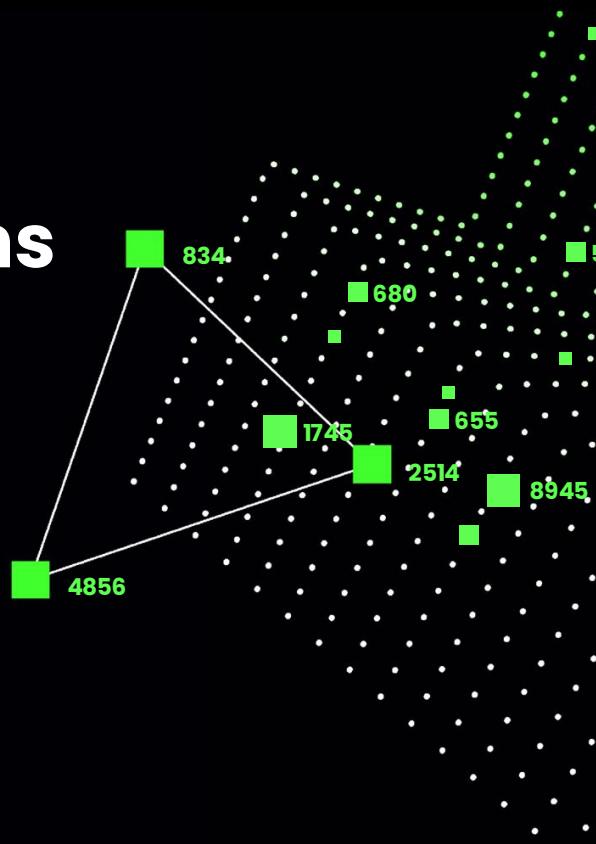
celonis

Chair of Process and Data Science | RWTH AACHEN UNIVERSITY

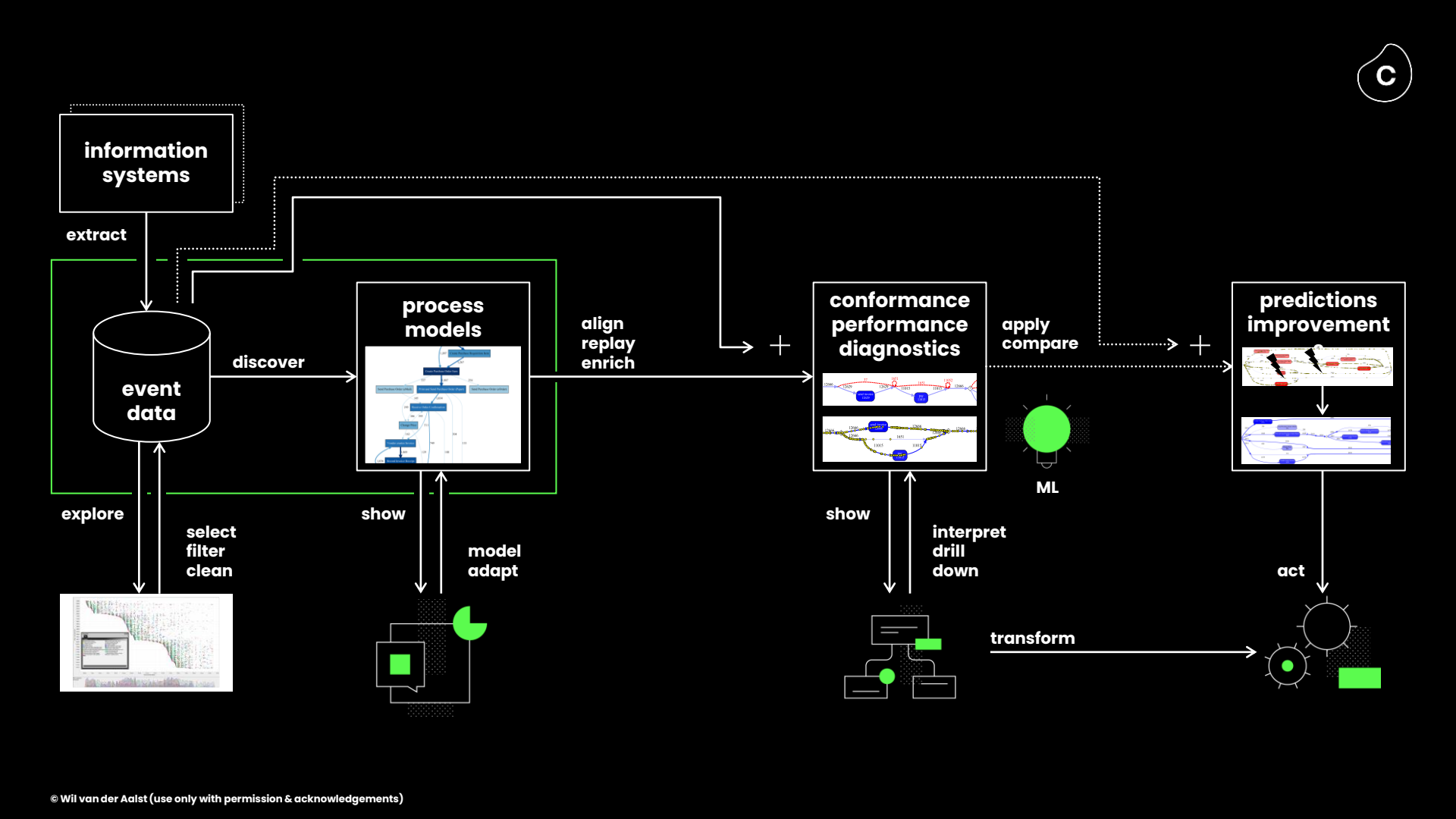
Process Discovery 1/2

Learning 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



© Wil van der Aalst (use only with permission & acknowledgements)

Remember: Event data

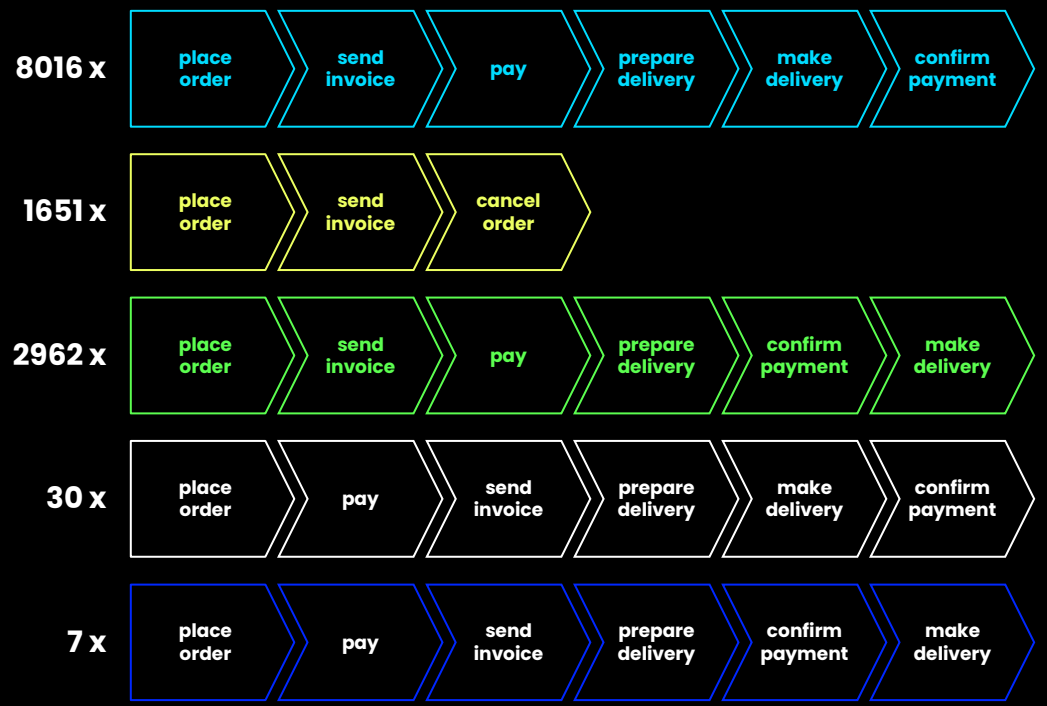
Case ID	Activity	Resource	Timestamp	Product	Prod-price	Quantity	Address
...	...	....	...	....	...	...	...
6350	place order	Aiden	2018/02/13 14:29:45.000	APPLE iPhone 6 16 GB	639,00 €	5	NL-7751DG-21
6283	pay	Lily	2018/02/13 14:39:25.000	SAMSUNG Galaxy S6 32 GB	543,99 €	3	NL-7828AM-11a
6253	prepare delivery	Sophia	2018/02/13 15:01:33.000	APPLE iPhone 6 16 GB	639,00 €	3	NL-7887AC-13
6257	prepare delivery	Aiden	2018/02/13 15:03:43.000	SAMSUNG Galaxy S6 32 GB	543.99 €	1	NL-9521KJ-34
6185	confirm payment	Emily	2018/02/13 15:05:36.000	SAMSUNG Galaxy S4	329,00 €	1	NL-9521GC-32
6218	confirm payment	Emily	2018/02/13 15:08:11.000	APPLE iPhone 6s 64 GB	969,00 €	2	NL-7948BX-10
6245	make delivery	Michael	2018/02/13 15:14:04.000	APPLE iPhone 6 16 GB	639,00 €	3	NL-7905AX-38
6272	pay	Emily	2018/02/13 15:20:36.000	APPLE iPhone 6 16 GB	639,00 €	1	NL-7821AC-3
6269	pay	Charlotte	2018/02/13 15:25:21.000	SAMSUNG Galaxy S4	329,00 €	1	NL-7907EJ-42
6212	prepare delivery	Sophia	2018/02/13 15:43:39.000	HUAWEI P8 Lite	169,00 €	1	NL-7905AX-38
6323	send invoice	Alexander	2018/02/13 15:46:08.000	APPLE iPhone 6 16 GB	639,00 €	1	NL-7833HT-15
6246	confirm payment	Jack	2018/02/13 15:56:03.000	SAMSUNG Galaxy S4	329,00 €	3	NL-7833HT-15
6347	send invoice	Jack	2018/02/13 15:57:42.000	SAMSUNG Galaxy S4	329,00 €	3	NL-7905AX-38
6351	place order	Zoe	2018/02/13 16:17:37.000	APPLE iPhone 6s 16 GB	549,00 €	3	NL-9521GC-32
6204	prepare delivery	Sophia	2018/02/13 16:31:28.000	SAMSUNG Core Prime G361	135,00 €	1	NL-7828AM-11a
6204	make delivery	Kaylee	2018/02/13 16:51:54.000	SAMSUNG Core Prime G361	135,00 €	1	NL-7828AM-11a
6265	confirm payment	Lily	2018/02/13 16:55:55.000	SAMSUNG Galaxy S4	329,00 €	4	NL-9521GC-32
6250	confirm payment	Jack	2018/02/13 17:03:26.000	MOTOROLA Moto G	199,00 €	4	NL-7942GT-2
6328	send invoice	Lily	2018/02/13 17:30:16.000	APPLE iPhone 6s 64 GB	858,00 €	4	NL-9514BV-16
6352	place order	Aiden	2018/02/13 17:53:22.000	APPLE iPhone 6 16 GB	639,00 €	2	NL-9514BV-16
6317	send invoice	Jack	2018/02/13 18:45:30.000	APPLE iPhone 6s 64 GB	858,00 €	5	NL-7907EJ-42
6353	place order	Sophia	2018/02/13 20:16:20.000	APPLE iPhone 5s 16 GB	449,00 €	4	NL-7751AR-19
...	...	....	...	....	...	...	...

