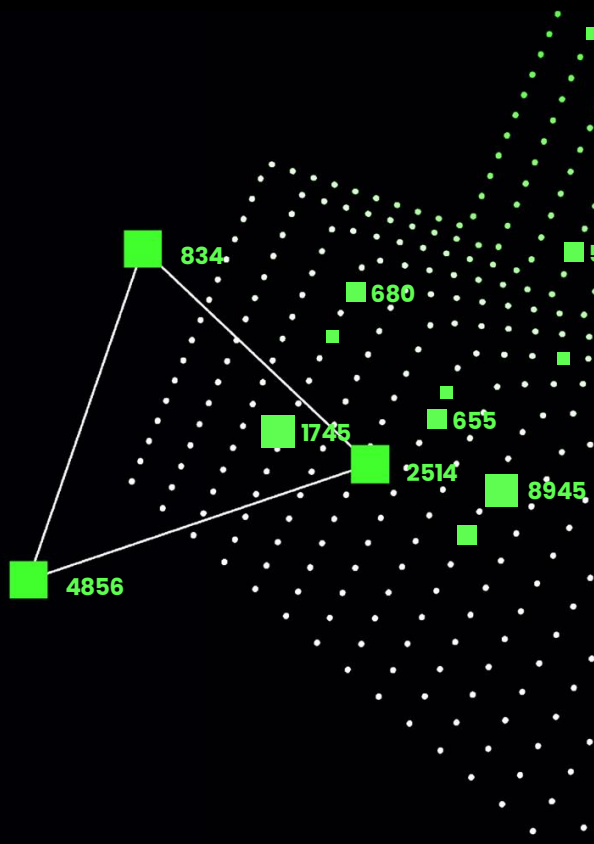
celonis

Chair of Process and Data Science | RWTH AACHEN UNIVERSITY

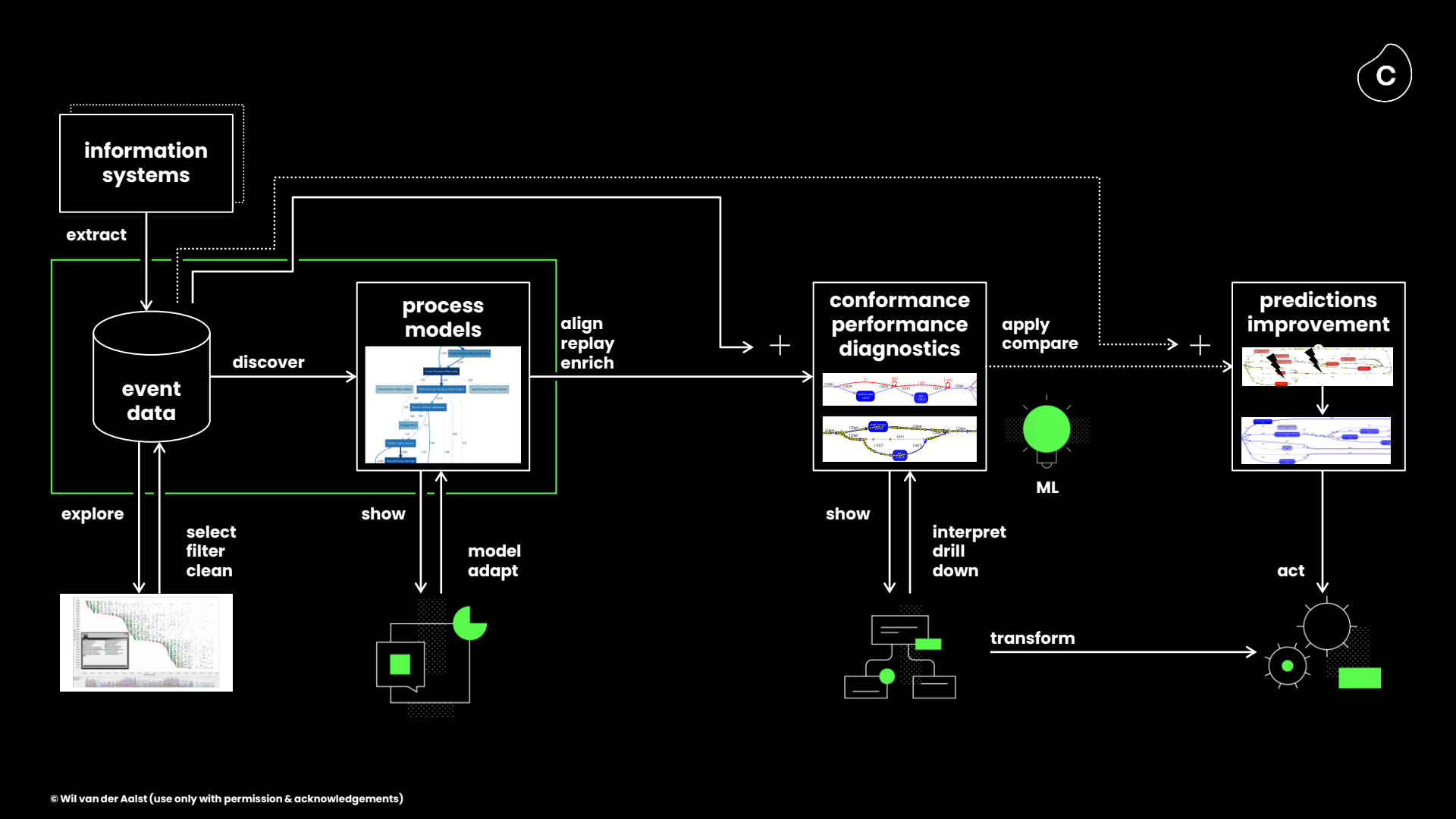
Process Discovery 2/2

Beyond Directly-Follows Graphs

Process mining: From Theory to Execution



prof.dr.ir. Wil van der Aalst
www.vdaalst.com @wvdaalst | www.pads.rwth-aachen.de

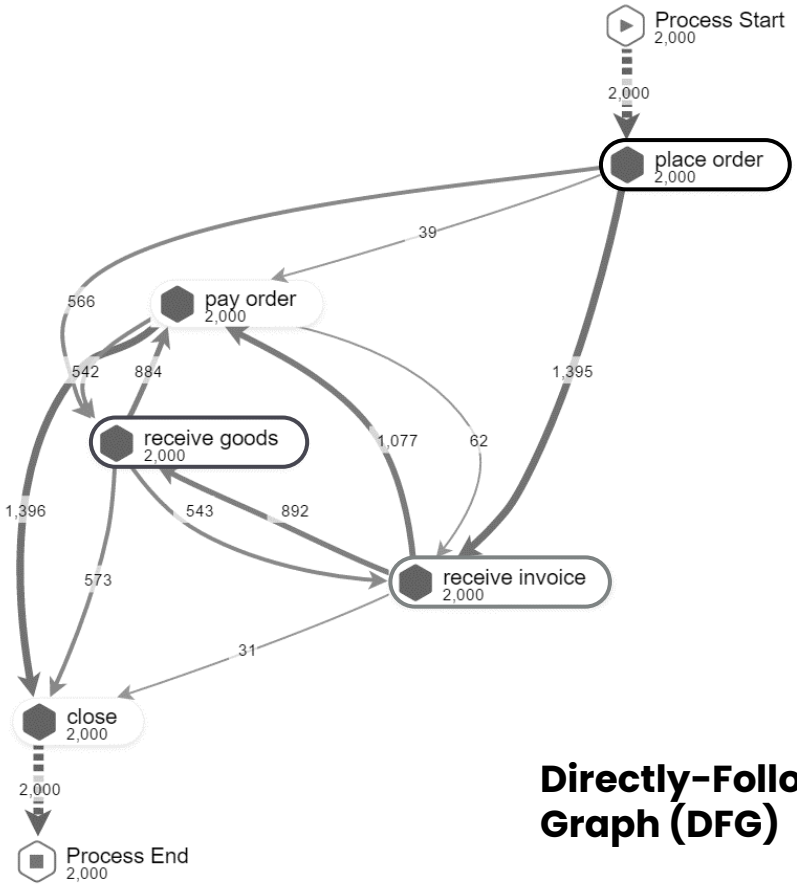


Recap Last Lecture

857 cases
42.85% of the log

DFGs cannot capture concurrency

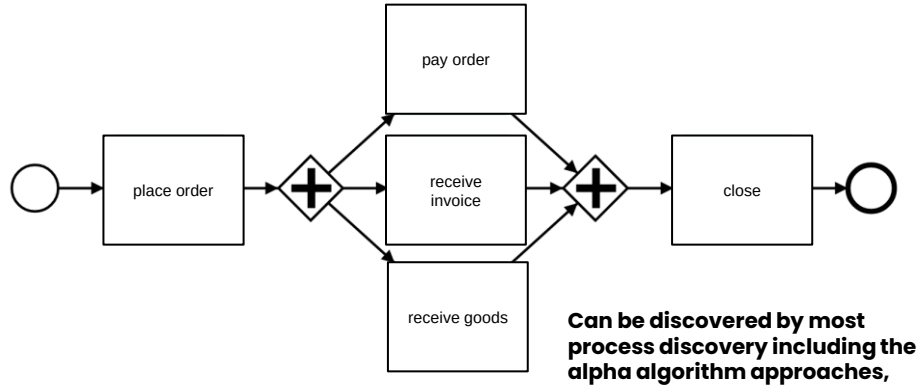
857 traces 42.85%	place order	receive invoice	receive goods	pay order	close
539 traces 26.95%	place order	receive goods	receive invoice	pay order	close
538 traces 26.90%	place order	receive invoice	pay order	receive goods	close
35 traces 1.75%	place order	pay order	receive invoice	receive goods	close
27 traces 1.35%	place order	receive goods	pay order	receive invoice	close
4 traces 0.20%	place order	pay order	receive goods	receive invoice	close



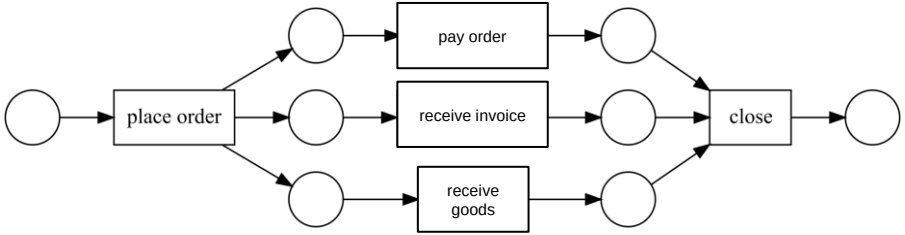
Directly-Follows Graph (DFG)

But more advanced techniques can!

857 traces 42.85%	place order	receive invoice	receive goods	pay order	close
539 traces 26.95%	place order	receive goods	receive invoice	pay order	close
538 traces 26.90%	place order	receive invoice	pay order	receive goods	close
35 traces 1.75%	place order	pay order	receive invoice	receive goods	close
27 traces 1.35%	place order	receive goods	pay order	receive invoice	close
4 traces 0.20%	place order	pay order	receive goods	receive invoice	close



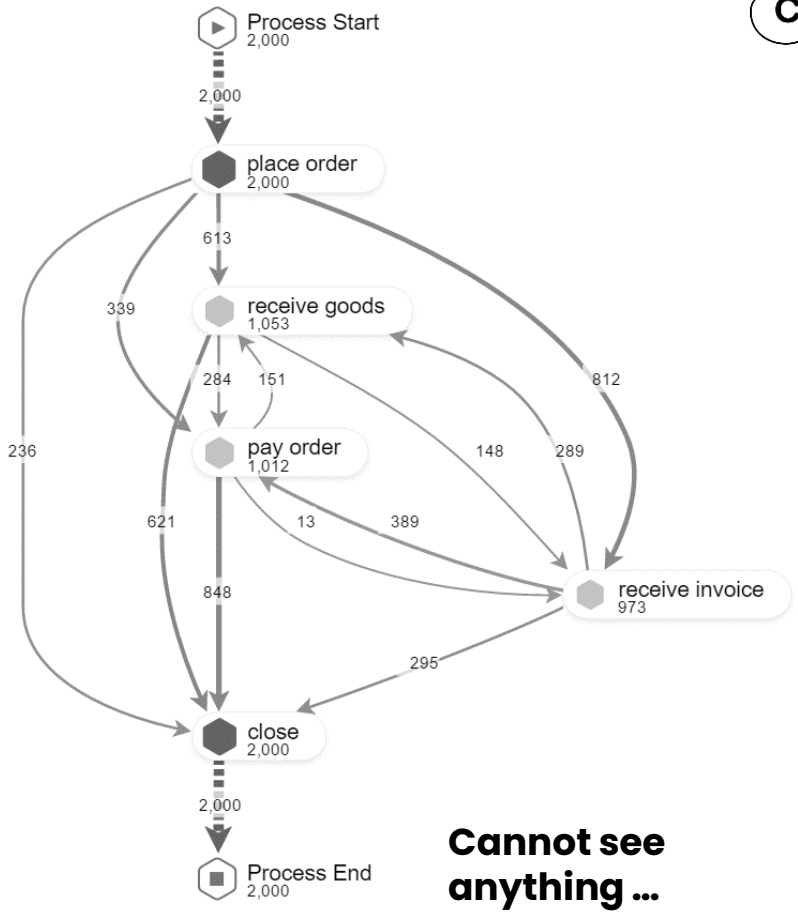
Can be discovered by most process discovery including the alpha algorithm approaches, region-based techniques, etc.



Making the problem a bit more challenging ...

