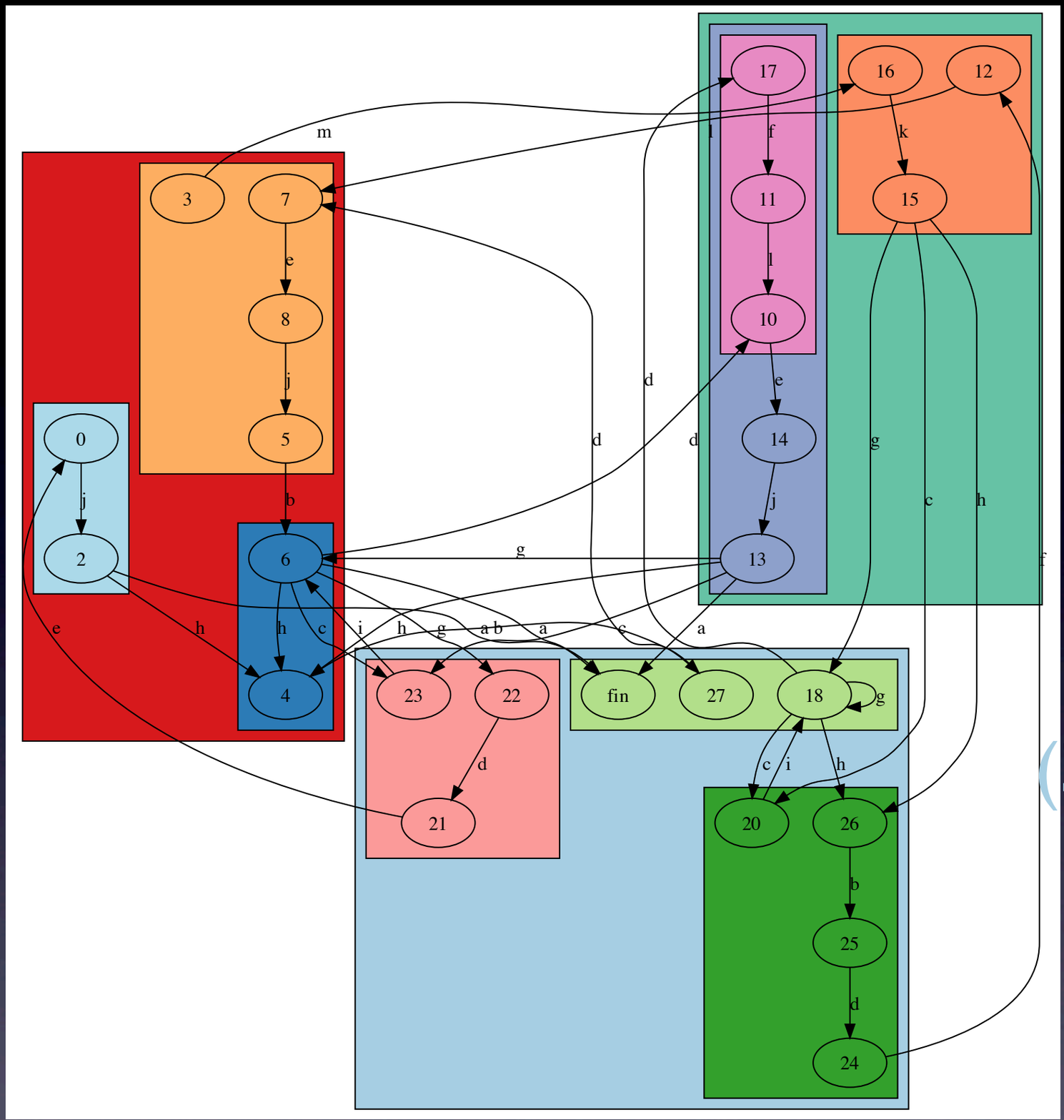
(1-3+2
+2-3+2)/3
+2-3+2/4

(2-4+2
+ 1-3+2
+1-3+2)
/3

HCC

2-4+2
1-2+2
+ 1-2+2)
/3

(2-3+2) / 2



HCC

(1-3+2