event =
case +
activity +
timestamp +

© Wil van der Aalst (use only with permission & acknowledgements)

Simplified Event data

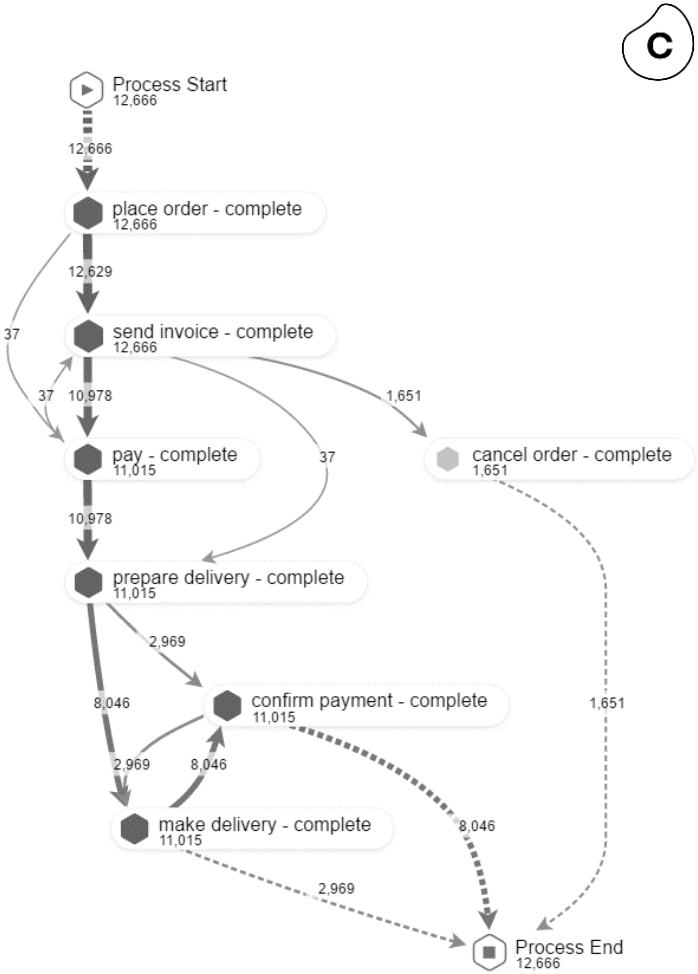
Case ID	Activity	Timestamp
...	...	...
6350	place order	2018/02/13 14:29:45.000
6283	pay	2018/02/13 14:39:25.000
6253	prepare delivery	2018/02/13 15:01:33.000
6257	prepare delivery	2018/02/13 15:03:43.000
6185	confirm payment	2018/02/13 15:05:36.000
6218	confirm payment	2018/02/13 15:08:11.000
6245	make delivery	2018/02/13 15:14:04.000
6272	pay	2018/02/13 15:20:36.000
6269	pay	2018/02/13 15:25:21.000
6212	prepare delivery	2018/02/13 15:43:39.000
6323	send invoice	2018/02/13 15:46:08.000
6246	confirm payment	2018/02/13 15:56:03.000
6347	send invoice	2018/02/13 15:57:42.000
6351	place order	2018/02/13 16:17:37.000
6204	prepare delivery	2018/02/13 16:31:28.000
6204	make delivery	2018/02/13 16:51:54.000
6265	confirm payment	2018/02/13 16:55:55.000
6250	confirm payment	2018/02/13 17:03:26.000
6328	send invoice	2018/02/13 17:30:16.000
6352	place order	2018/02/13 17:53:22.000
6317	send invoice	2018/02/13 18:45:30.000
6353	place order	2018/02/13 20:16:20.000
...	...	...



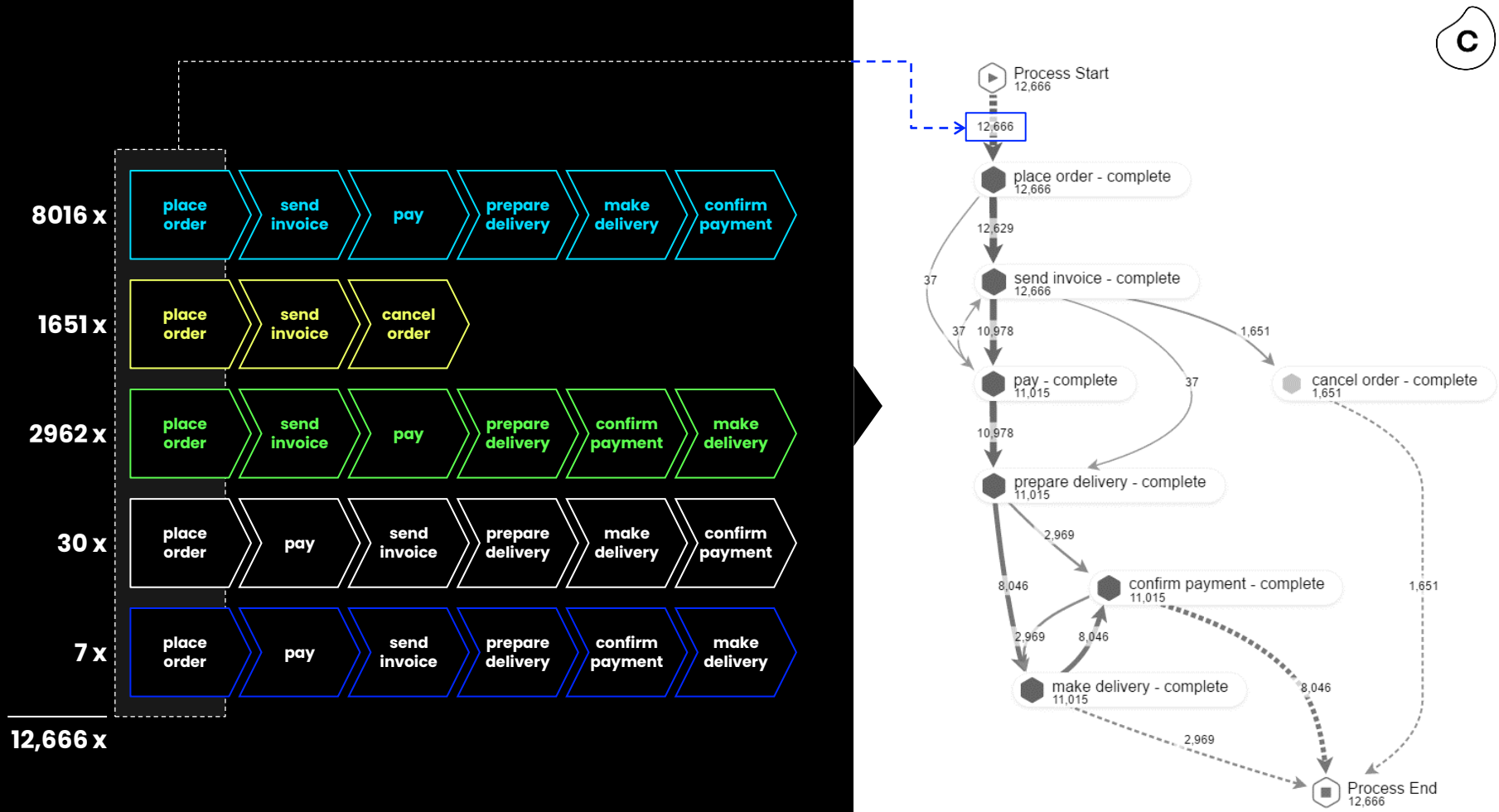
© Wil van der Aalst (use only with permission & acknowledgements)

Directly-Follows Graph (DFG)

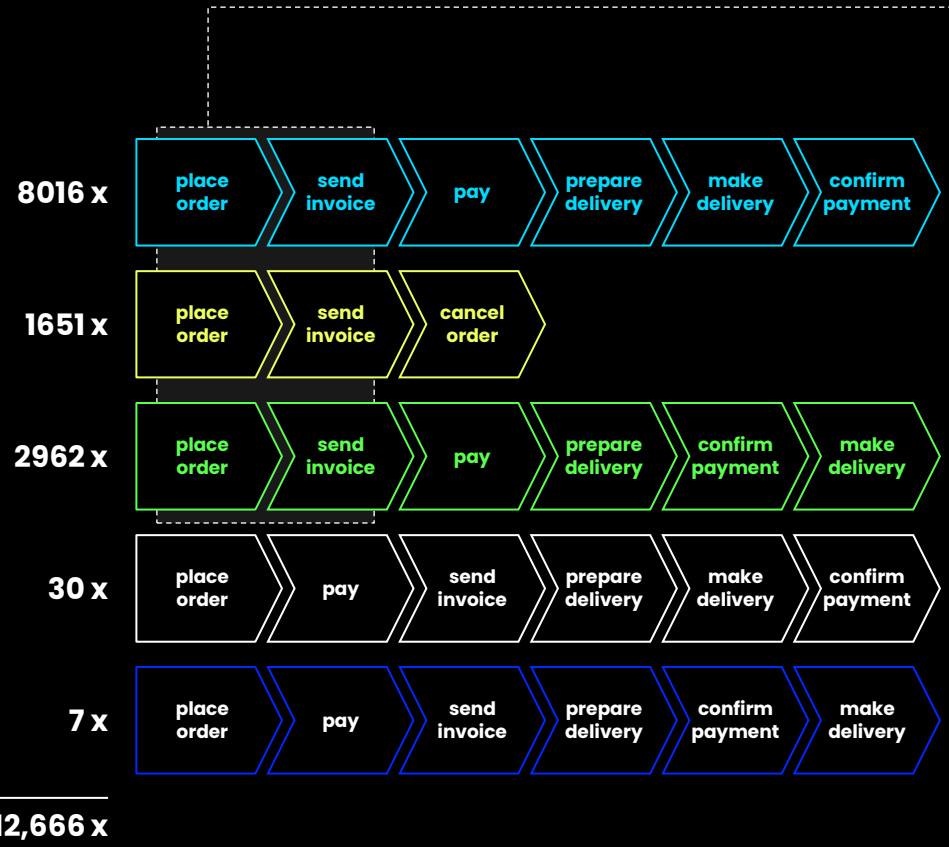
8016 x	place order	send invoice	pay	prepare delivery	make delivery	confirm payment
1651 x	place order	send invoice	cancel order			
2962 x	place order	send invoice	pay	prepare delivery	confirm payment	make delivery
30 x	place order	pay	send invoice	prepare delivery	make delivery	confirm payment
7 x	place order	pay	send invoice	prepare delivery	confirm payment	make delivery
12,666 x						