285 traces 14.25%	place order	receive goods	close		
251 traces 12.55%	place order	pay order	close		
238 traces 11.90%	place order	receive invoice	pay order	close	
236 traces 11.80%	place order	close			
216 traces 10.80%	place order	receive goods	close		
183 traces 9.15%	place order	receive invoice	receive goods	close	
177 traces 8.85%	place order	receive goods	pay order	close	
103 traces 5.15%	place order	receive invoice	receive goods	pay order	close
79 traces 3.95%	place order	receive goods	receive invoice	pay order	close
78 traces 3.90%	place order	pay order	receive goods	close	
72 traces 3.60%	place order	receive invoice	pay order	receive goods	close
68 traces 3.40%	place order	receive goods	receive invoice	close	
6 traces 0.30%	place order	pay order	receive invoice	close	
4 traces 0.20%	place order	receive goods	pay order	receive invoice	close
3 traces 0.15%	place order	pay order	receive invoice	receive goods	close
1 trace 0.05%	place order	pay order	receive goods	receive invoice	close

The three middle activities are now optional ...



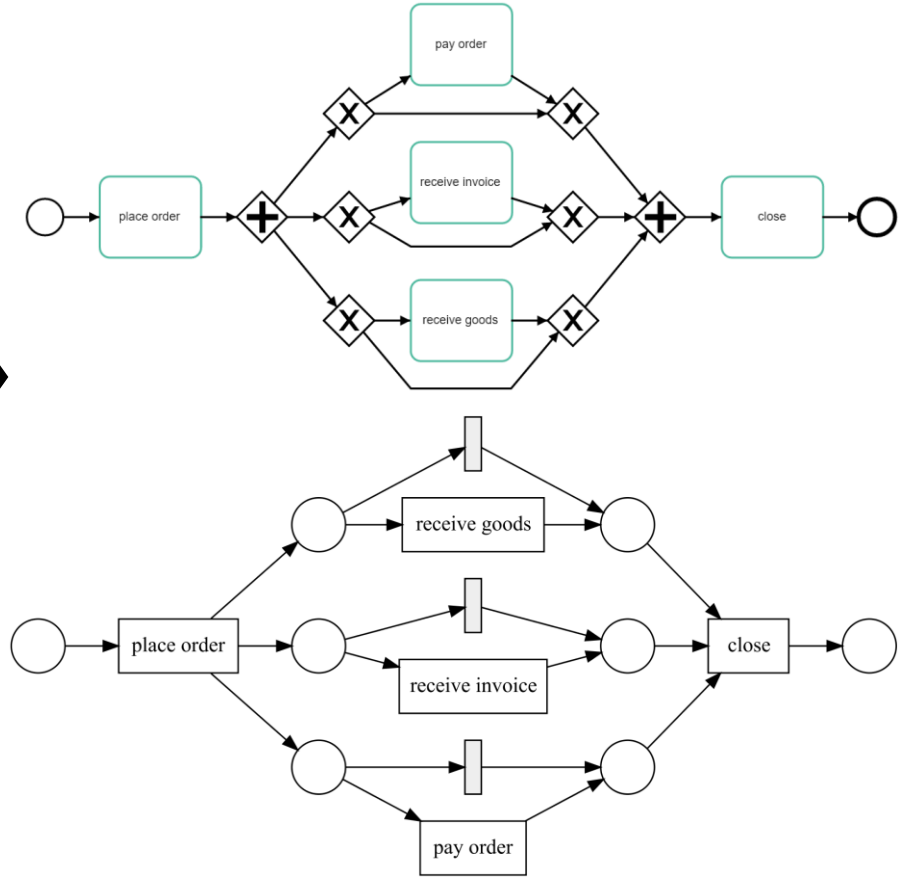
Cannot see anything ...

Making the problem a bit more challenging …

285 traces 14.25%	place order	receive goods	close	
251 traces 12.55%	place order	pay order	close	
238 traces 11.90%	place order	receive invoice	pay order	close
236 traces 11.80%	place order	close		
216 traces 10.80%	place order	receive goods	close	
183 traces 9.15%	place order	receive invoice	receive goods	close
177 traces 8.85%	place order	receive goods	pay order	close
103 traces 5.15%	place order	receive invoice	receive goods	pay order	close
79 traces 3.95%	place order	receive goods	receive invoice	pay order	close
78 traces 3.90%	place order	pay order	receive goods	close
72 traces 3.60%	place order	receive invoice	pay order	receive goods	close
68 traces 3.40%	place order	receive goods	receive invoice	close
6 traces 0.30%	place order	pay order	receive invoice	close
4 traces 0.20%	place order	receive goods	pay order	receive invoice	close
3 traces 0.15%	place order	pay order	receive invoice	receive goods	close
1 trace 0.05%	place order	pay order	receive goods	receive invoice	close

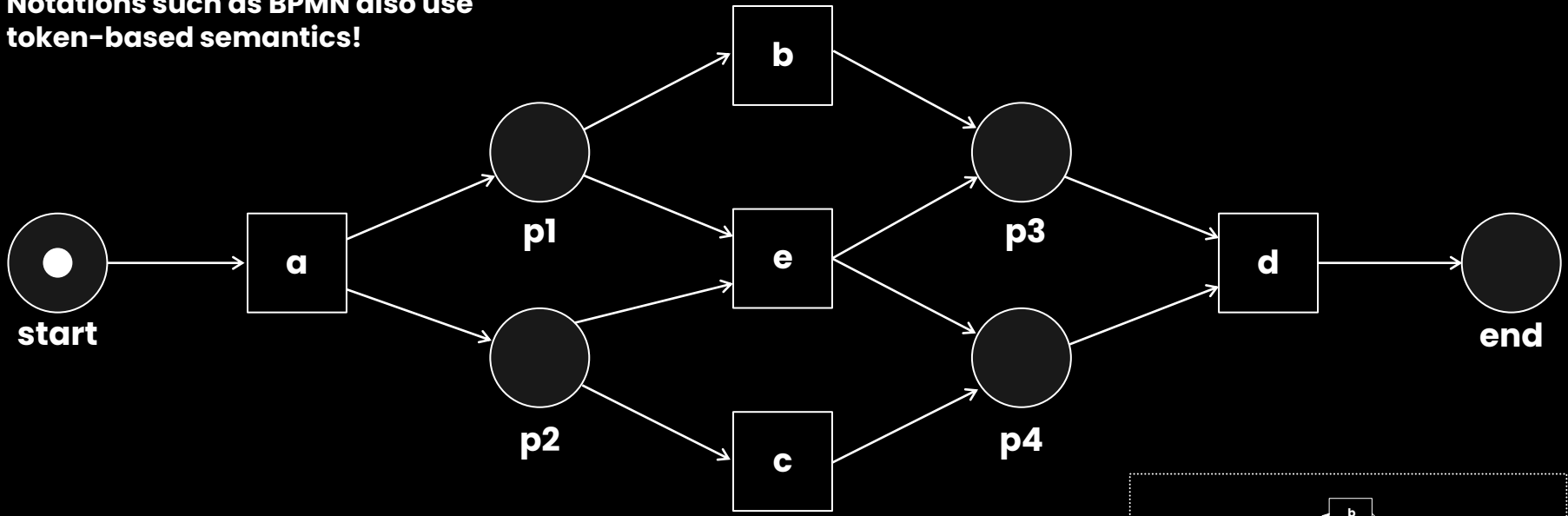
The three middle activities are now optional …

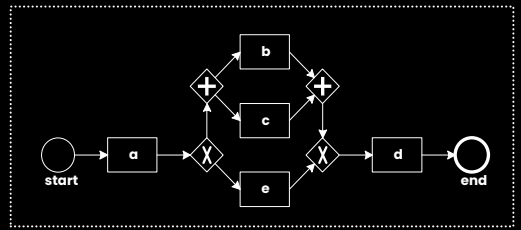
Model discovered by the inductive mining technique first implemented in ProM and later added to Celonis.



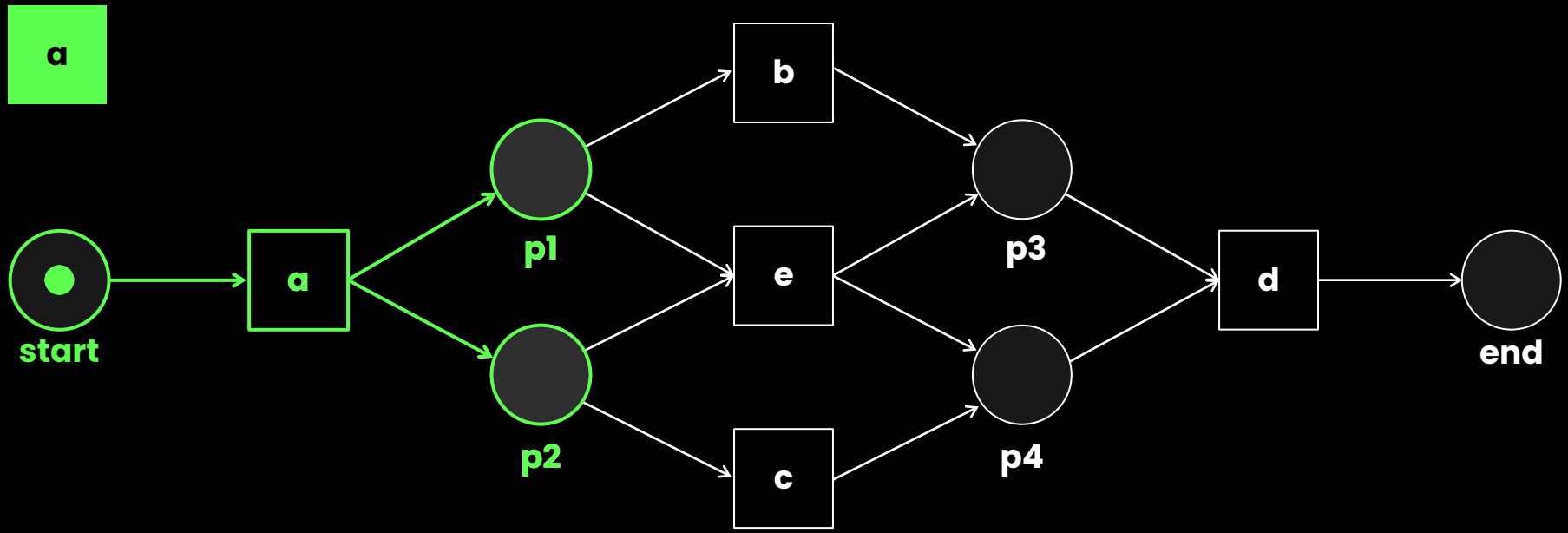
Petri nets: The basics

Notations such as BPMN also use token-based semantics!