+2-3+2)/3
+2-3+2/4

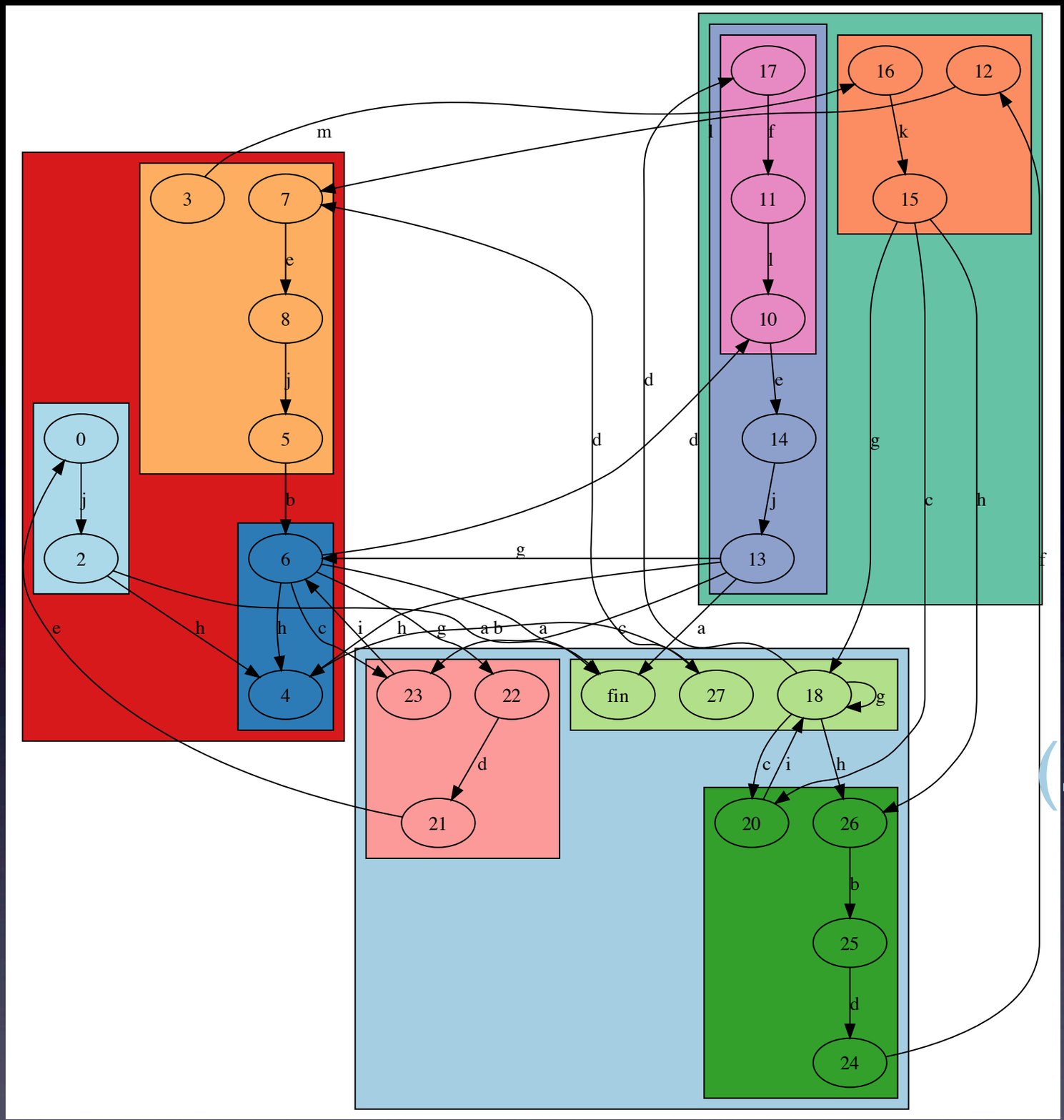
(0-2+2)/2

(3-3+2)/2

(2-4+2
+ 1-3+2
+1-3+2)
/3

2-4+2
1-2+2
+ 1-2+2)
/3

(2-3+2) / 2