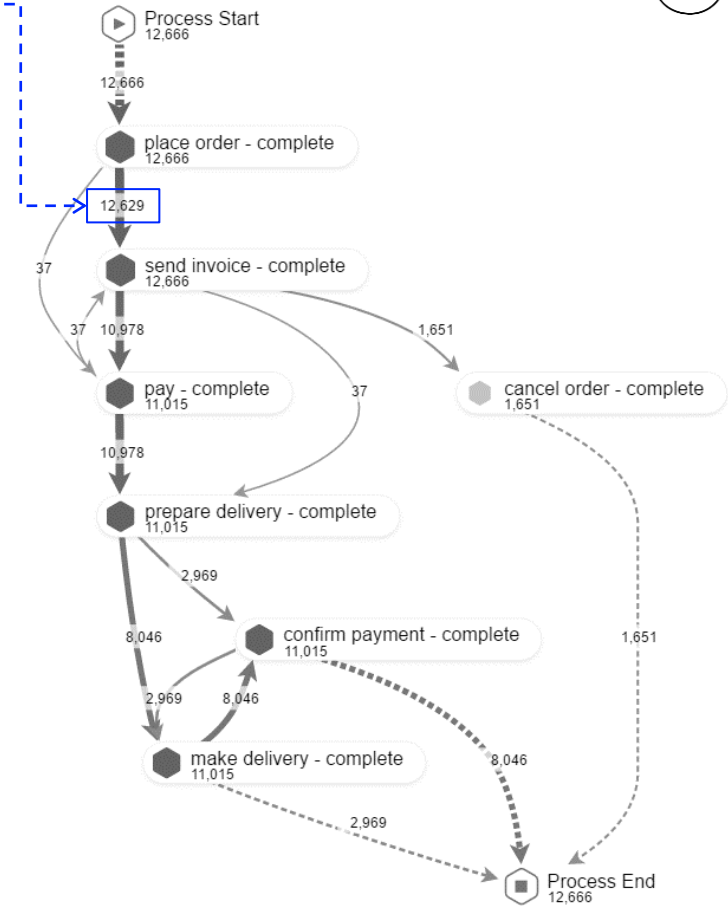


© Wil van der Aalst (use only with permission & acknowledgements)

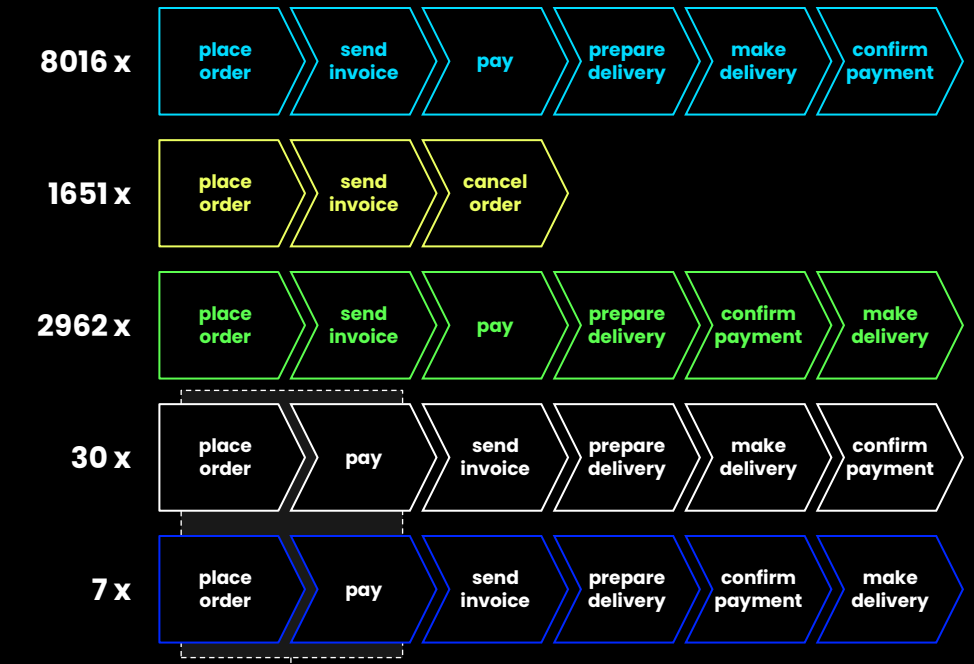


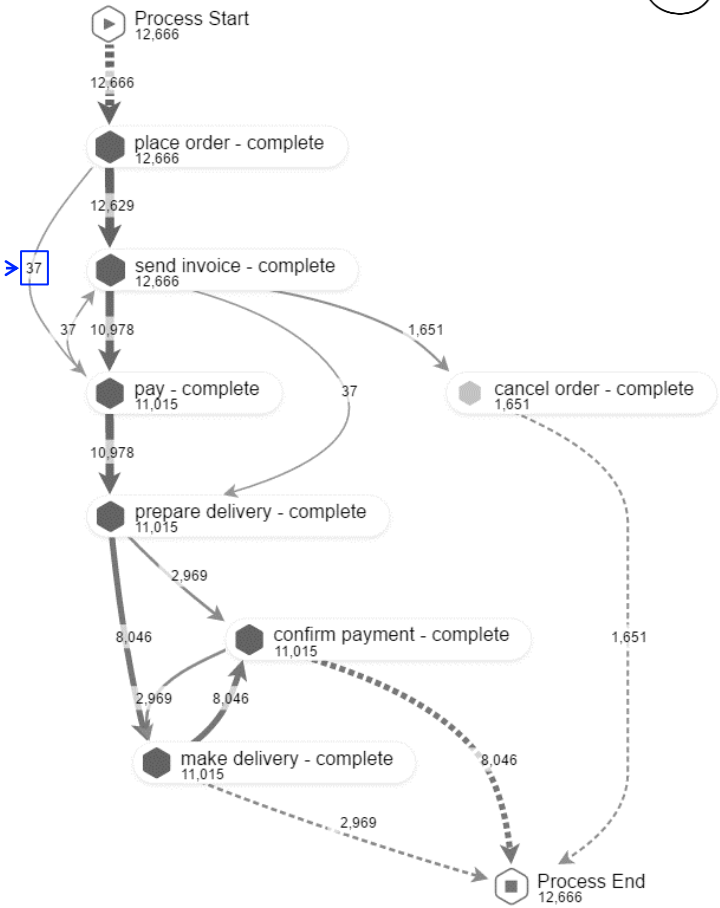
© Wil van der Aalst (use only with permission & acknowledgements)



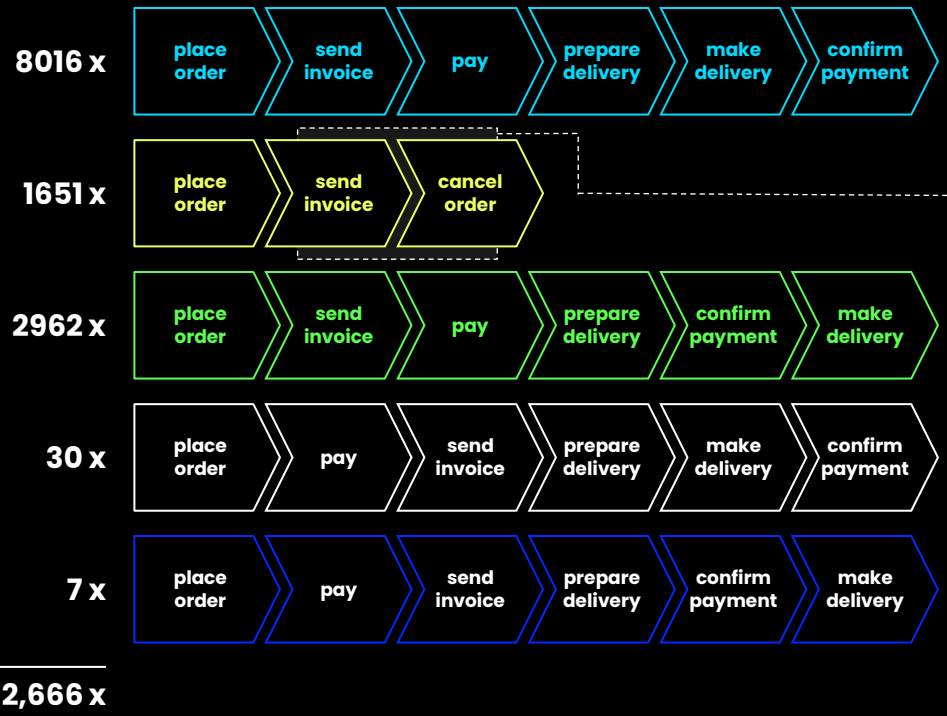


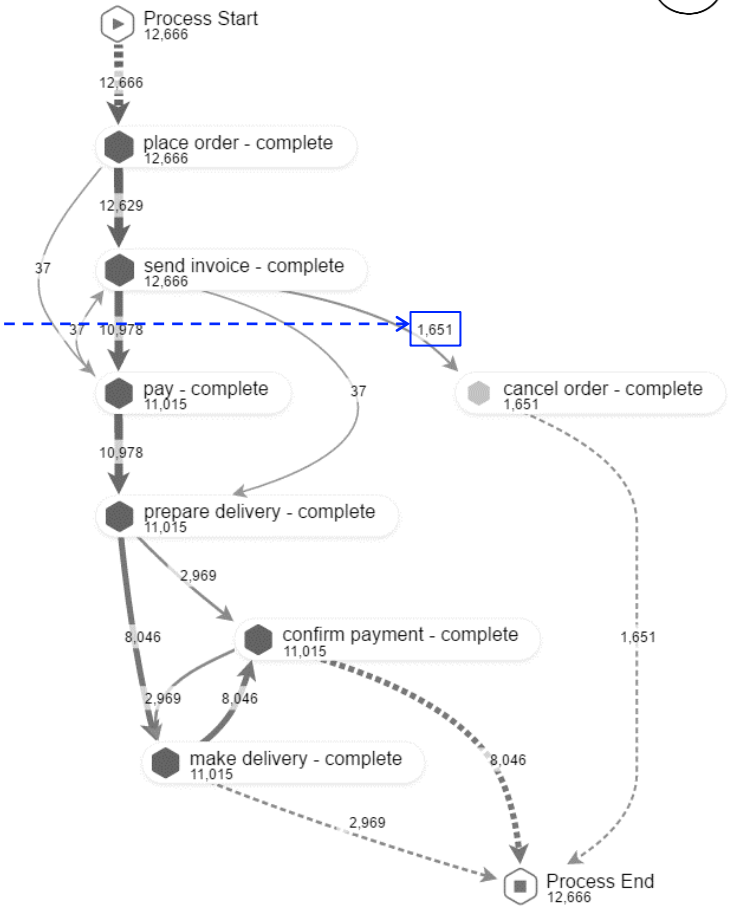
© Wil van der Aalst (use only with permission & acknowledgements)