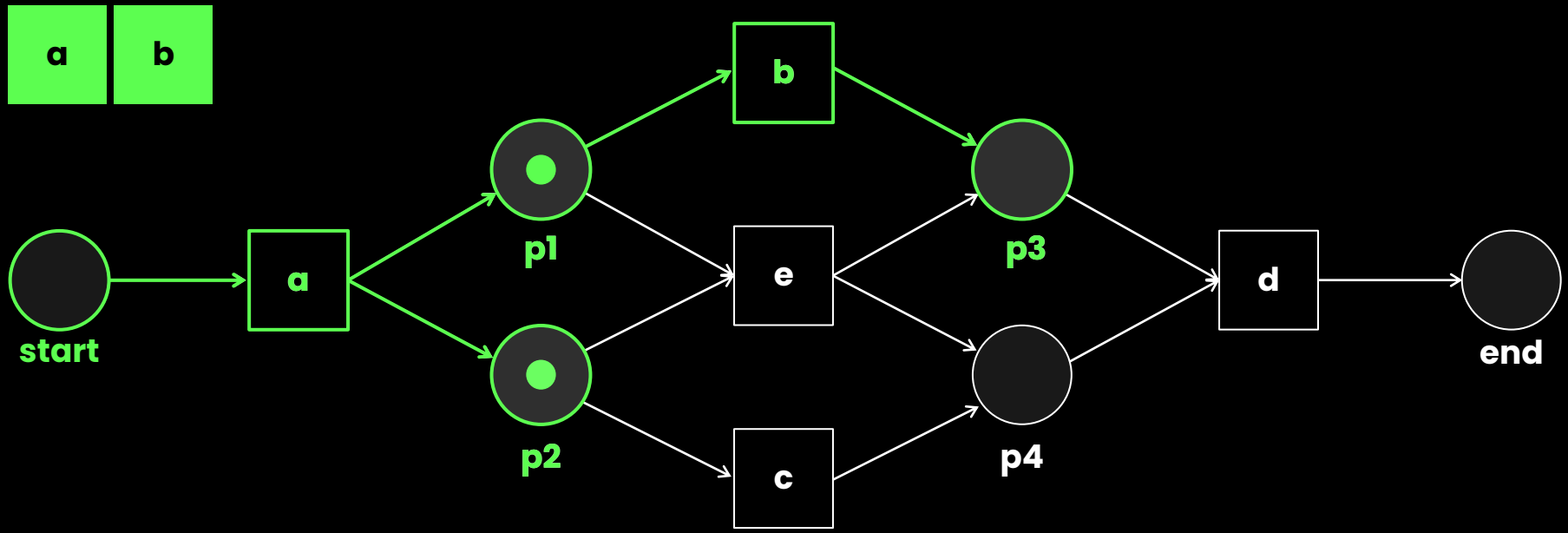




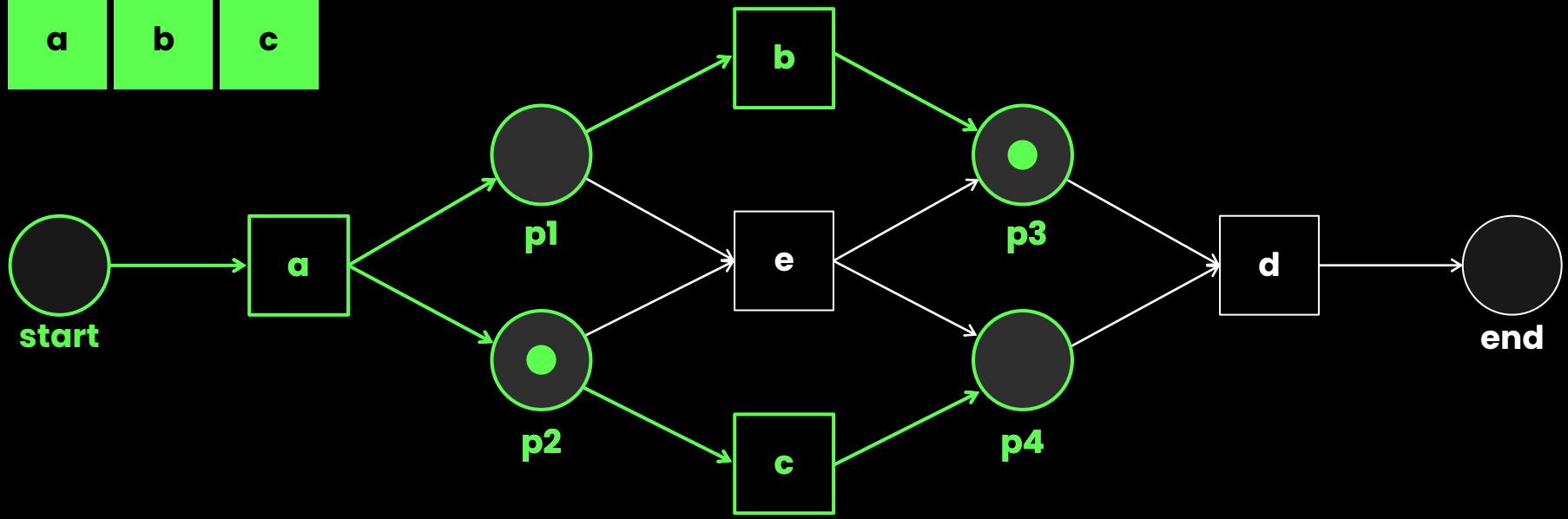
Petri nets: The basics



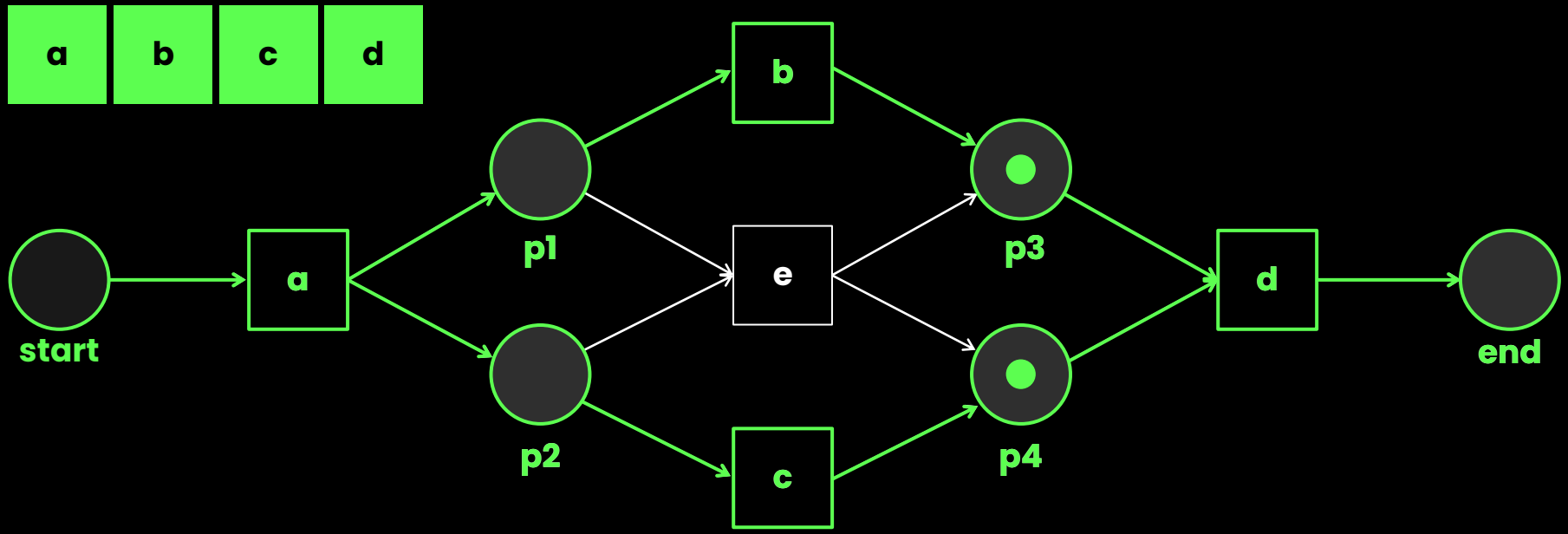
Petri nets: The basics



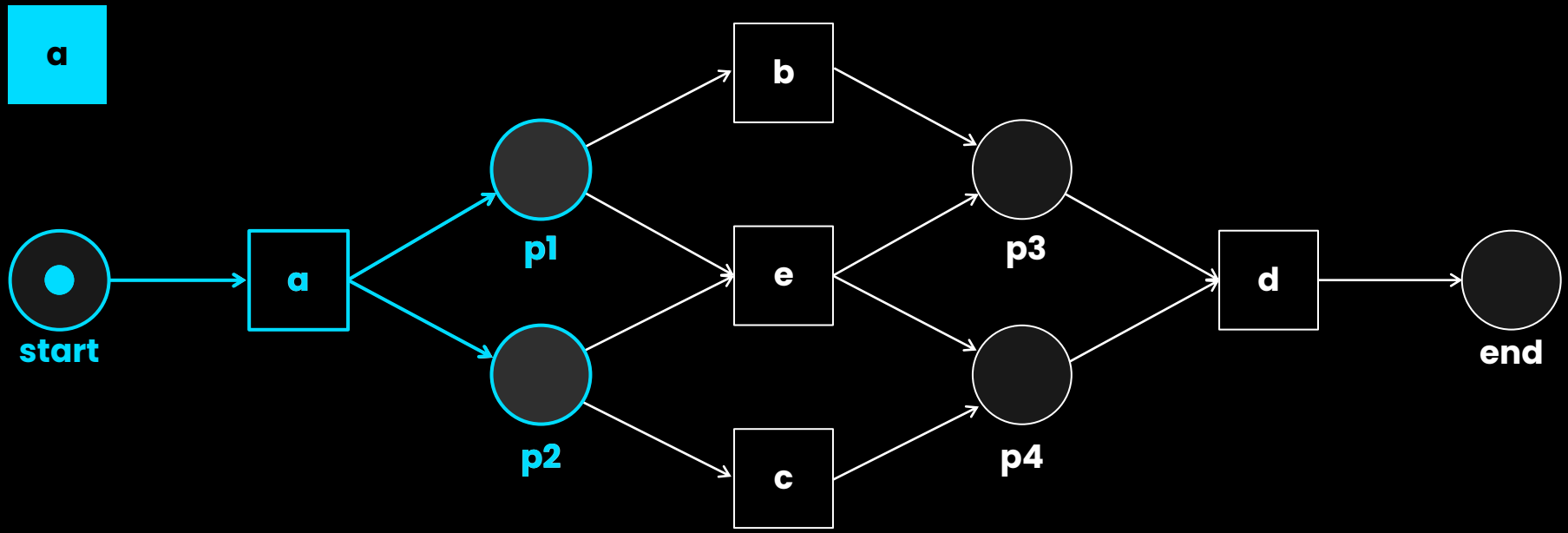
Petri nets: The basics



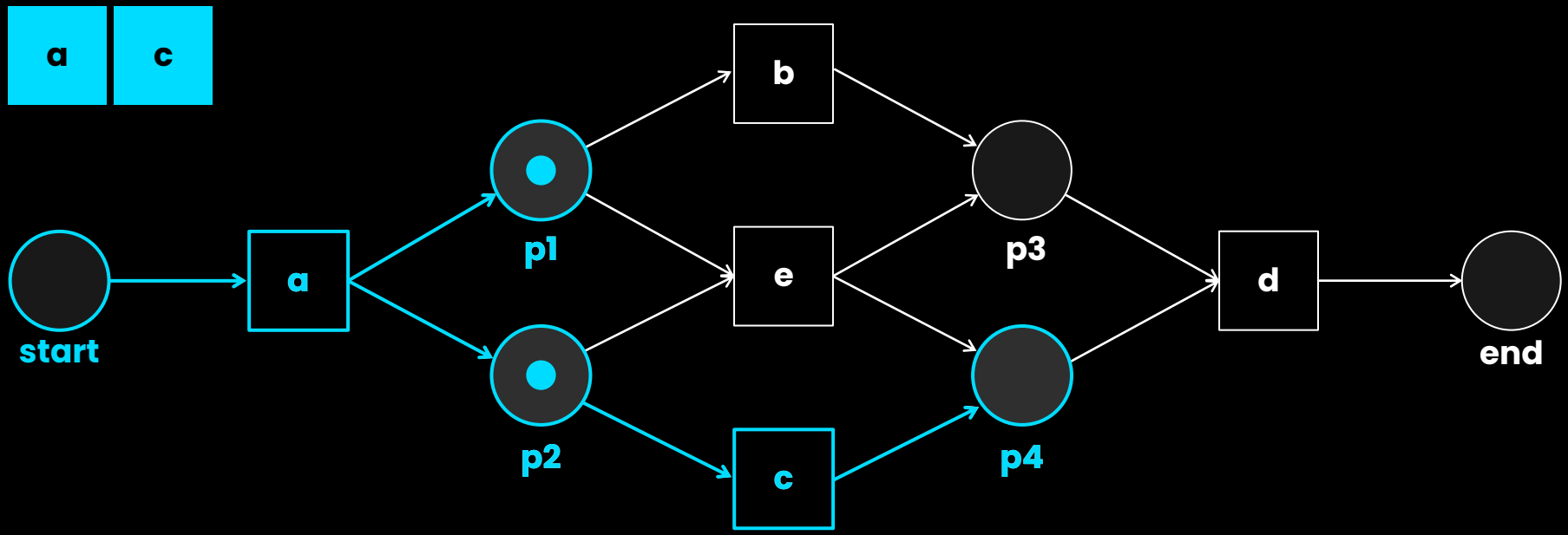
Petri nets: The basics



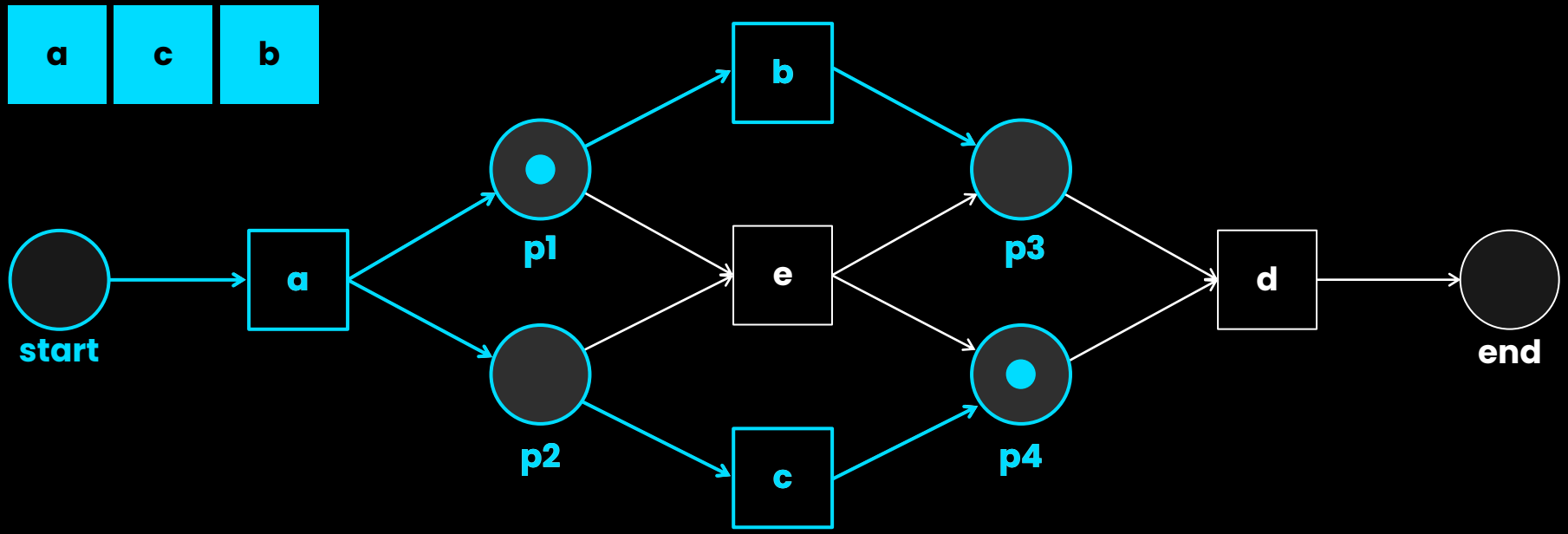
Petri nets: The basics



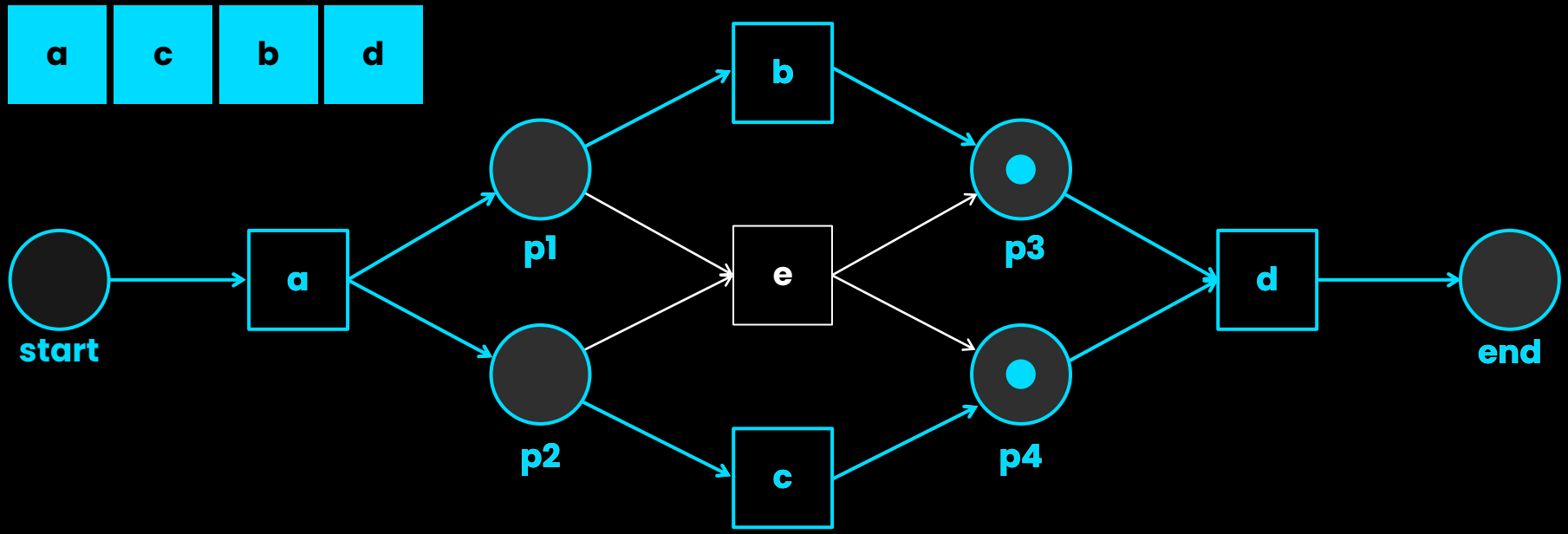
Petri nets: The basics



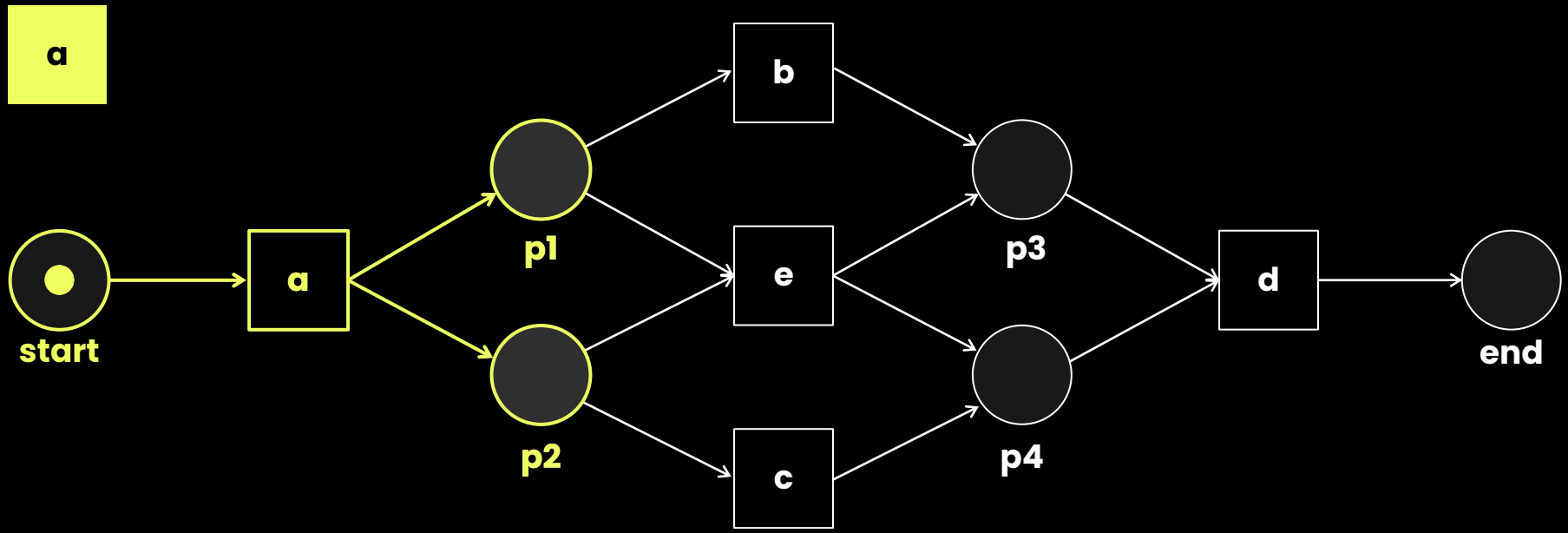
Petri nets: The basics



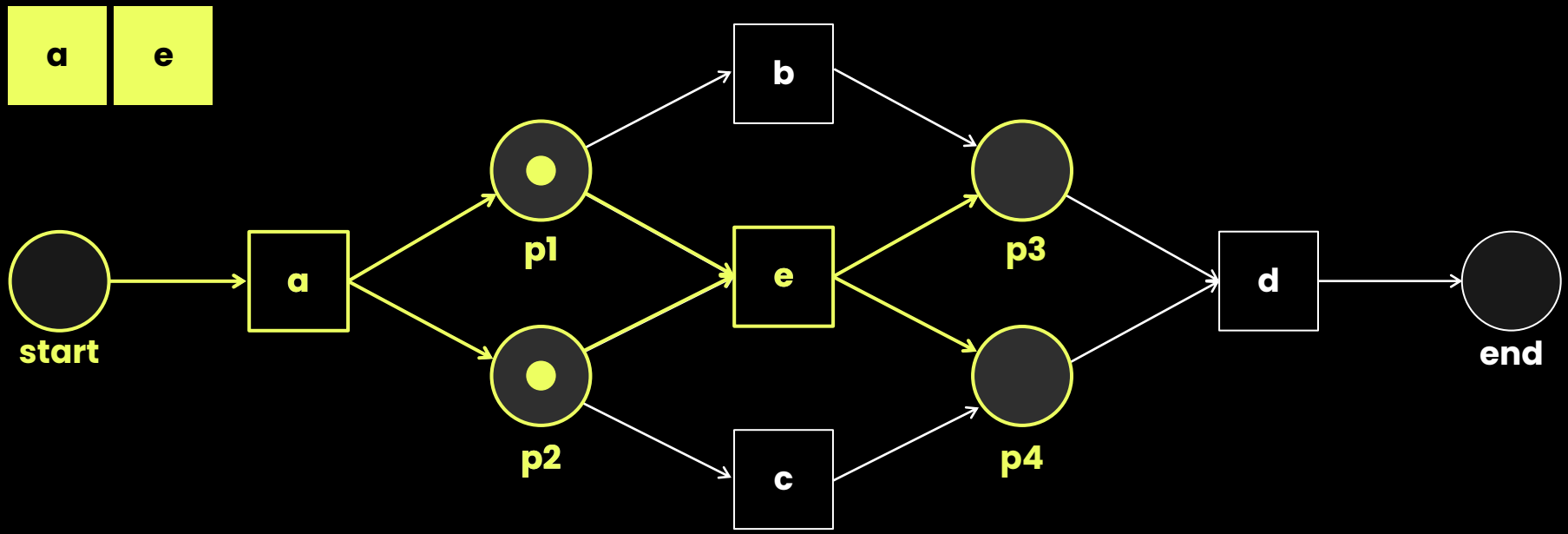
Petri nets: The basics



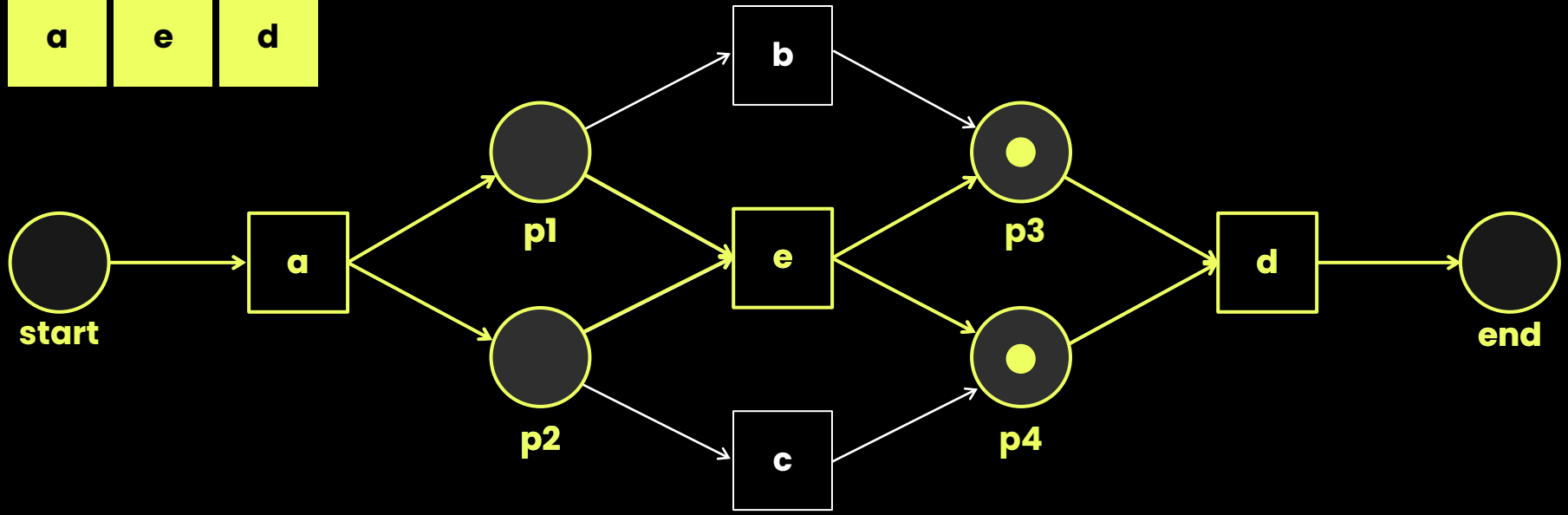
Petri nets: The basics



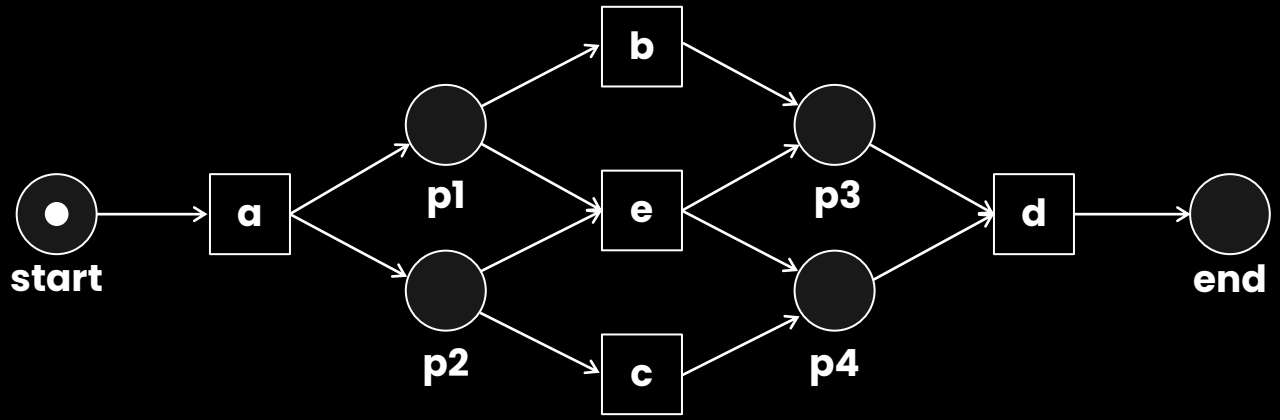