(1-3+2
+2-3+2)/3
+2-3+2/4

(0-2+2)/2

(3-3+2)/2

(2-4+2
+ 1-3+2
+1-3+2)
/3

HCC

19 + (1+0+2)/2 + (2+2+0)/3 + 1/4

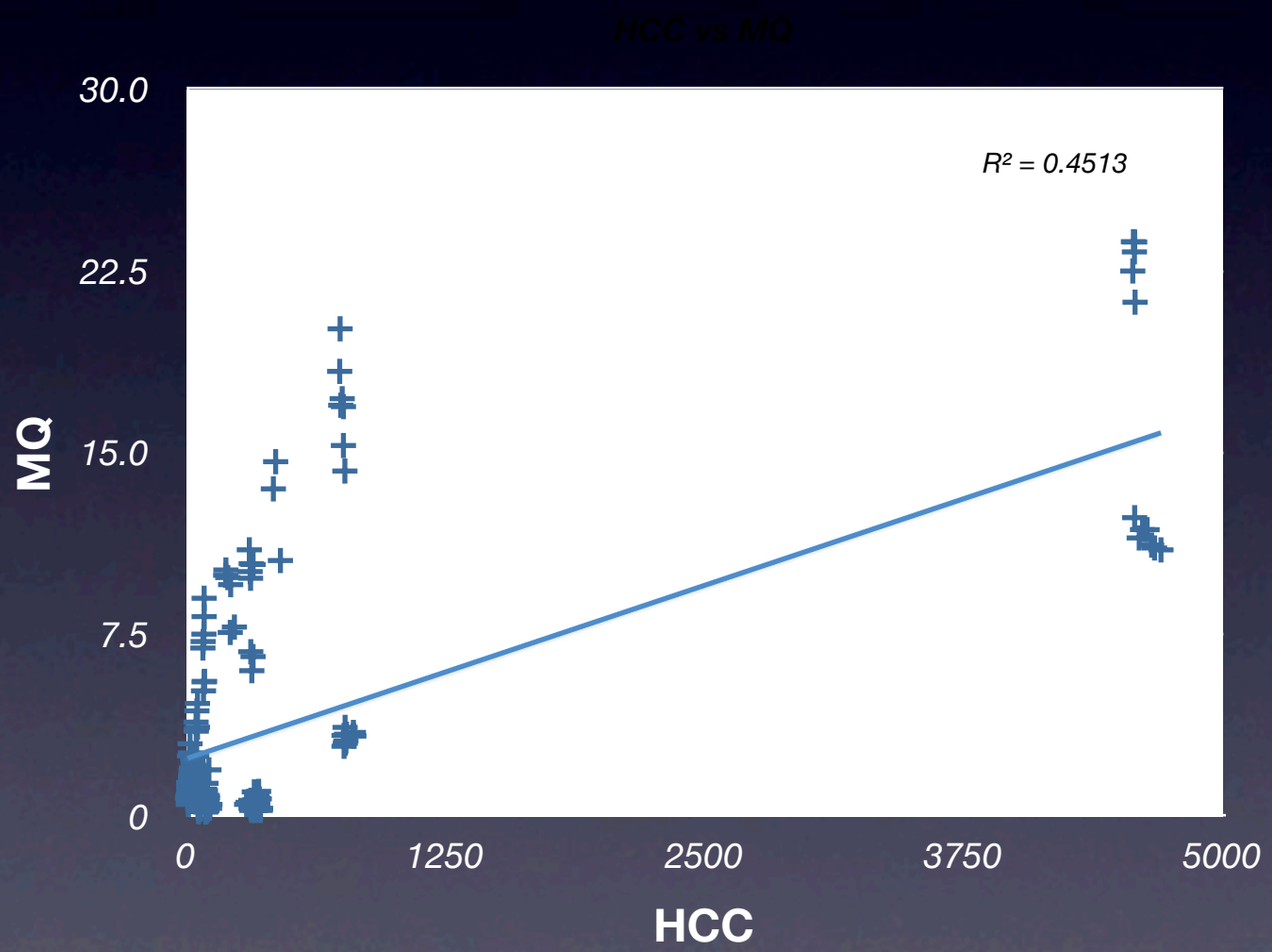