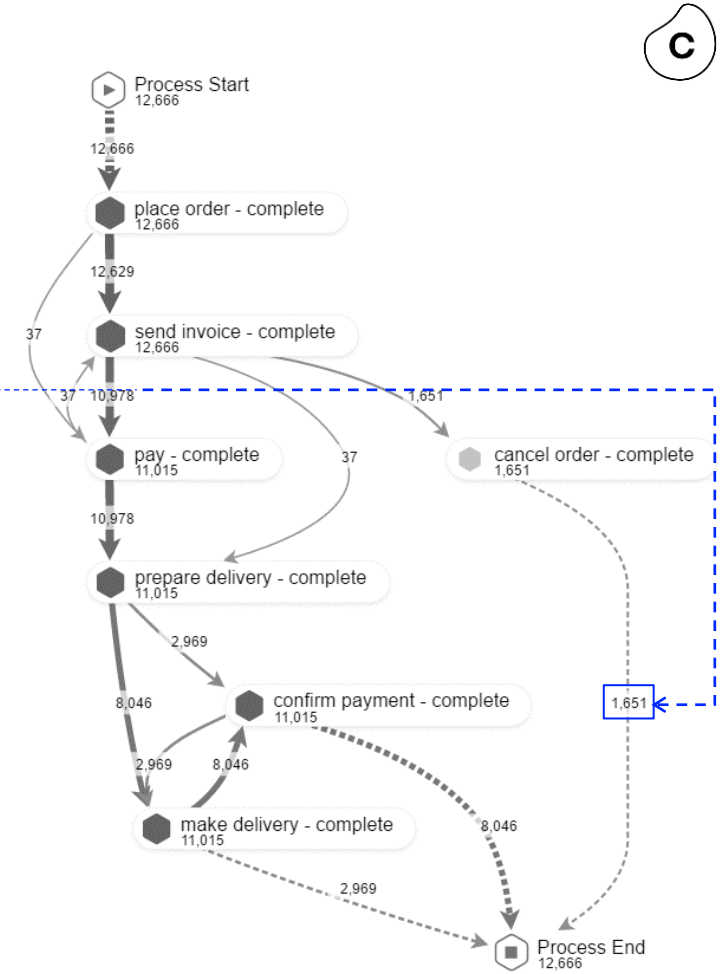
© Wil van der Aalst (use only with permission & acknowledgements)





© Wil van der Aalst (use only with permission & acknowledgements)

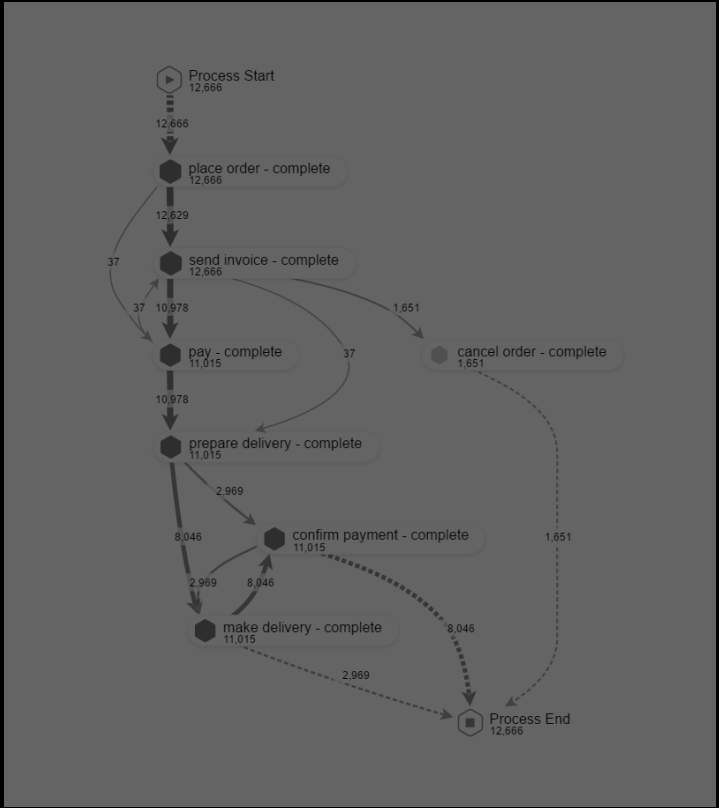
8016 x	place order	send invoice	pay	prepare delivery	make delivery	confirm payment
1651 x	place order	send invoice	cancel order			
2962 x	place order	send invoice	pay	prepare delivery	confirm payment	make delivery
30 x	place order	pay	send invoice	prepare delivery	make delivery	confirm payment
7 x	place order	pay	send invoice	prepare delivery	confirm payment	make delivery
12,666 x						

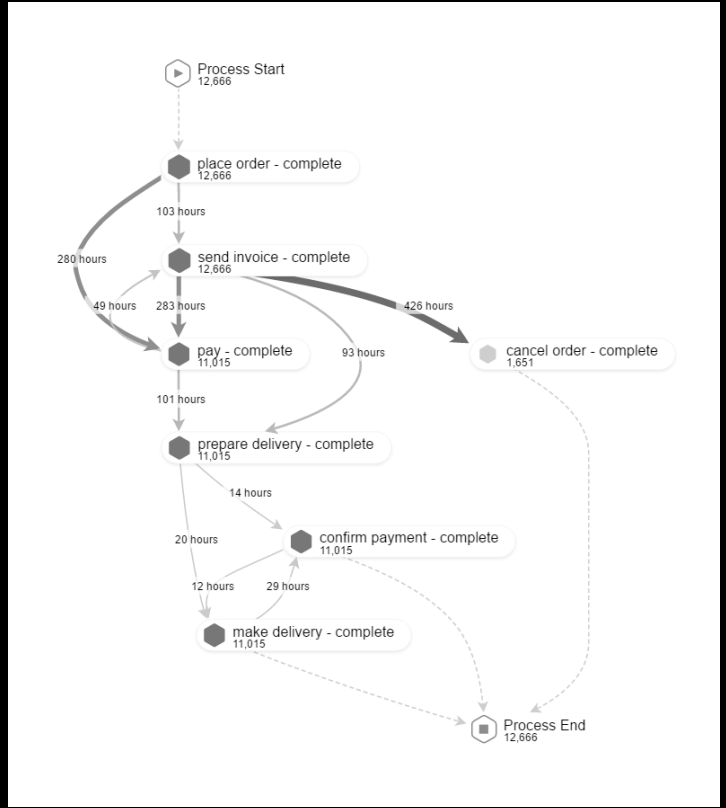


© Wil van der Aalst (use only with permission & acknowledgements)