Petri nets: The basics



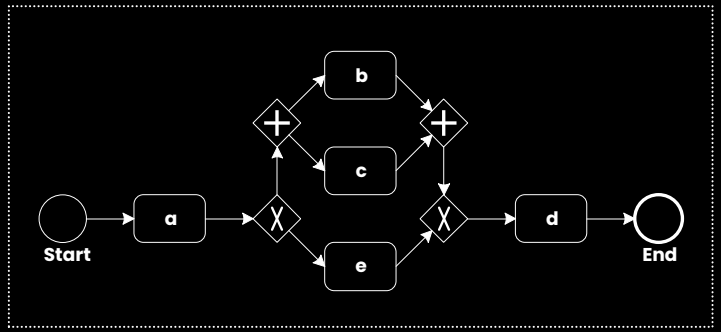
Petri nets: The basics



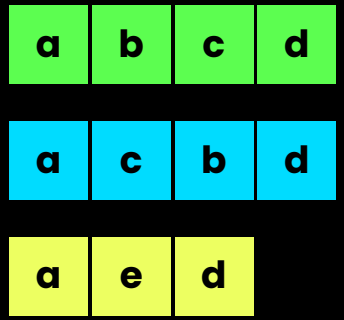
Petri nets: The basics

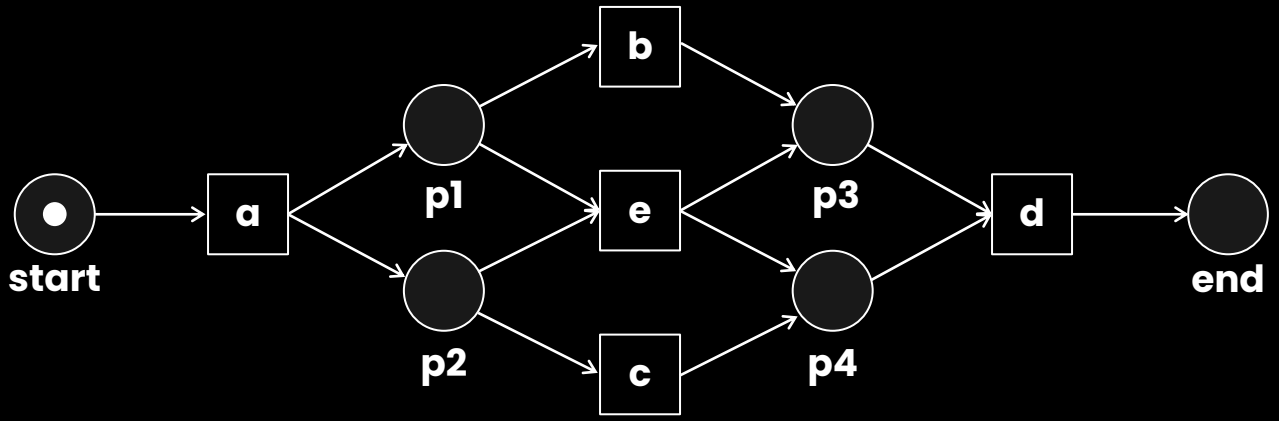


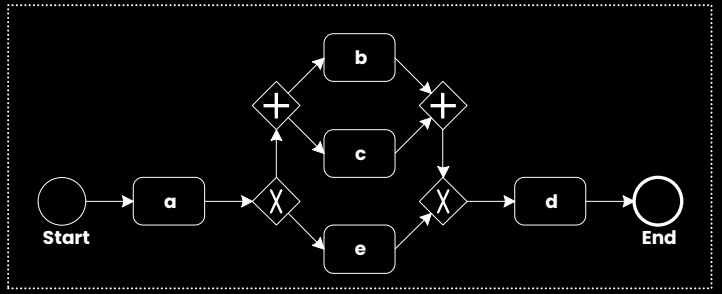
a	b	c	d
a	c	b	d
a	e	d	



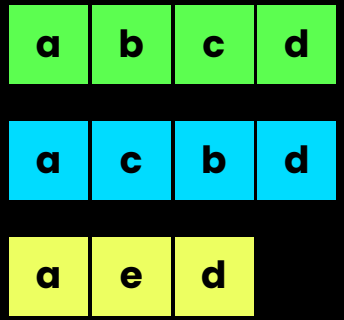
Process discovery

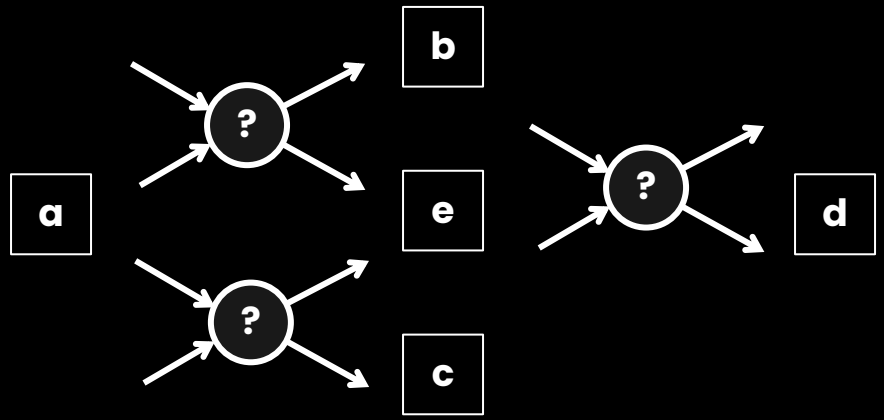






Bottom-up discovery: Finding places

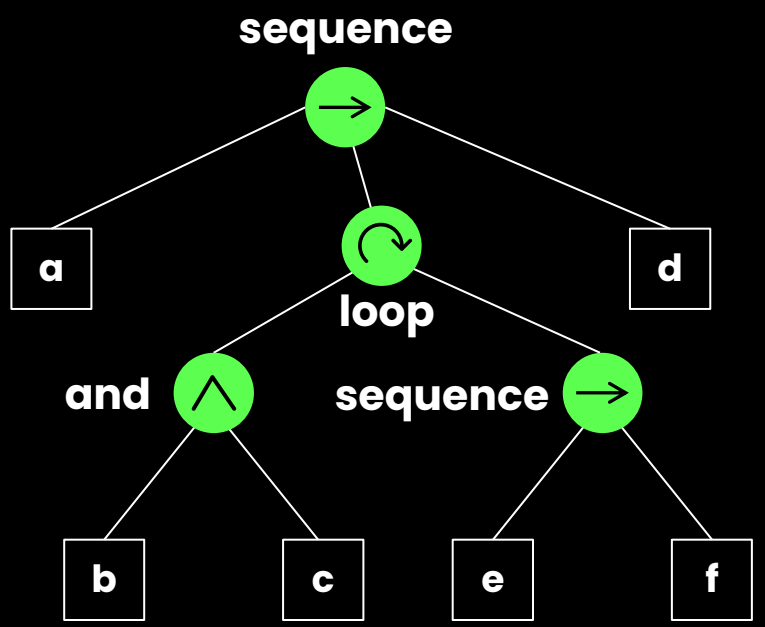




This is how the alpha algorithm, region-based techniques, etc. work!

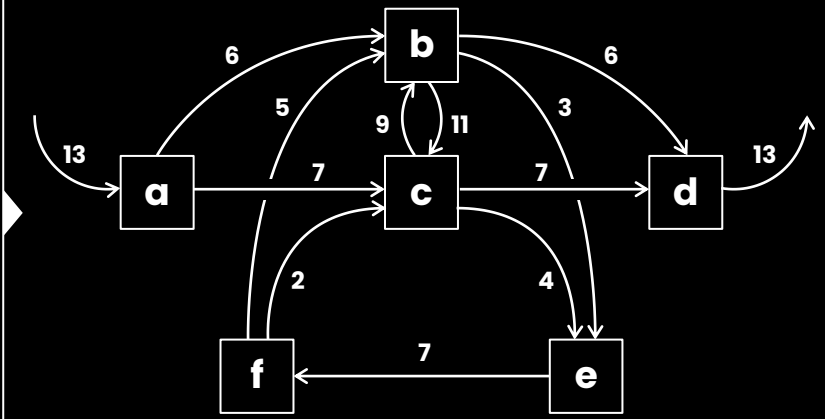
Top-down discovery: Inductive miner

3x	a	b	c	d							
4x	a	c	b	d							
2x	a	b	c	e	f	b	c	d			