Problems with HCC

- Easy to cheat
- Still based on CC
- Subject to same criticisms

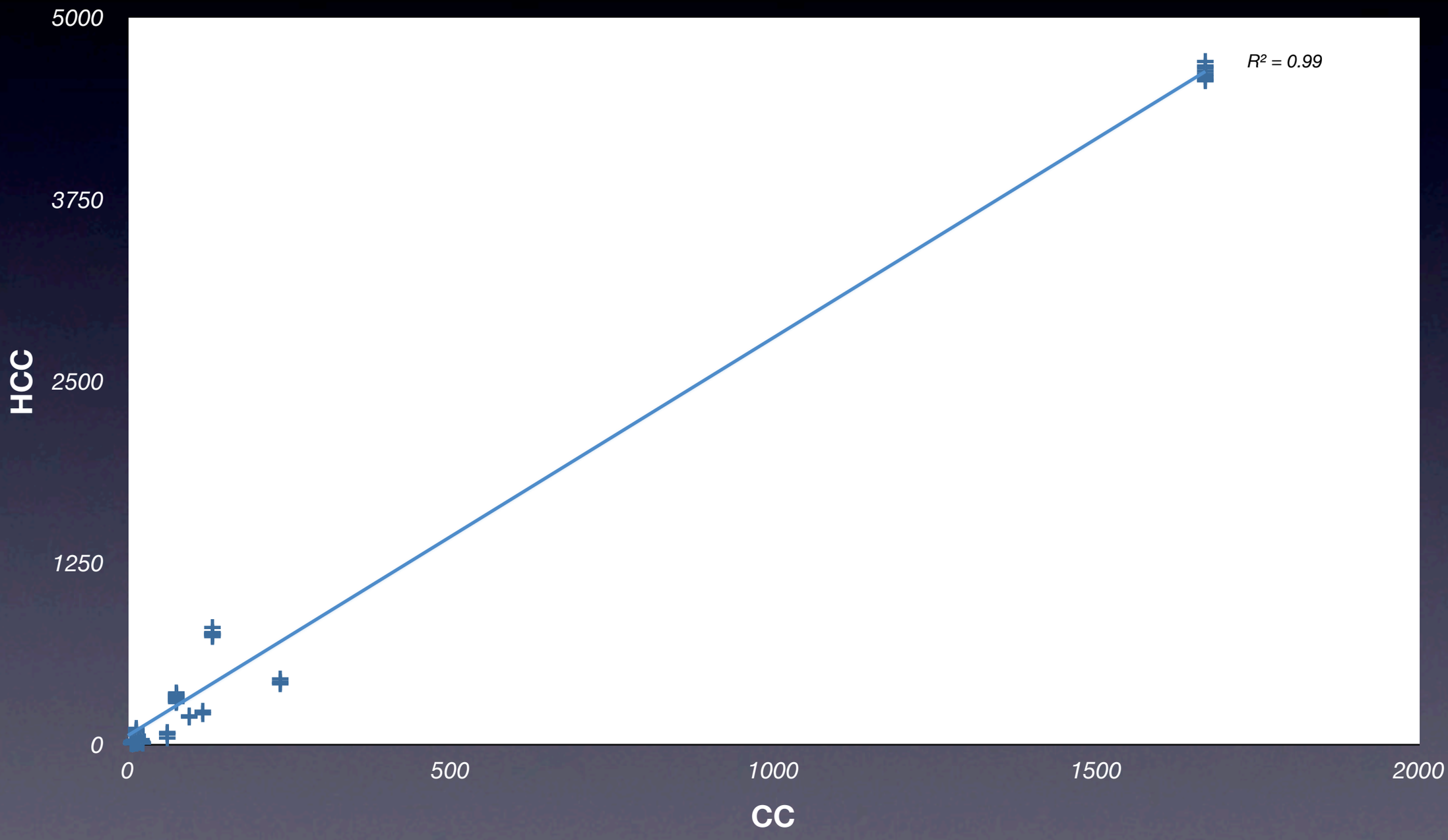
HCC vs MQ

HCC vs MQ