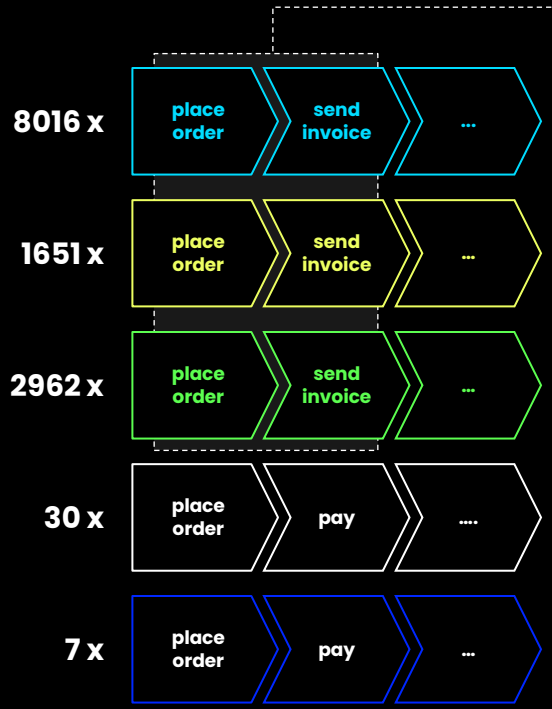
frequency

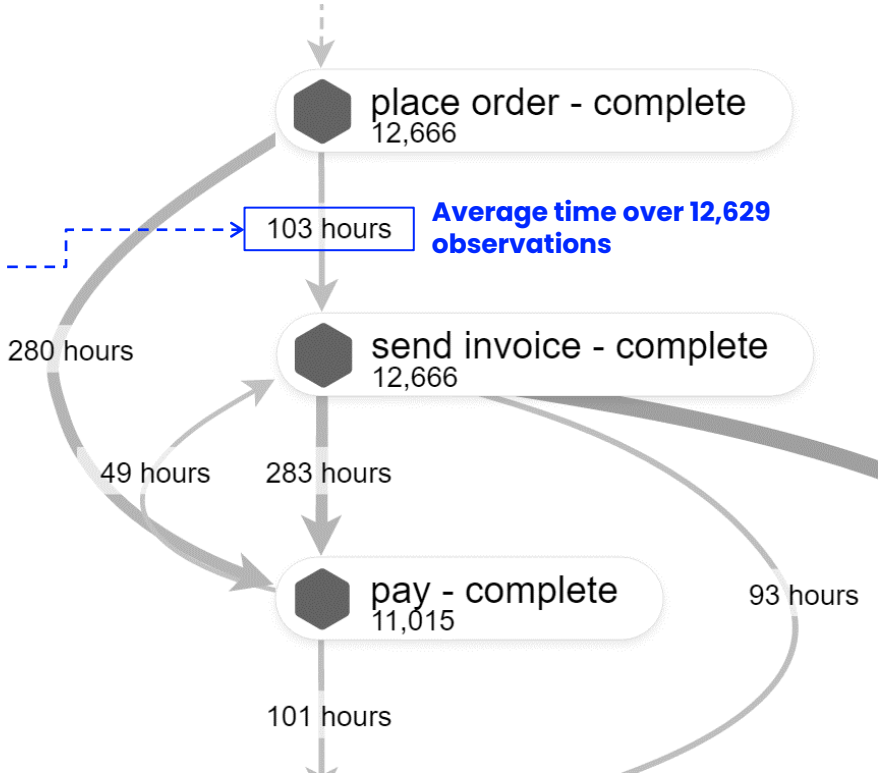
time



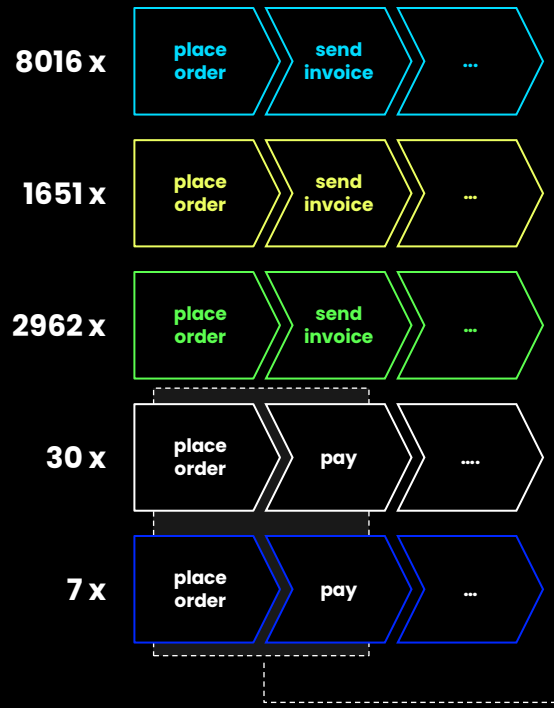


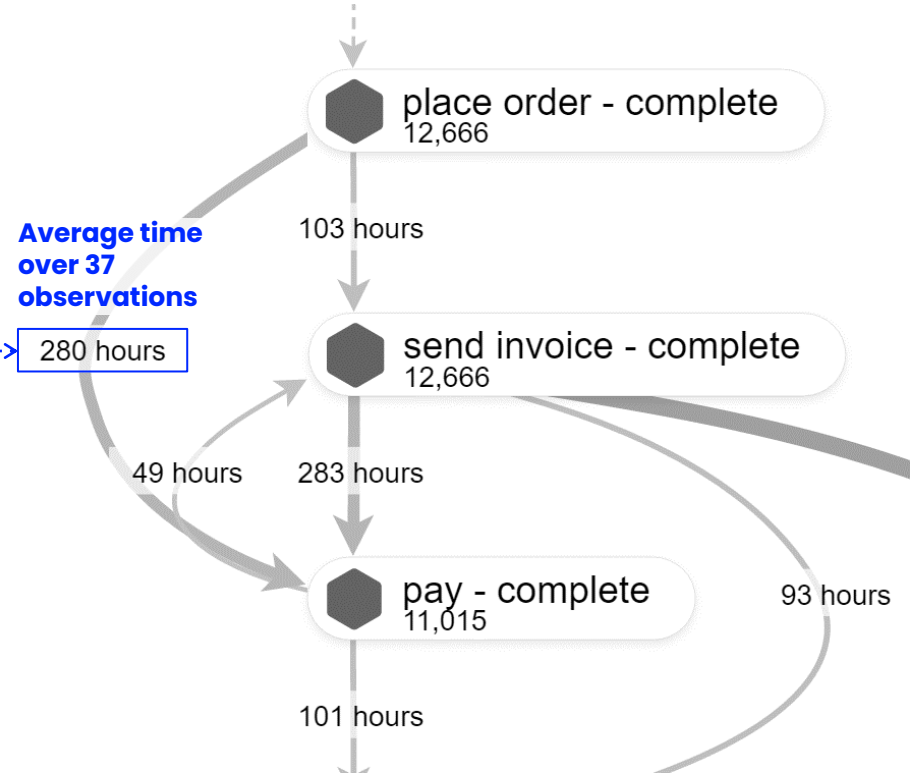
© Wil van der Aalst (use only with permission & acknowledgements)





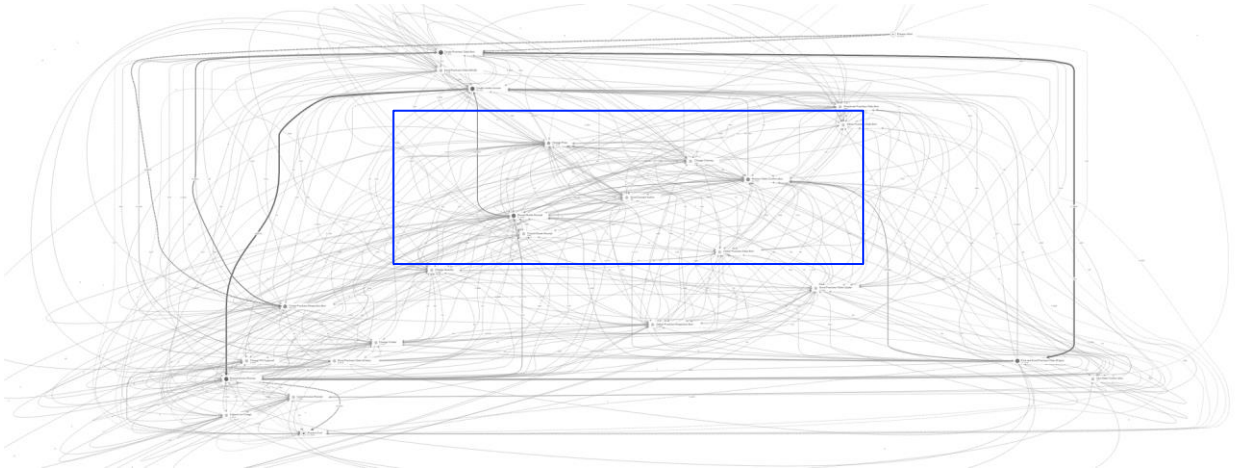
© Wil van der Aalst (use only with permission & acknowledgements)





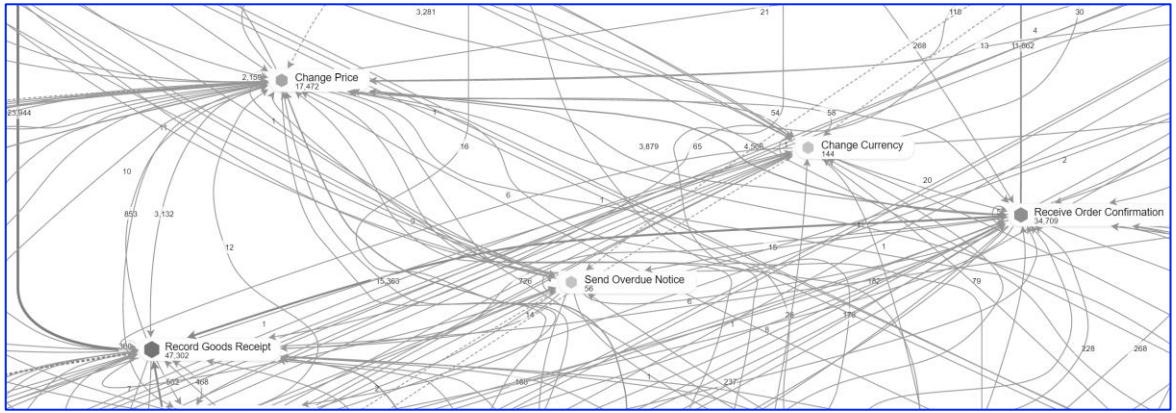
© Wil van der Aalst (use only with permission & acknowledgements)

What if?



© Wil van der Aalst (use only with permission & acknowledgements)

What if?



© Wil van der Aalst (use only with permission & acknowledgements)

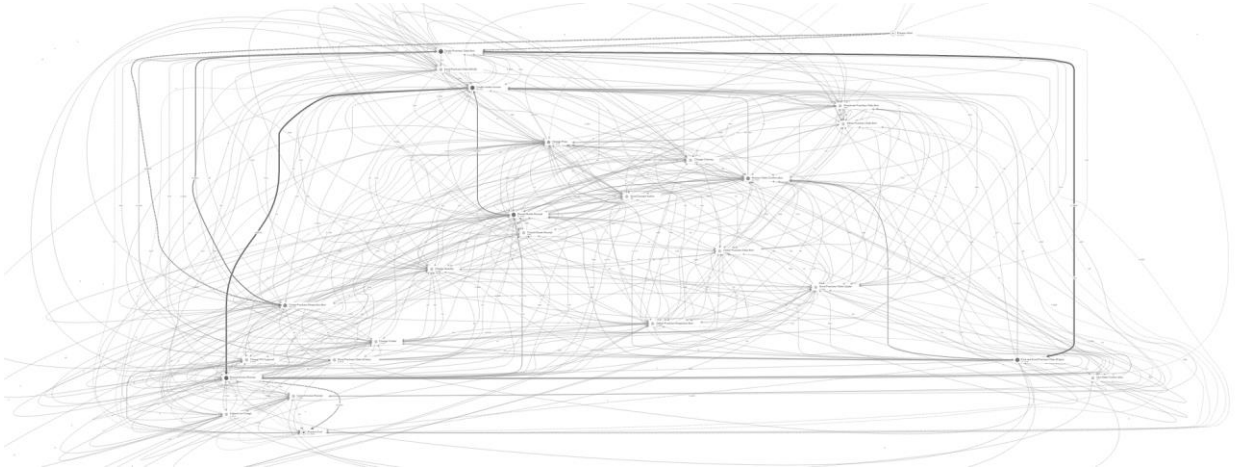
Three types of filtering to simplify DFGs

1 Activity-based

2 Arc-based

3 Variant-based

Use with care!