2x	a	c	b	e	f	b	c	d			
1x	a	b	c	e	f	c	b	d			
1x	a	c	b	e	f	b	c	e	f	c	b	d

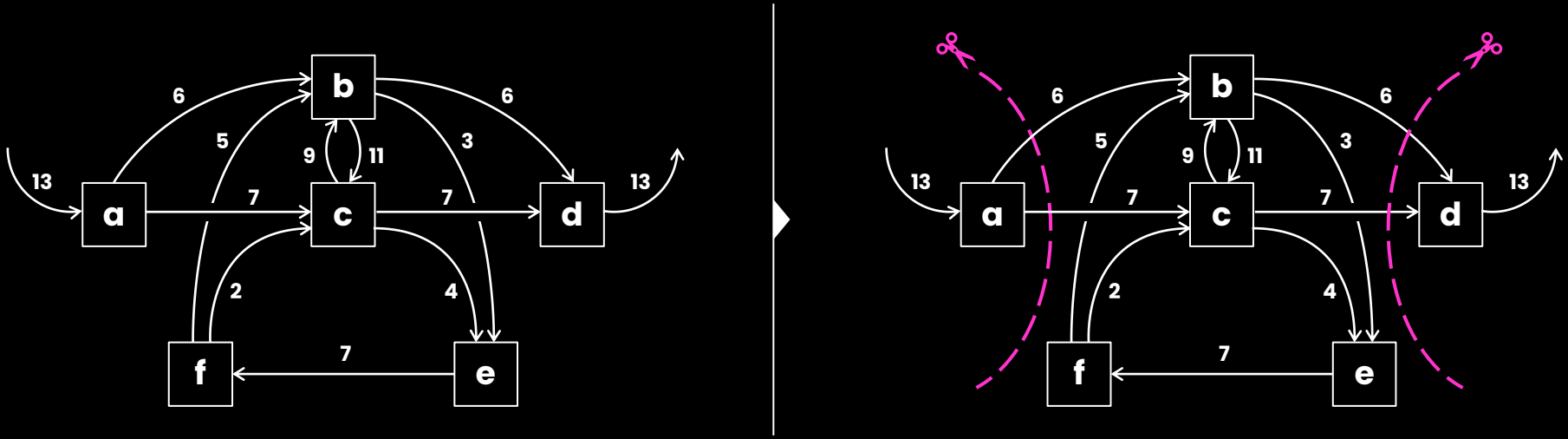


Directly-follows graph based on event log

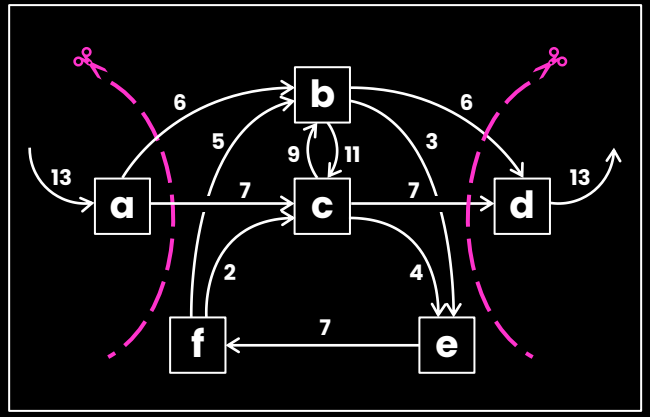
3x	a	b	c	d							
4x	a	c	b	d							
2x	a	b	c	e	f	b	c	d			
2x	a	c	b	e	f	b	c	d			
1x	a	b	c	e	f	c	b	d			
1x	a	c	b	e	f	b	c	e	f	c	b	d

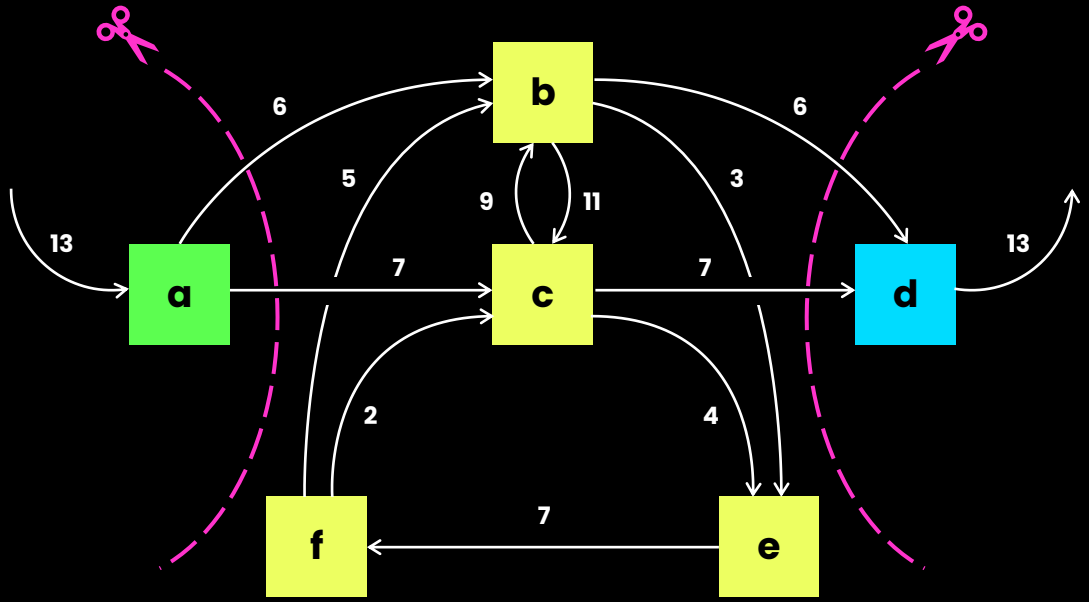


Sequence cut

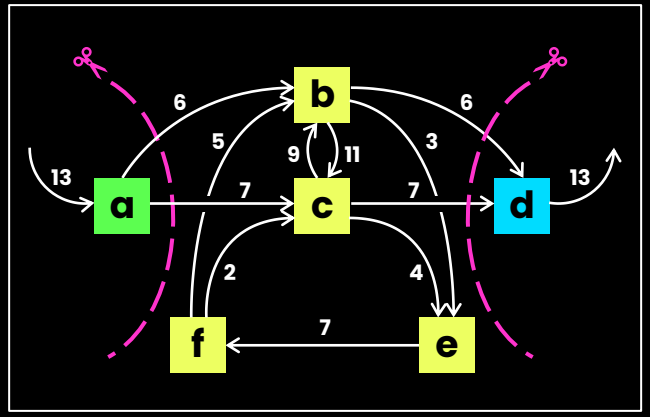


Partition activities based on sequence cut



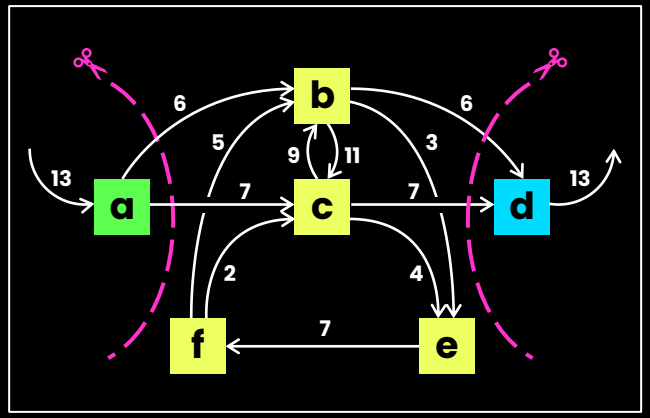


Partition events based on sequence cut



3x	a	b	c	d							
4x	a	c	b	d							
2x	a	b	c	e	f	b	c	d			
2x	a	c	b	e	f	b	c	d			
1x	a	b	c	e	f	c	b	d			
1x	a	c	b	e	f	b	c	e	f	c	b	d