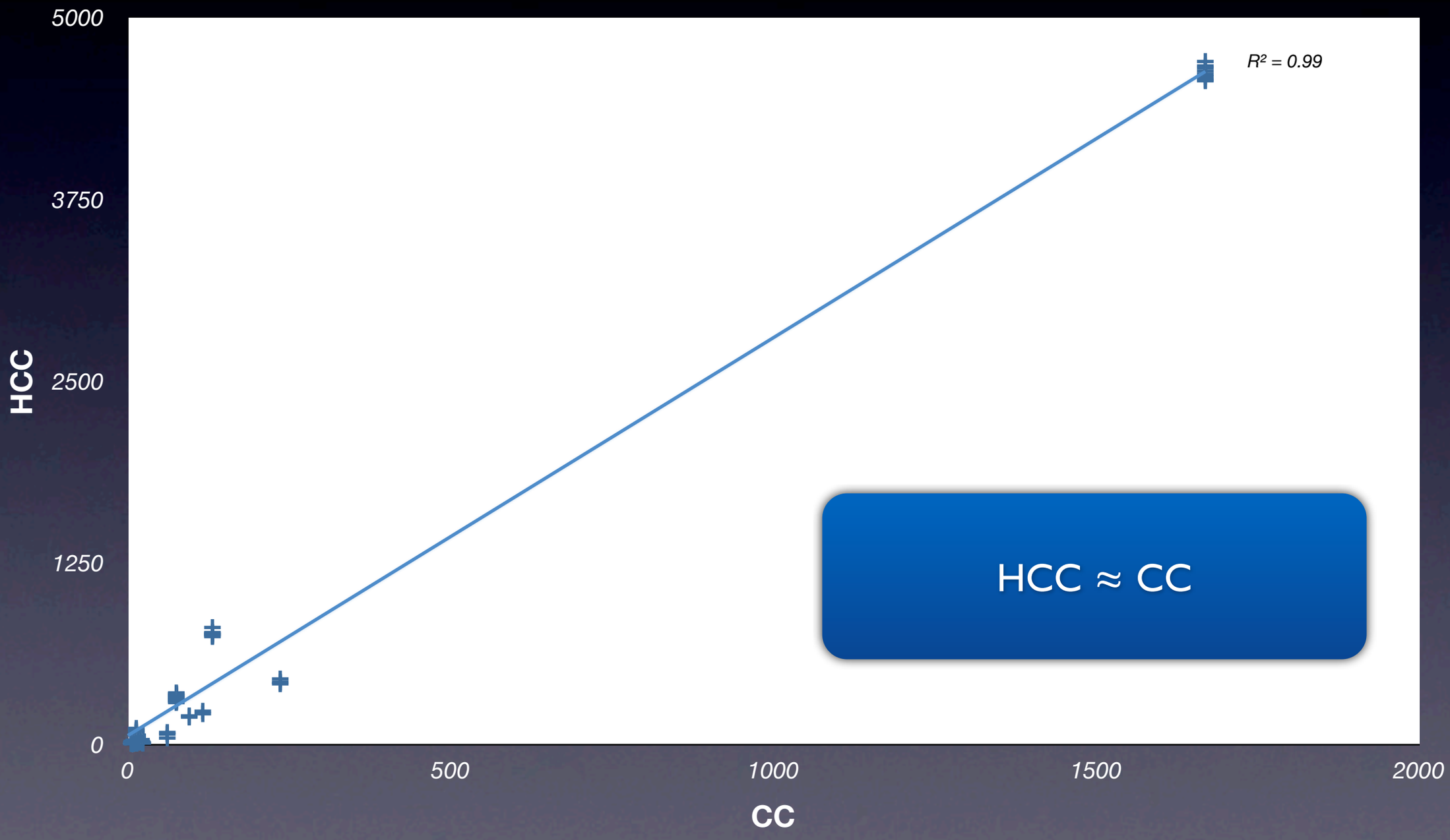
- 27 FSMs
- size between 5 and 1400 states
- results from 5 different fitness functions
- 202 points total