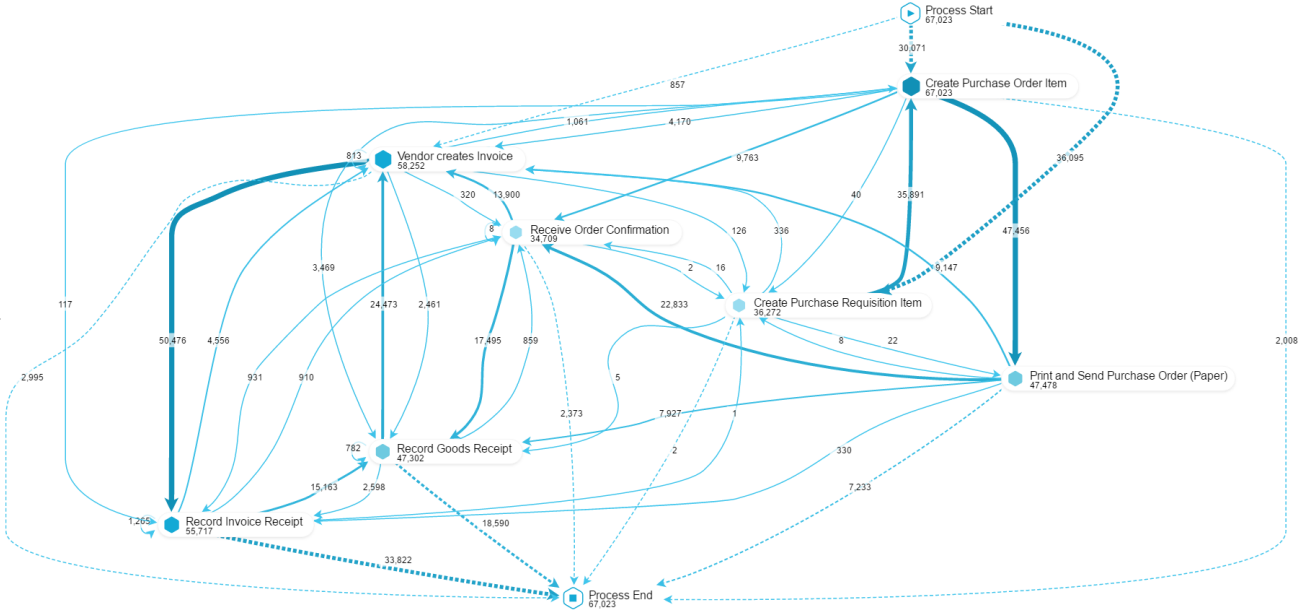


© Wil van der Aalst (use only with permission & acknowledgements)

Arc-based filtering of DFGs

Top 7 activities (>30,000 times)

To be safe: Remove activities from the event log!



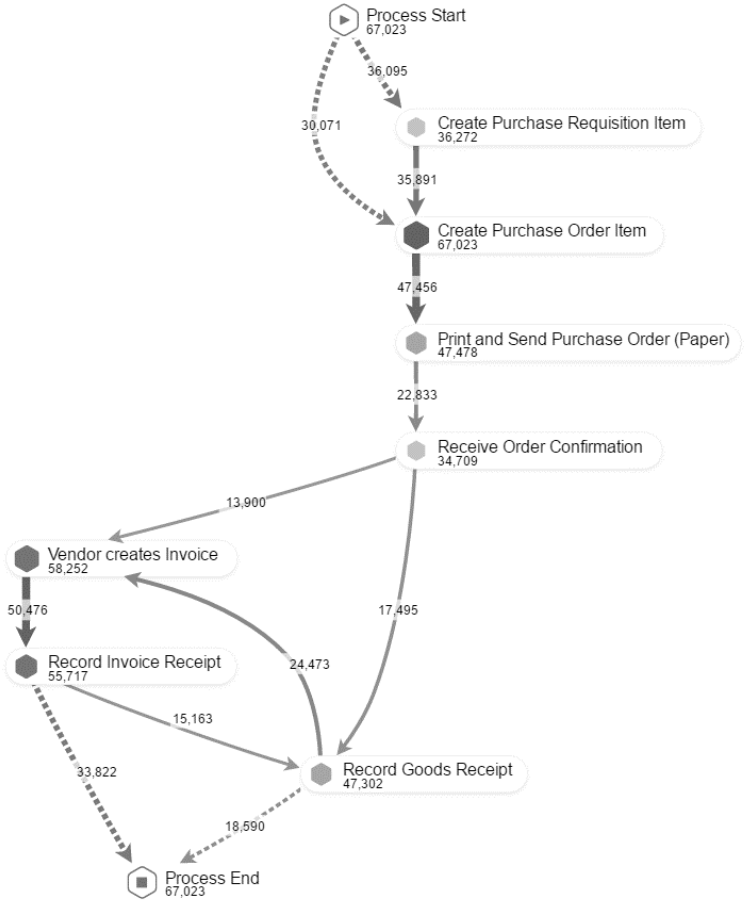
© Wil van der Aalst (use only with permission & acknowledgements)

Activity-based filtering of DFGs

Top 7 activities (>30,000 times) and arcs with frequency above 10,000

Important

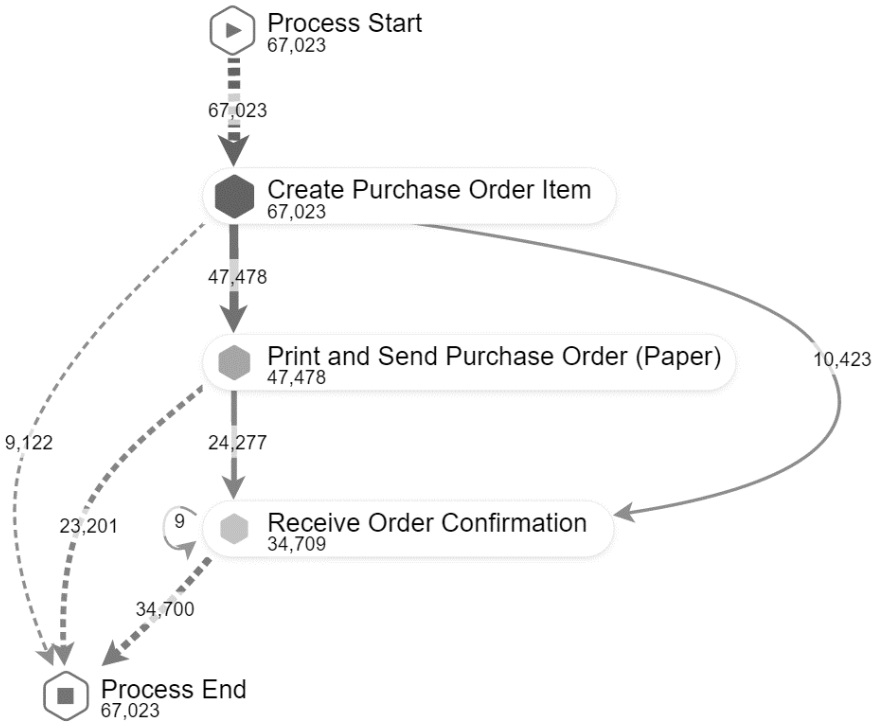
- **If you do not remove activities, they will influence the result.**
- **Hidden activities influence directly-follows frequency.**
- **Numbers do not need to add up after removing arcs.**



© Wil van der Aalst (use only with permission & acknowledgements)

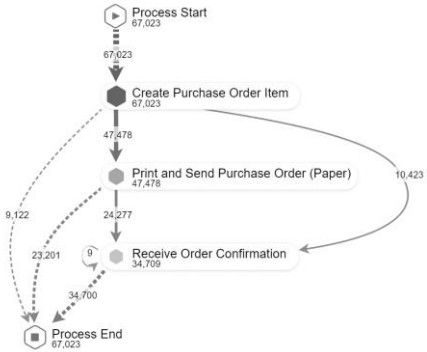
Considering just 3 activities

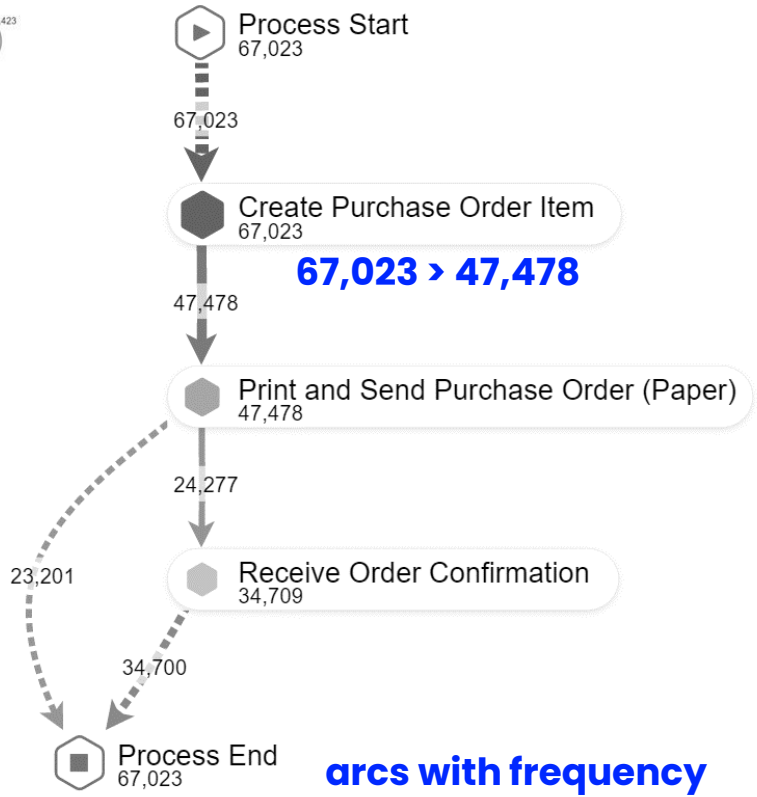
67,023 = 47,478 + 10,423 + 9,122



© Wil van der Aalst (use only with permission & acknowledgements)

Considering just 3 activities

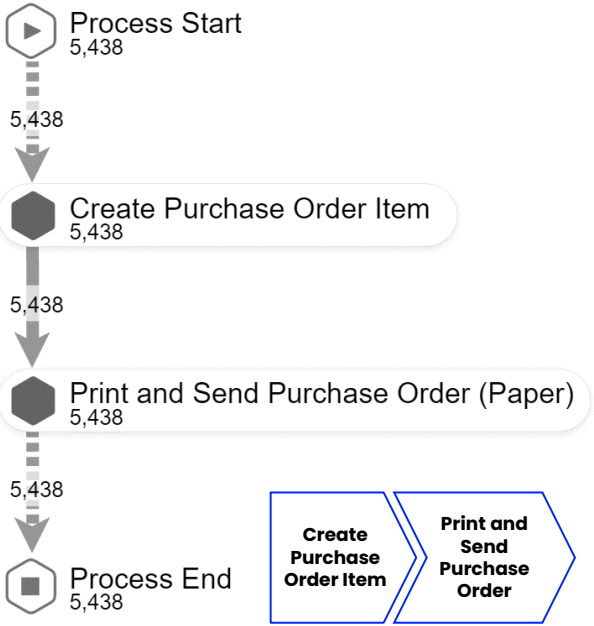




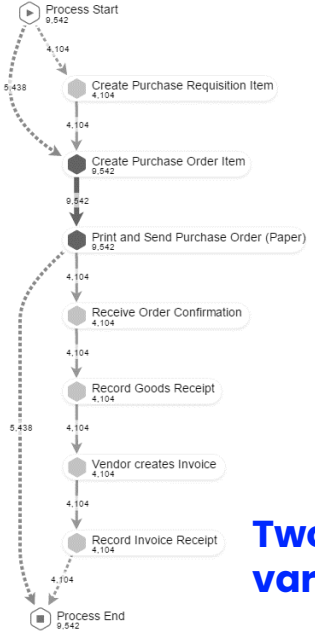
© Wil van der Aalst (use only with permission & acknowledgements)