Partition events based on sequence cut



3x a b c d

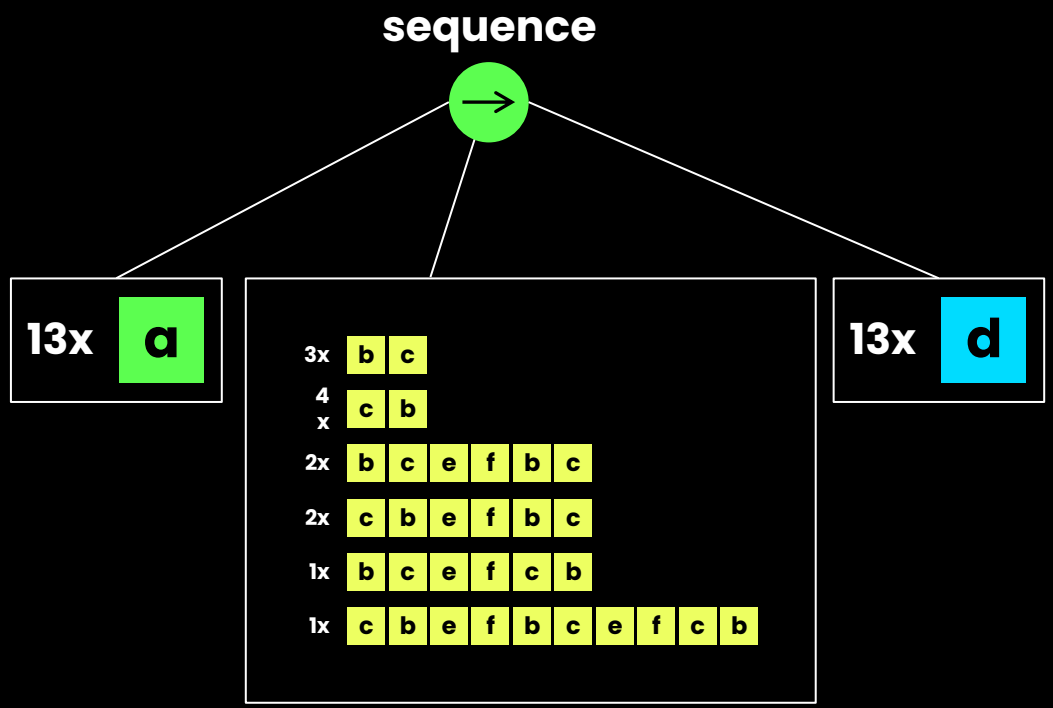
4x a c b d

2x a b c e f b c d

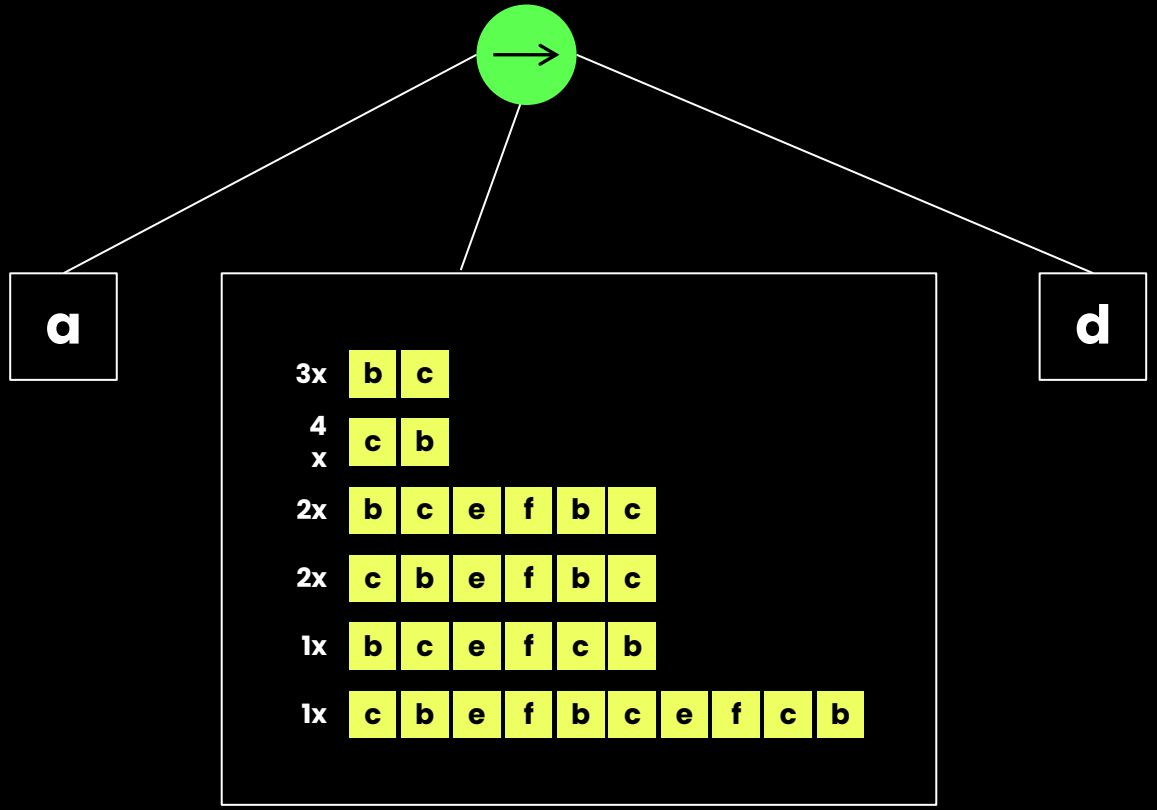
2x a c b e f b c d

1x a b c e f c b d

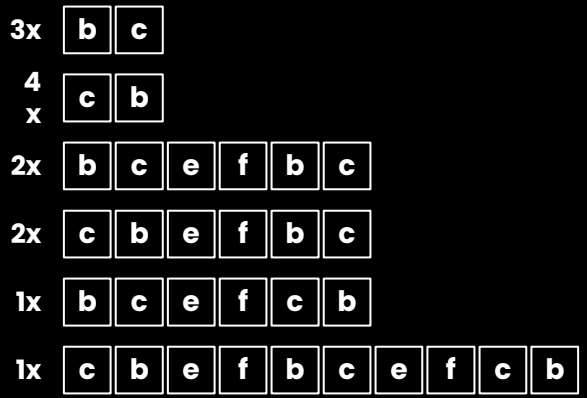
1x a c b e f b c e f c b d

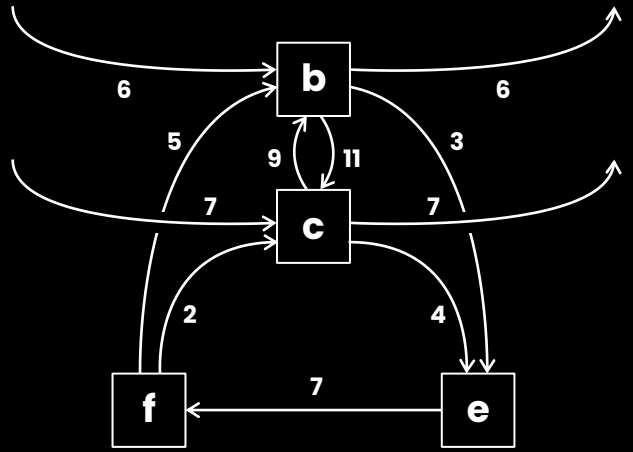


Recurse on non-base cases

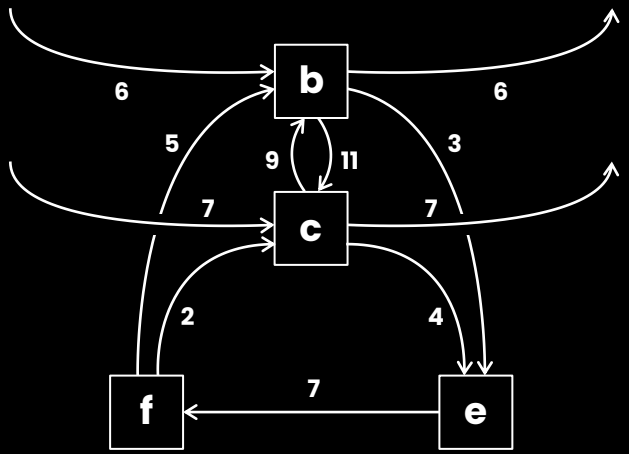


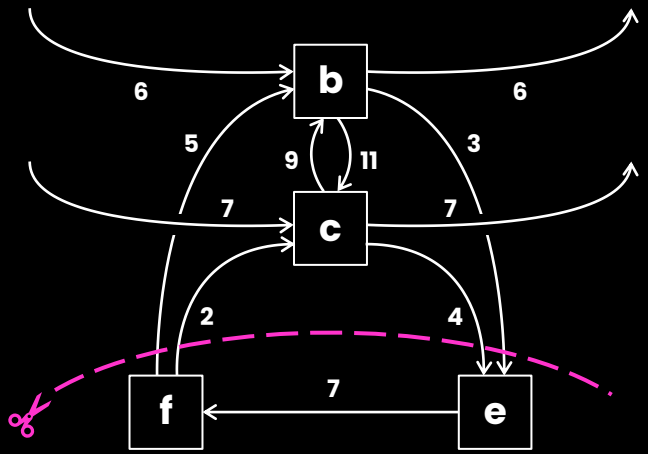
Directly-follows graph based on sublog



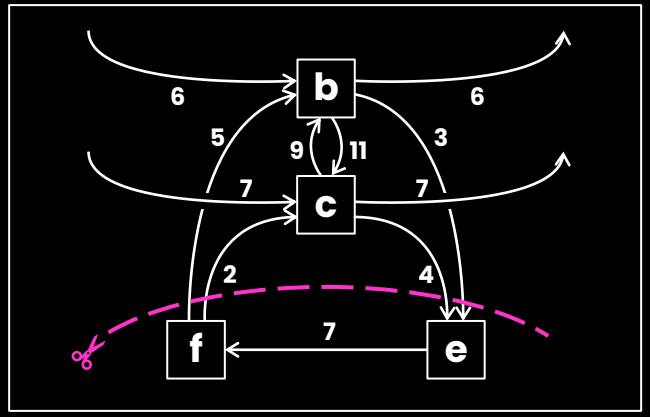


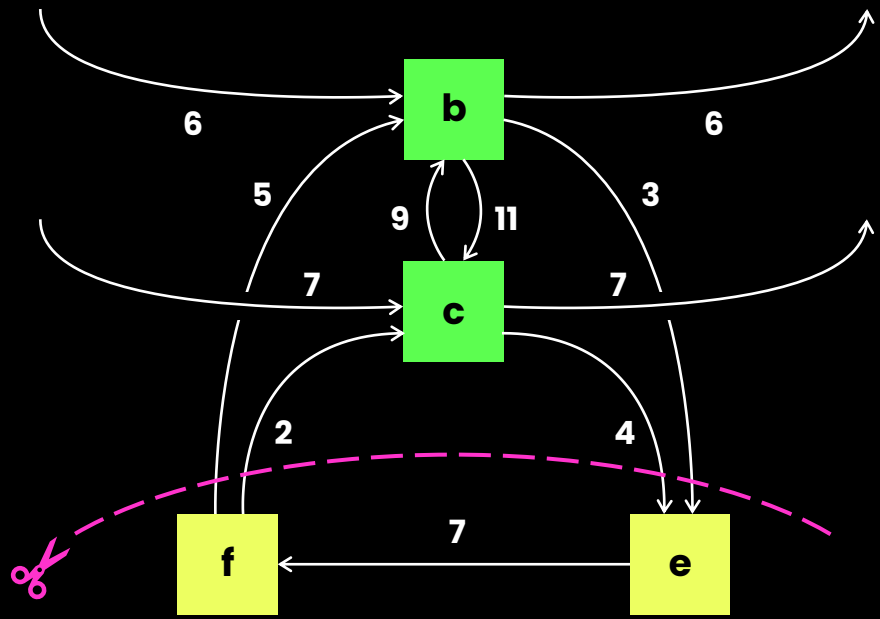