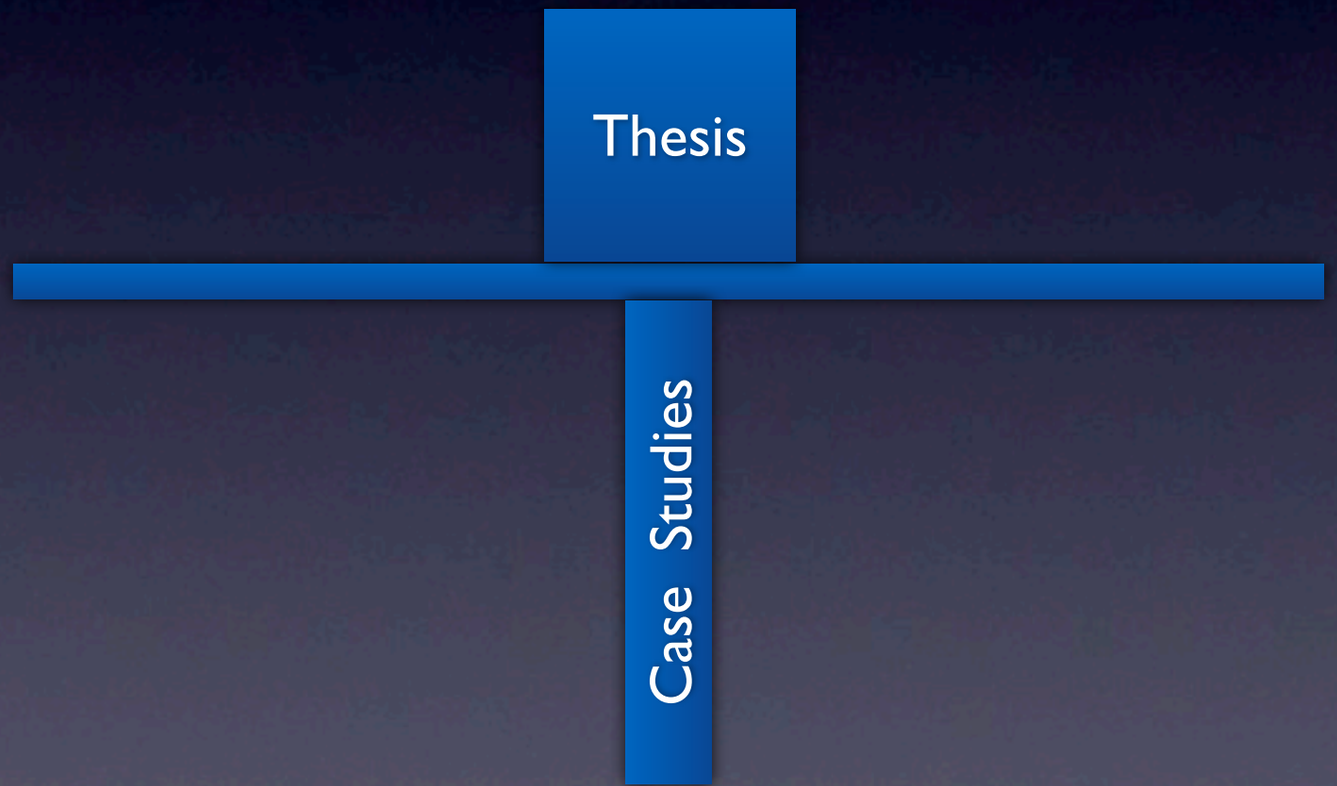
HCC vs CC



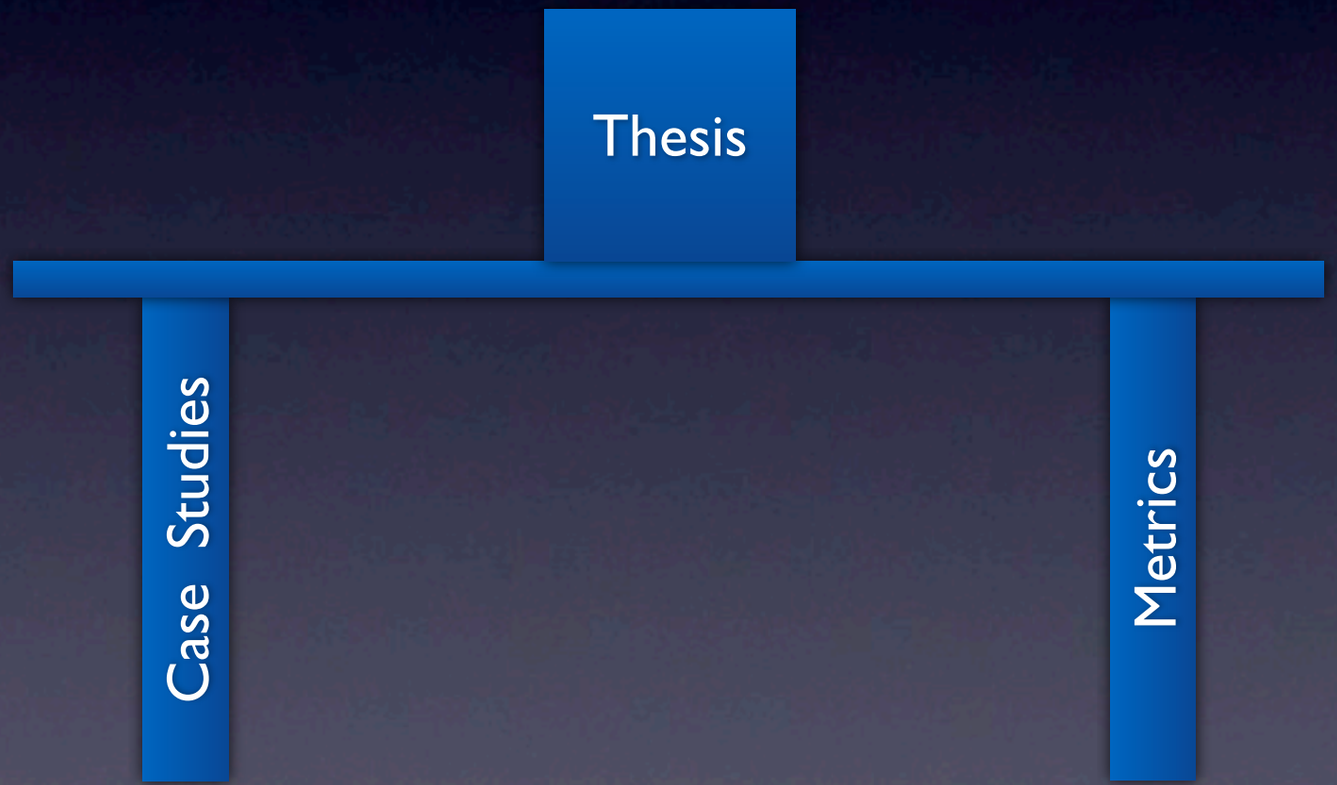
HCC vs CC



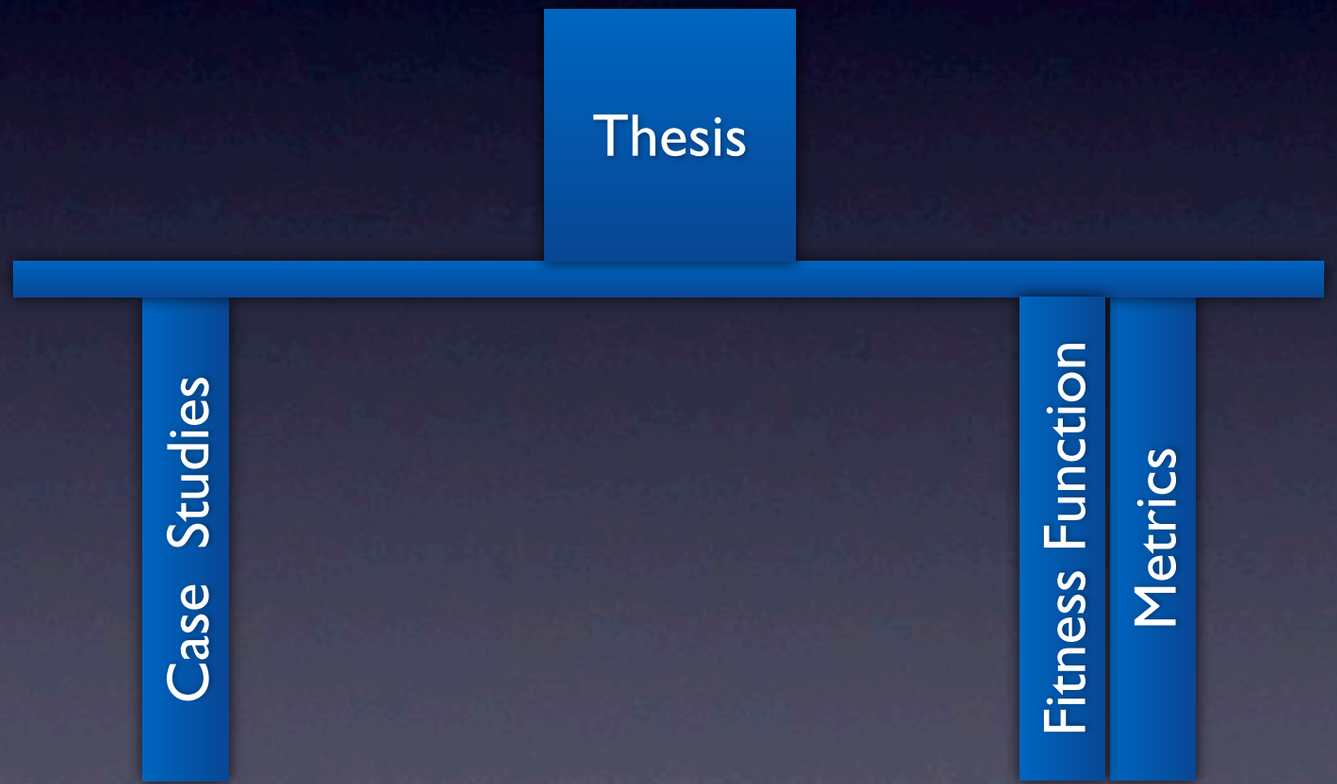
PhD



PhD



PhD



Metrics

- Looking into SE metrics to adapt to HSMs
- Many depend on unavailable information
- Can also be used as a fitness function

Metrics

- Looking into SE metrics to adapt to HSMs
- Many depend on unavailable information
- Can also be used as a fitness function

Current Work

- Look… HSMs
- Many … tion
- Can also be used as a fitness function

Need to find or develop an appropriate metric

Considerations

- How big can a superstate be before it's considered too large?
- Is there enough information in the machine structure?
- Is it enough to maximise some metric?
- Multiobjective?

PhD: Intended Contributions

- Crunch: search-based graph clustering tool
- Survey and evaluation of metrics
- *Superstate Name Generation*
- *Human study*

Conclusion

- Reverse engineered state machines are too large
- It's possible to reduce CC through hierarchy introduction
- Introduced a way to compare CC of HSMs
- ...but HCC still subject to same problems as CC