Variant-based filtering of DFGs

8% of cases



14% of cases

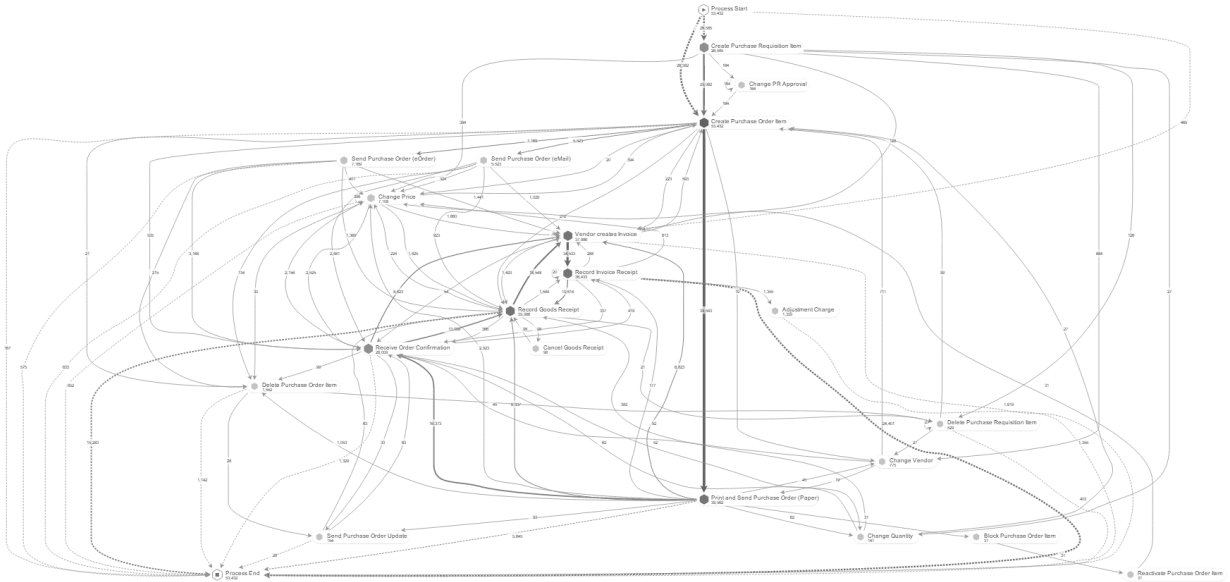


Two variants

© Wil van der Aalst (use only with permission & acknowledgements)

Variant-based filtering of DFGs

80% of cases described by **4.3%** of all variants!



298
of 6,897 variants

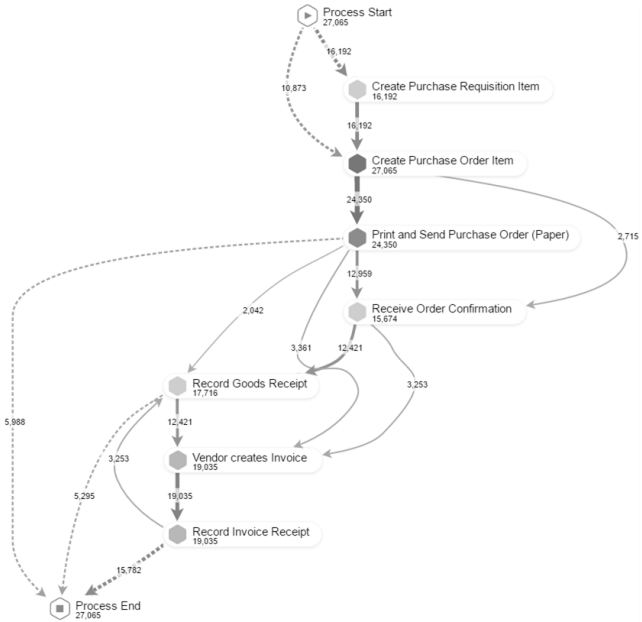
80%
of cases covered

© Wil van der Aalst (use only with permission & acknowledgements)

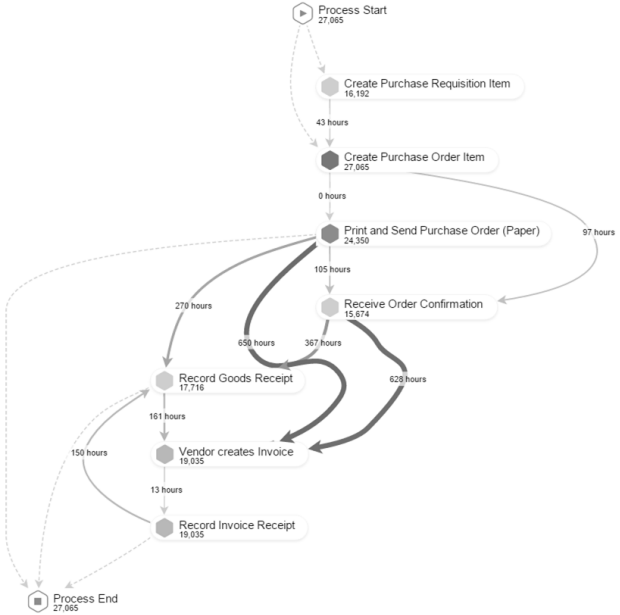
Different filters can be combined!

DFG based on top 7 activities and top 7 variants

frequency view

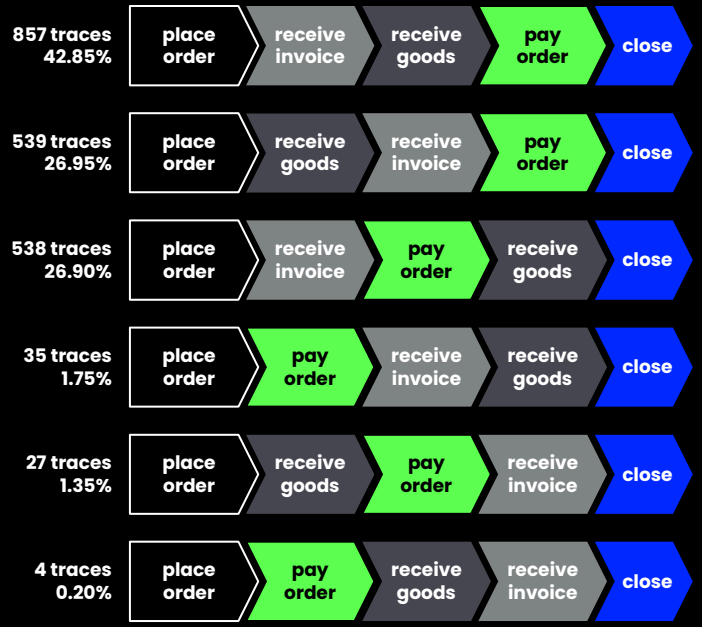


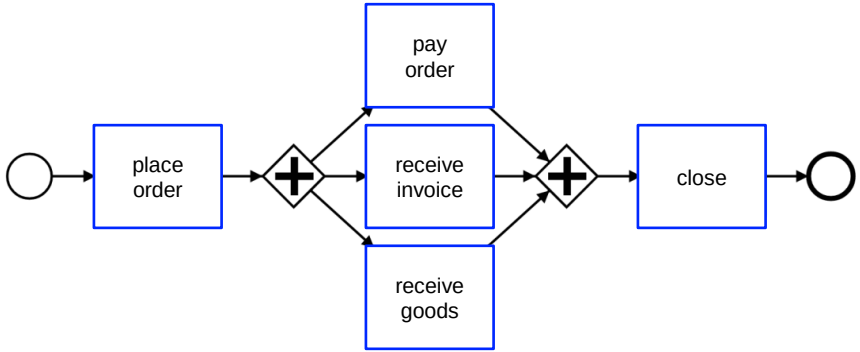
time view



© Wil van der Aalst (use only with permission & acknowledgements)

DFGs cannot capture concurrency



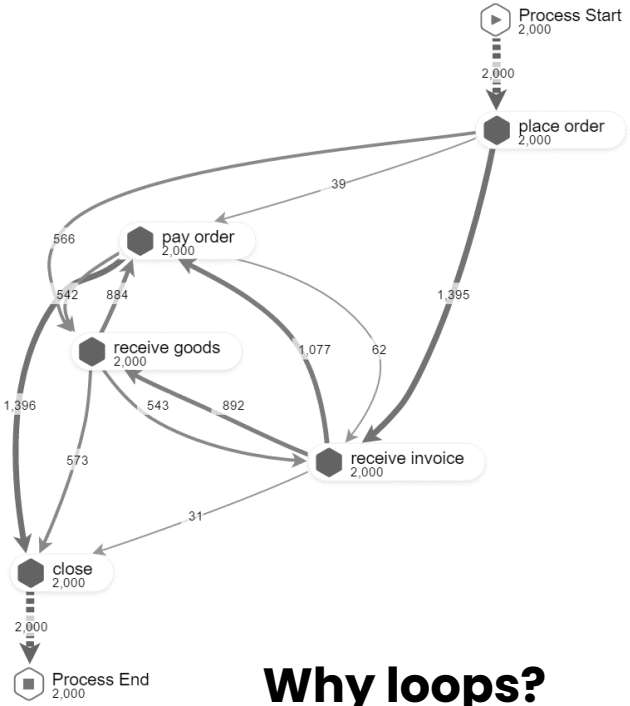


- **2,000 procurement orders**
- **5 activities**
- **6 unique traces**
- **Unbalanced**

© Wil van der Aalst (use only with permission & acknowledgements)

DFGs cannot capture concurrency

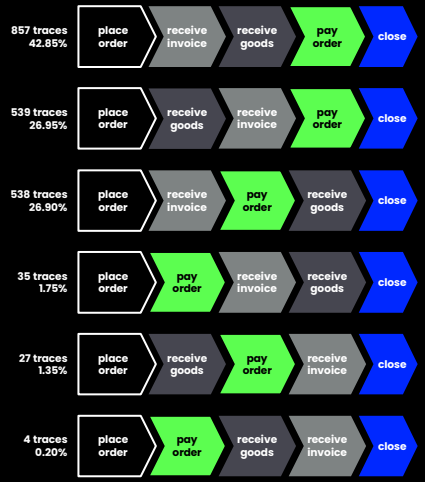
857 traces 42.85%	place order	receive invoice	receive goods	pay order	close
539 traces 26.95%	place order	receive goods	receive invoice	pay order	close
538 traces 26.90%	place order	receive invoice	pay order	receive goods	close
35 traces 1.75%	place order	pay order	receive invoice	receive goods	close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



Why loops?