Loop cut



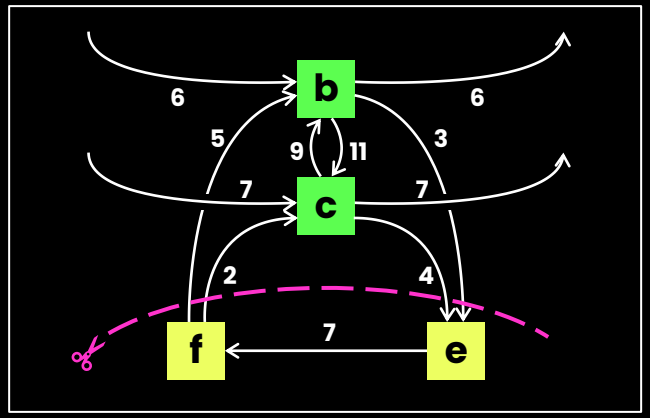


Partition activities based on loop cut



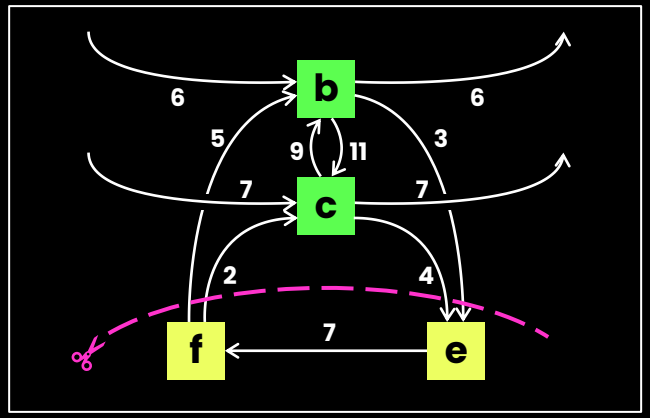


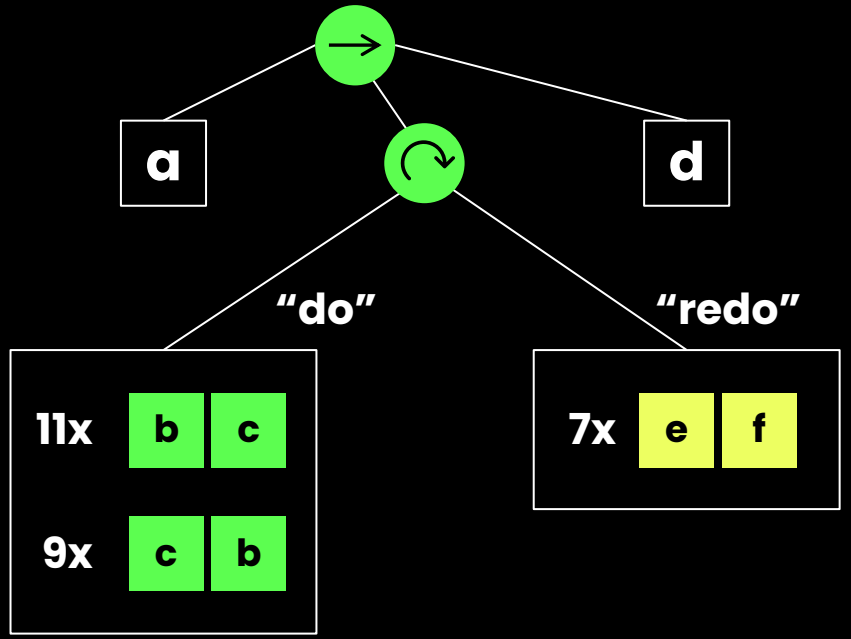
Partition events based on loop cut



3x	b	c							
4x	c	b							
2x	b	c	e	f	b	c			
2x	c	b	e	f	b	c			
1x	b	c	e	f	c	b			
1x	c	b	e	f	b	c	e	f	c	b

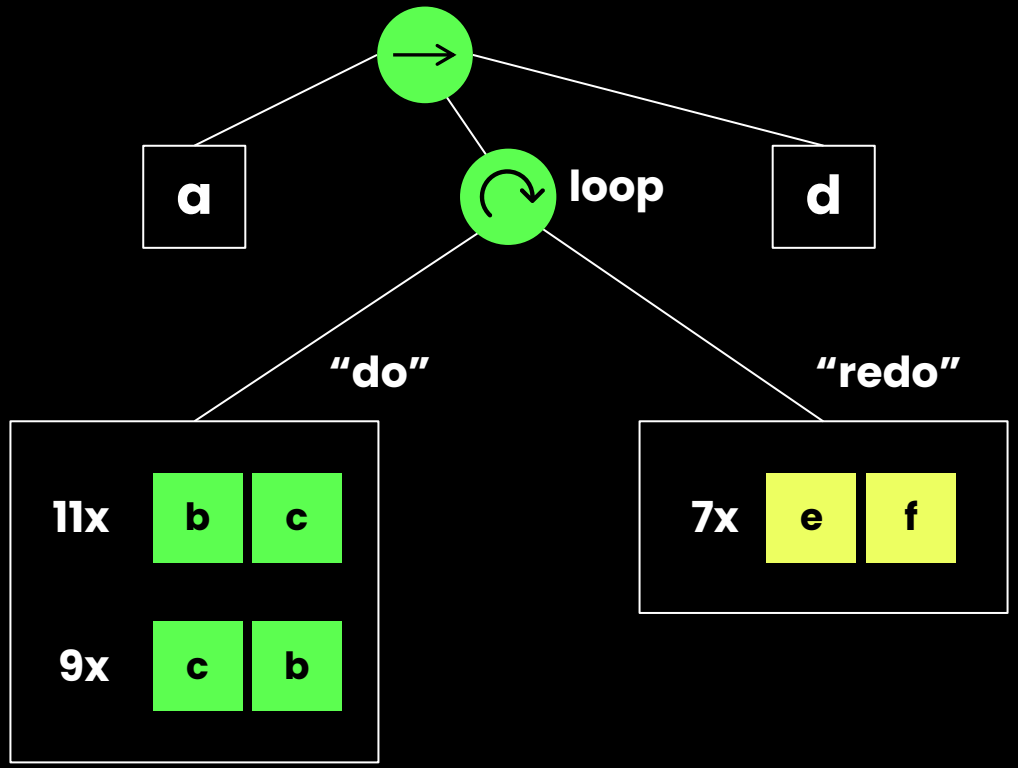
Partition events based on loop cut



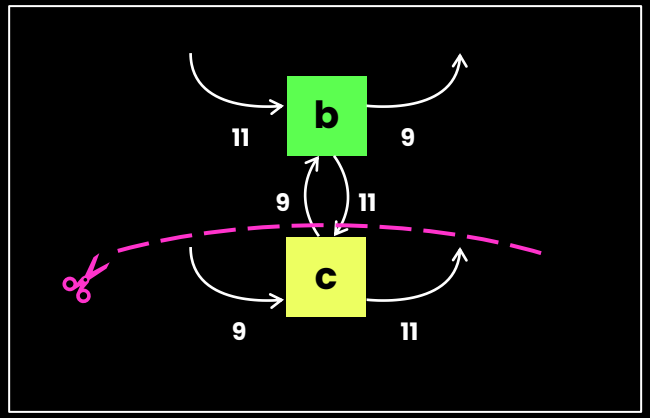


3 x	b	c							
4 x	c	b							
2 x	b	c	e	f	b	c			
2 x	c	b	e	f	b	c			
1x	b	c	e	f	c	b			
1x	c	b	e	f	b	c	e	f	c	b