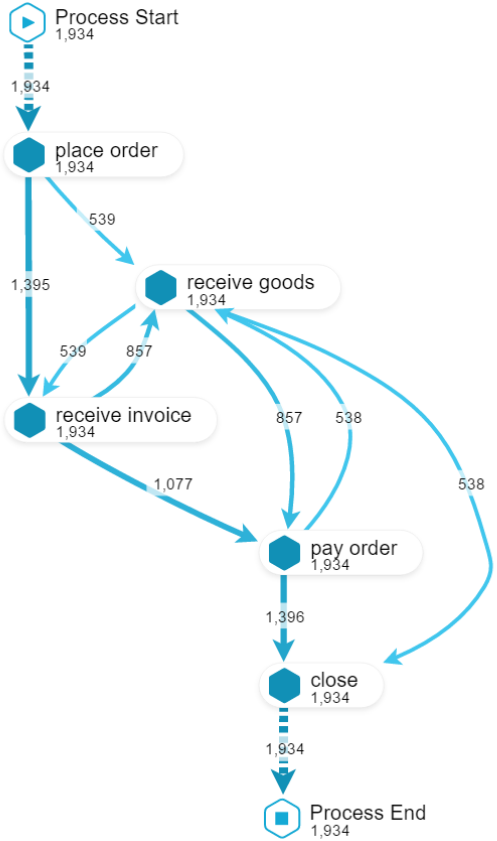
© Wil van der Aalst (use only with permission & acknowledgements)

Filtering does not help



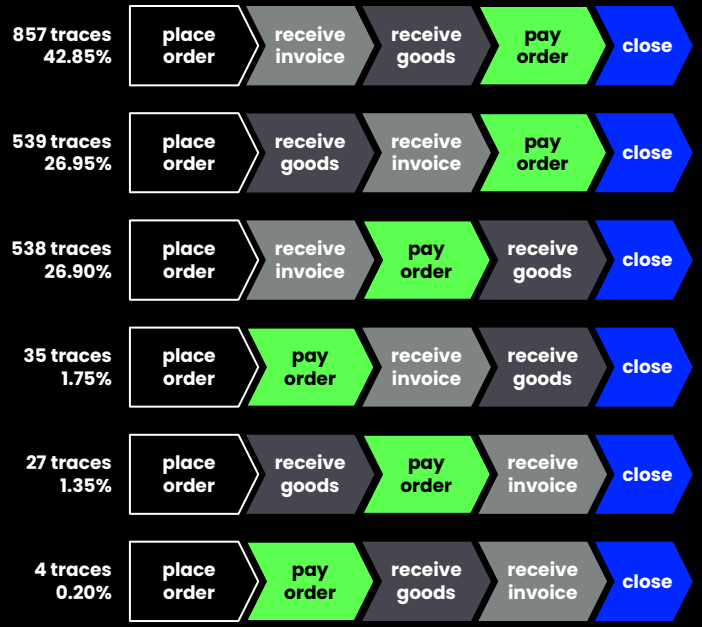
Arc-based filtering (>750)

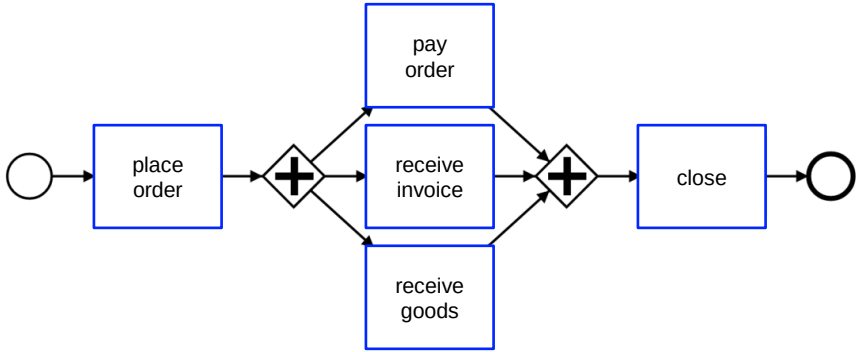
Variant-based filtering (top 3 variants)



© Wil van der Aalst (use only with permission & acknowledgements)

Inductive mining finds the right model





BPMN notation
(Petri nets and process trees under the hood)

© Wil van der Aalst (use only with permission & acknowledgements)

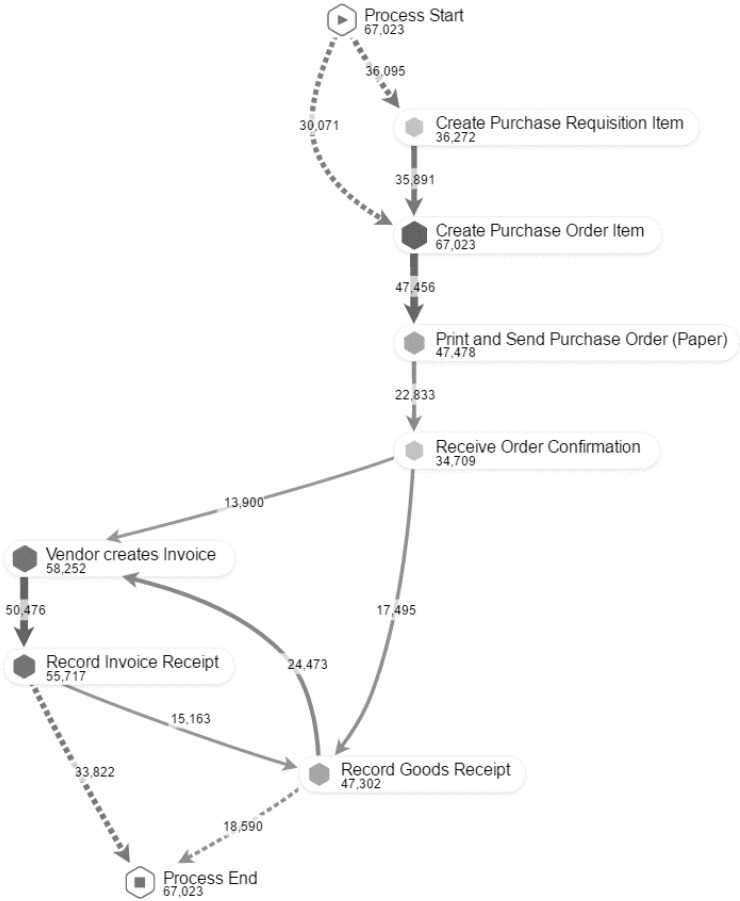
Directly-Follows Graphs (DFGs)

+

- Easy to create and scalable.
- Show paths, frequencies, and times.
- Can be simplified easily using filters.

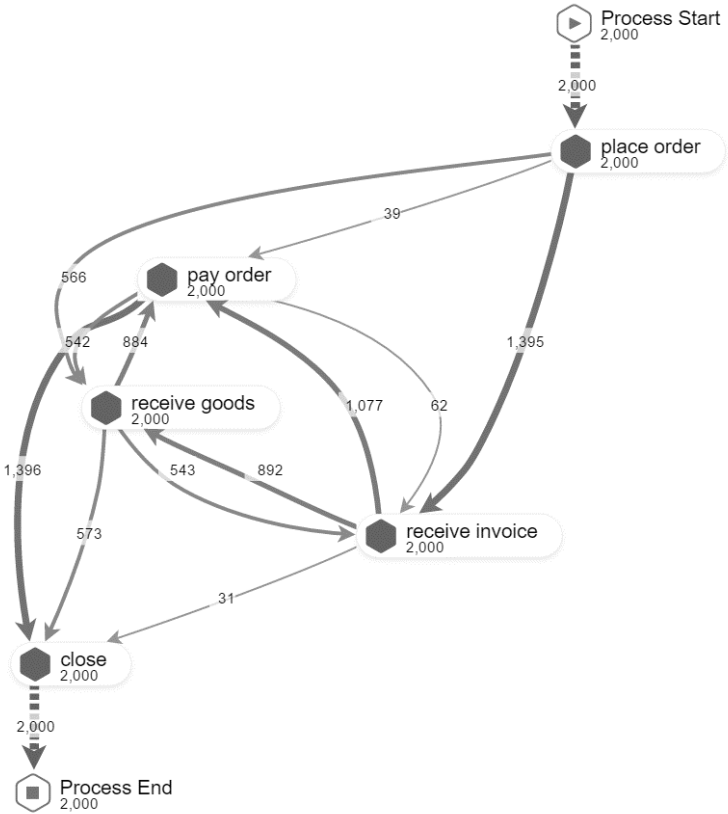
—

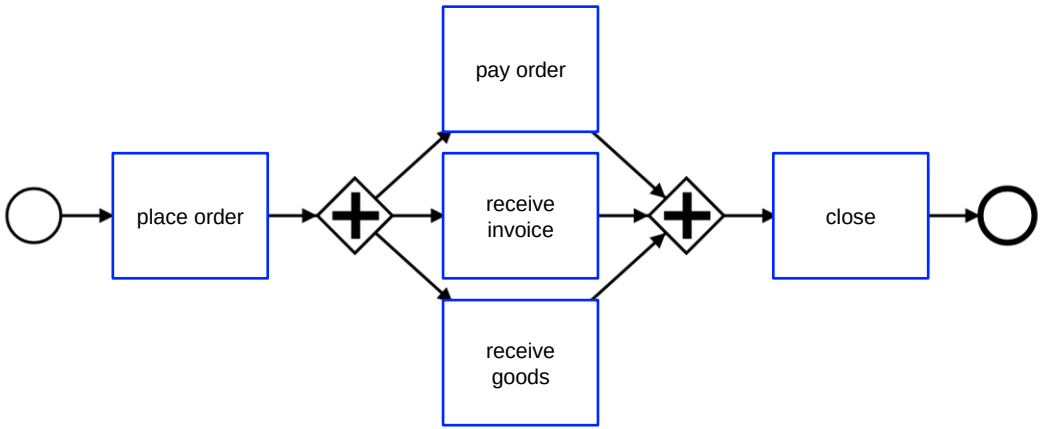
- Filtering needs to be done with the utmost care.
- DFGs cannot capture concurrency.



© Wil van der Aalst (use only with permission & acknowledgements)

Next lecture: Improved process discovery





© Wil van der Aalst (use only with permission & acknowledgements)