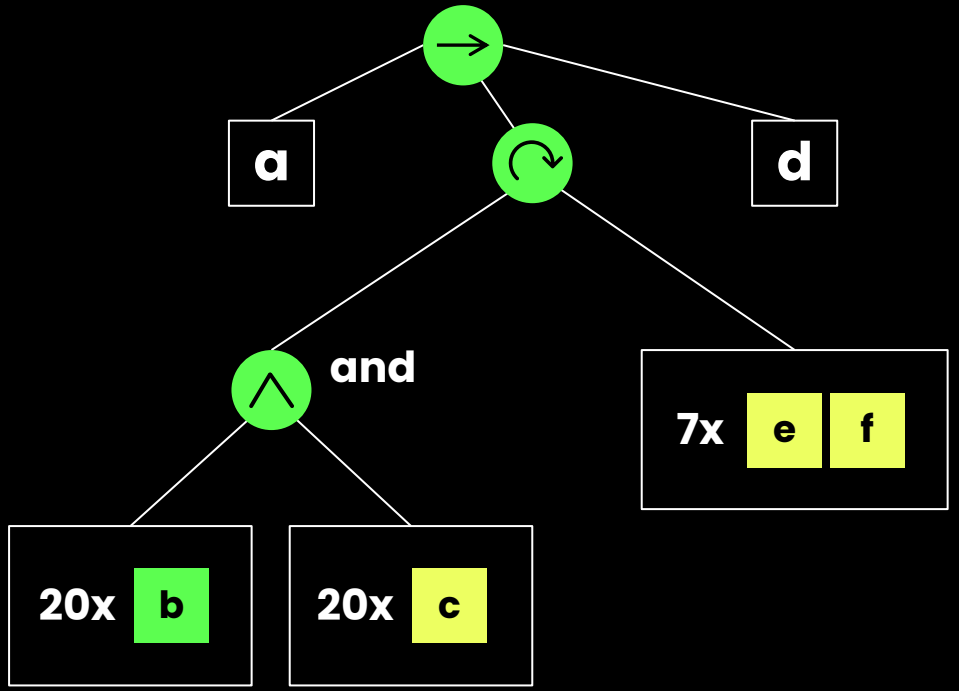
Recurse on the two sublogs



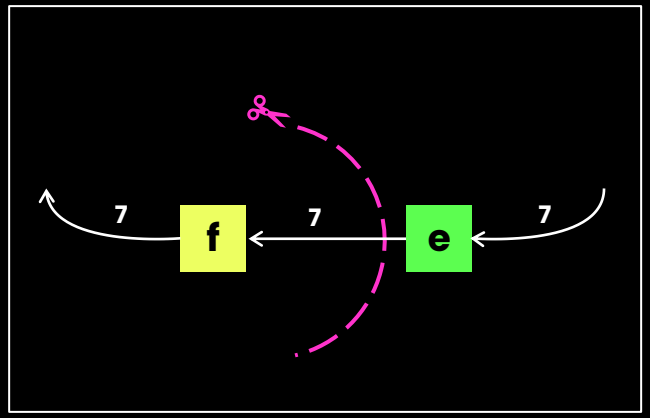
Partition events based on and cut



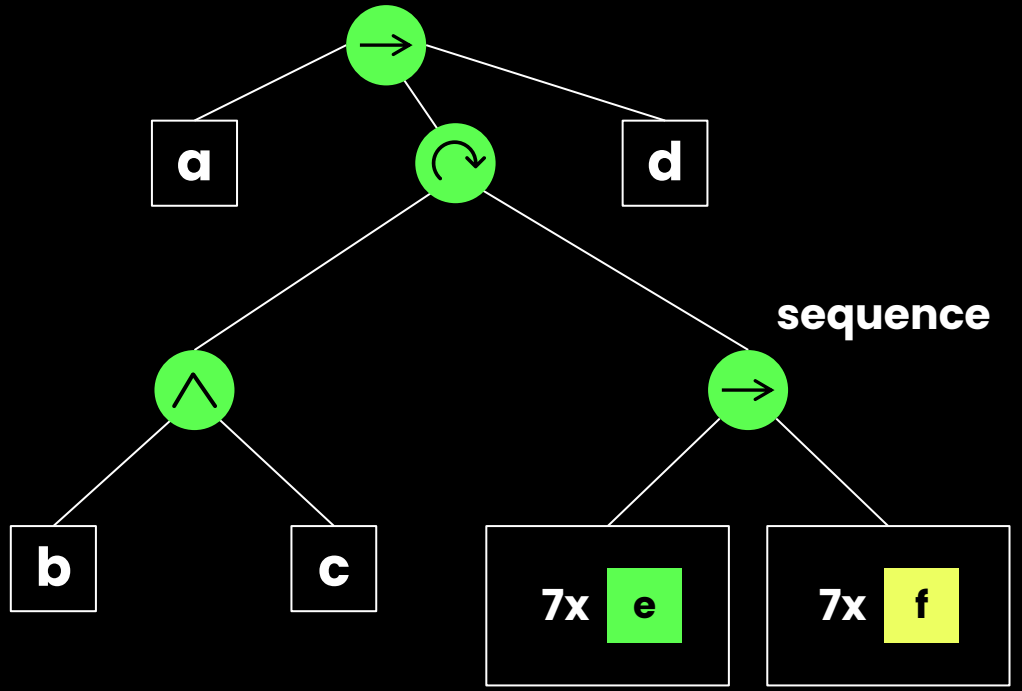
and cut



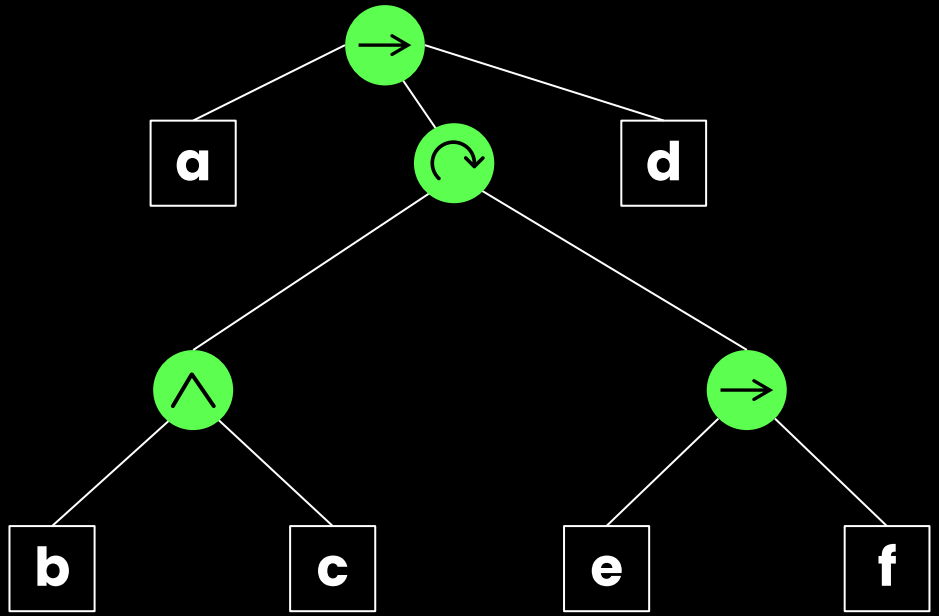
Partition events based on sequence cut



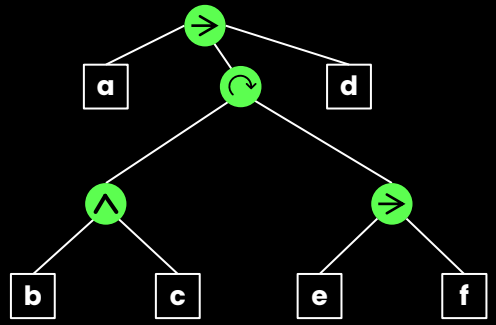
sequence cut



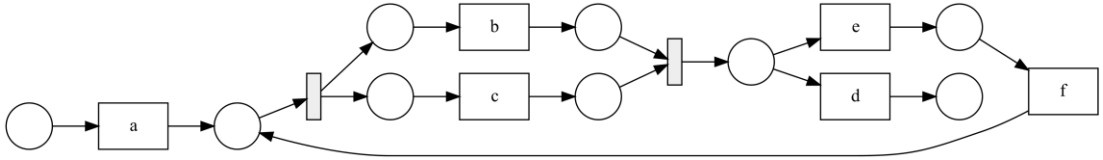
Final model



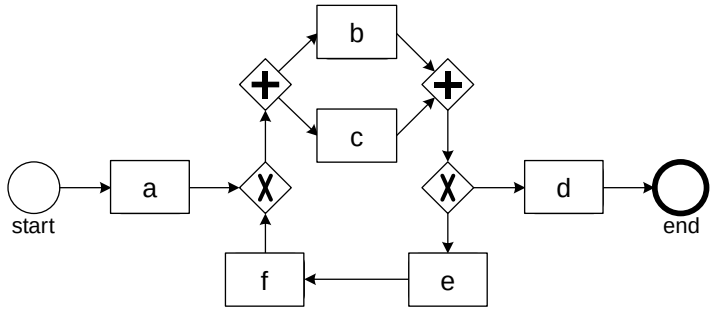
Different visualizations of the same discovered model



Petri net

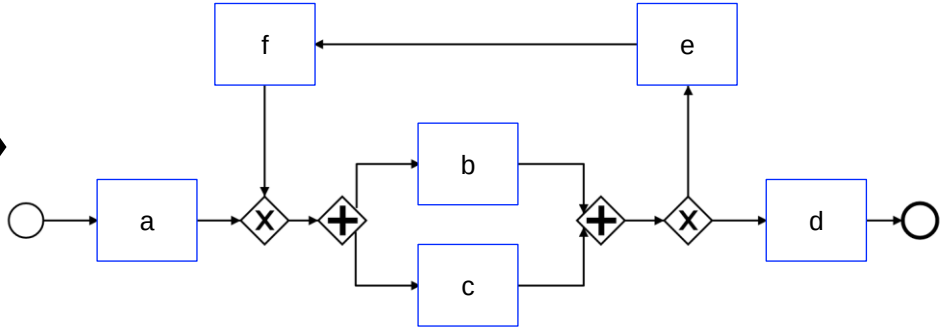


BPMN

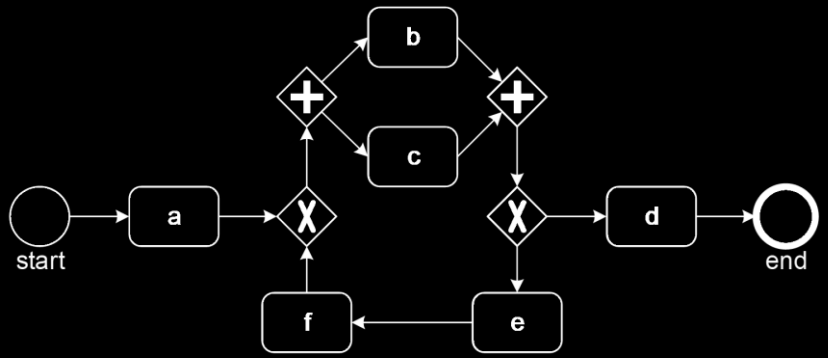


Celonis visualizes the discovered process tree as a BPMN model

3x	a	b	c	d								
4x	a	c	b	d								
2x	a	b	c	e	f	b	c	d				
2x	a	c	b	e	f	b	c	d				
1x	a	b	c	e	f	c	b	d				
1x	a	c	b	e	f	b	c	e	f	c	b	d



DFG is severely underfitting



3x a b c d

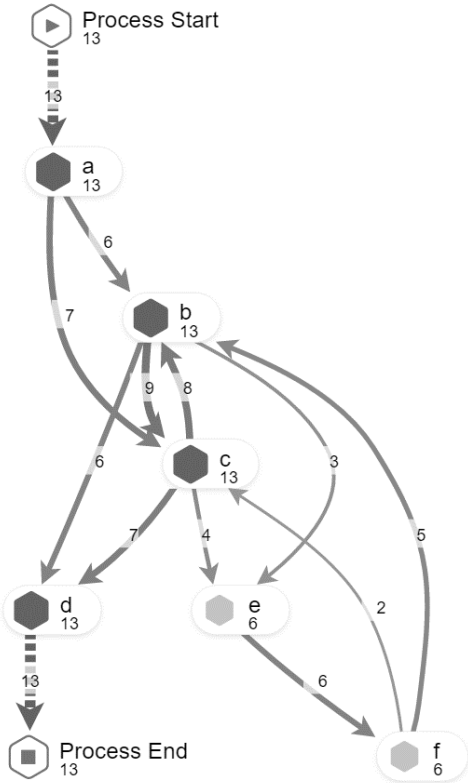
4x a c b d

2x a b c e f b c d

2x a c b e f b c d

1x a b c e f c b d

1x a c b e f b c e f c b d



Examples of unobserved traces allowed by the DFG

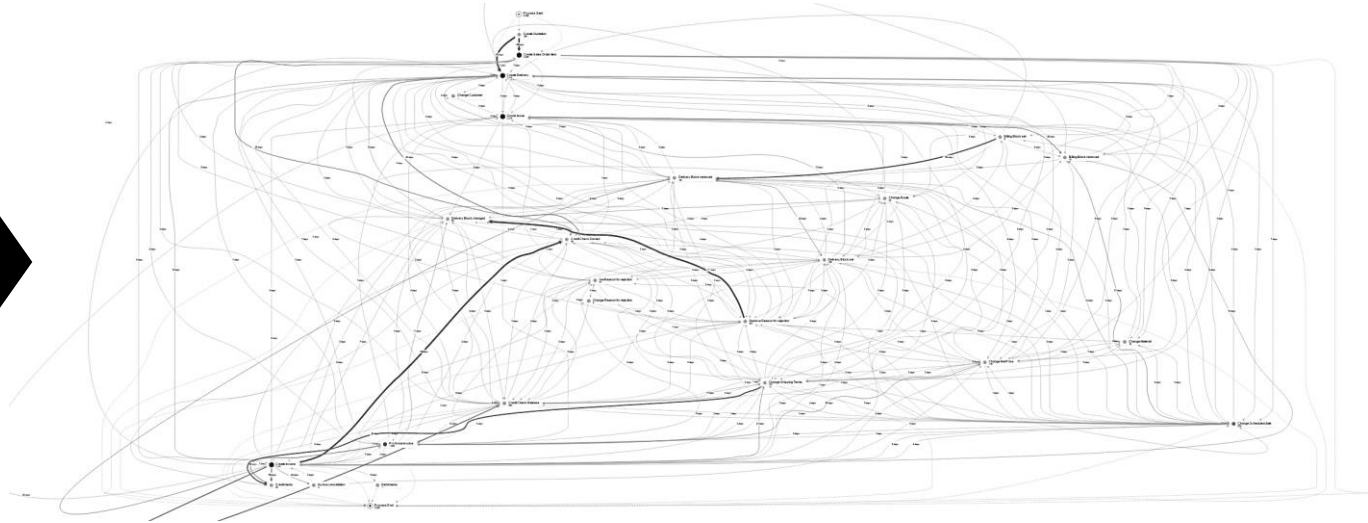
a c d

a b c b c b d

Larger example: O2C

3,308 cases

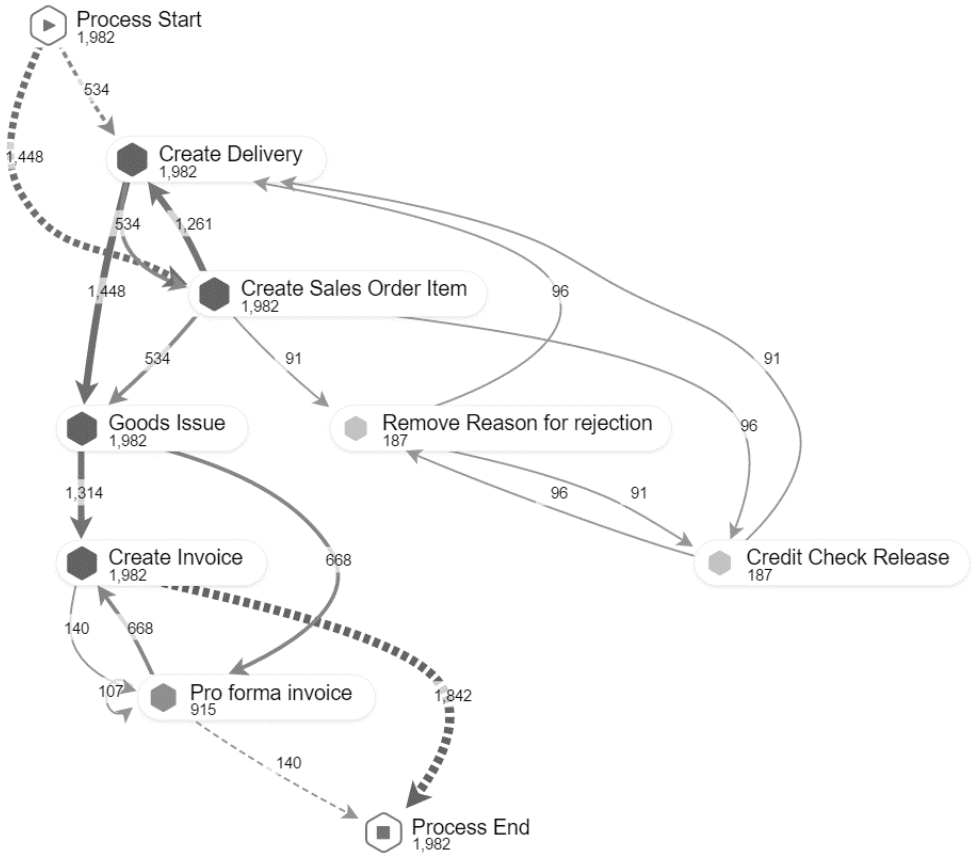
501 variants



DFG based on top 8 variants

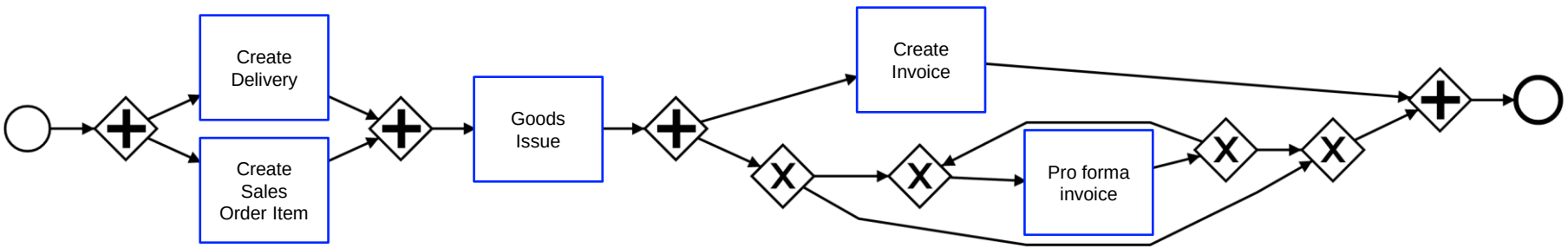
60% of cases

1.6% of variants



BPMN based on top 5 variants

Also covers 60% of cases, but is much simpler and does not show non-existent loops



Next lecture

Conformance checking