UNIVERSITY OF BIRMINGHAM

A Framework for the Analysis of Process Mining Algorithms

Phil Weber, Behzad Bordbar, Peter Tino
School of Computer Science

What is Process Mining?

Learning **models** of **business processes**, from the **log files** produced by their information systems:

People and systems interact according to some **underlying process**.
This may not be the same as the **intention**.
IT systems always **write log files.**
We can use this data to generate **useful information.**

Business Processes

Events, activities and how related:
To **achieve a business need or goal.**

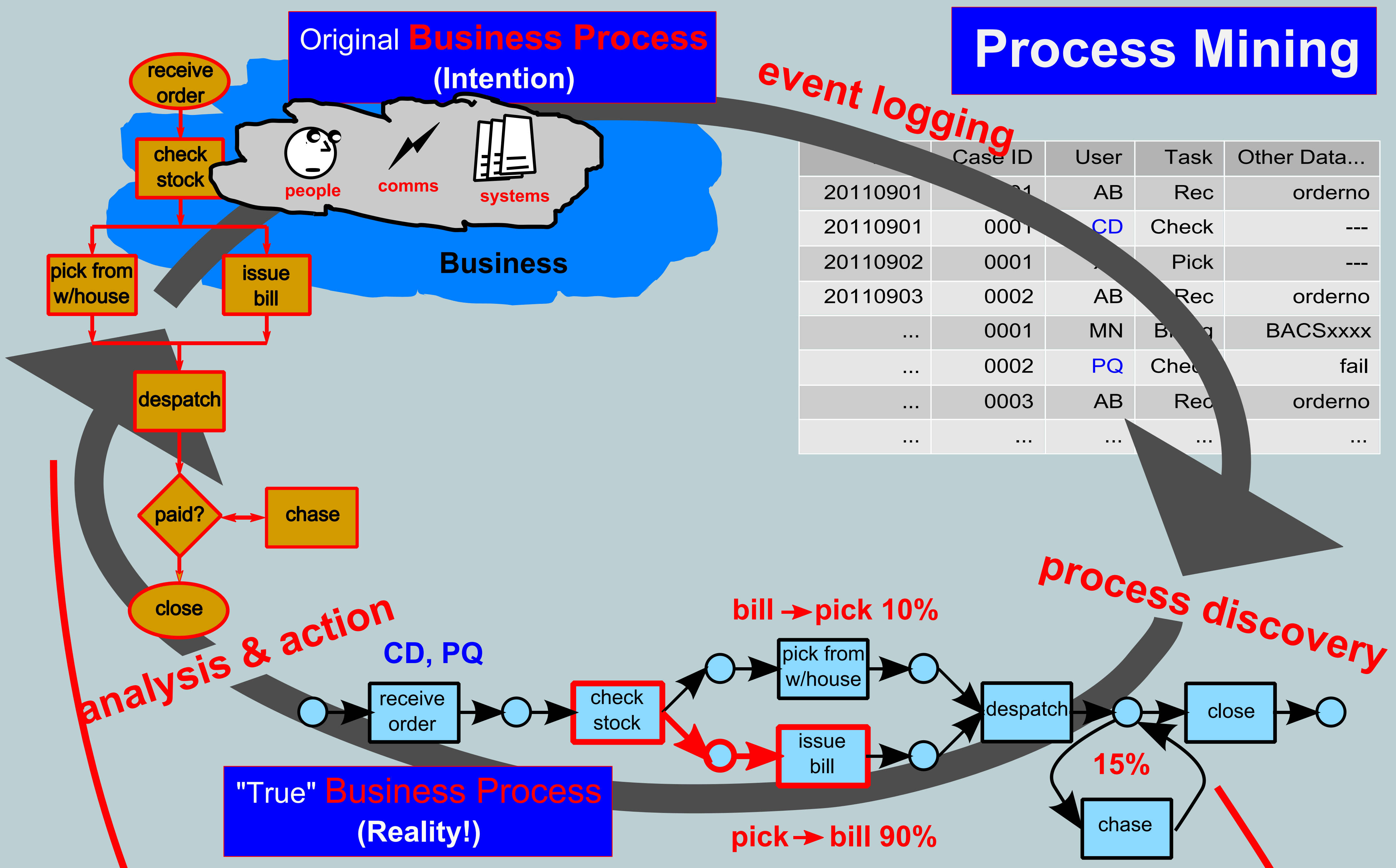
Used to control or dictate business operations:
invoicing, purchasing, web services, ...

Enforce business **rules**, **audit** requirements, ...
Improve **efficiency**, B2B, web services, ...

Process discovery algorithms attempt to recover **model** of the true **underlying process**.

Process conformance compares reality (mined model) with intention.

Process extension adds information on **decision rules**, **performance**, **planning** through **simulation**, “what-if?” modelling.



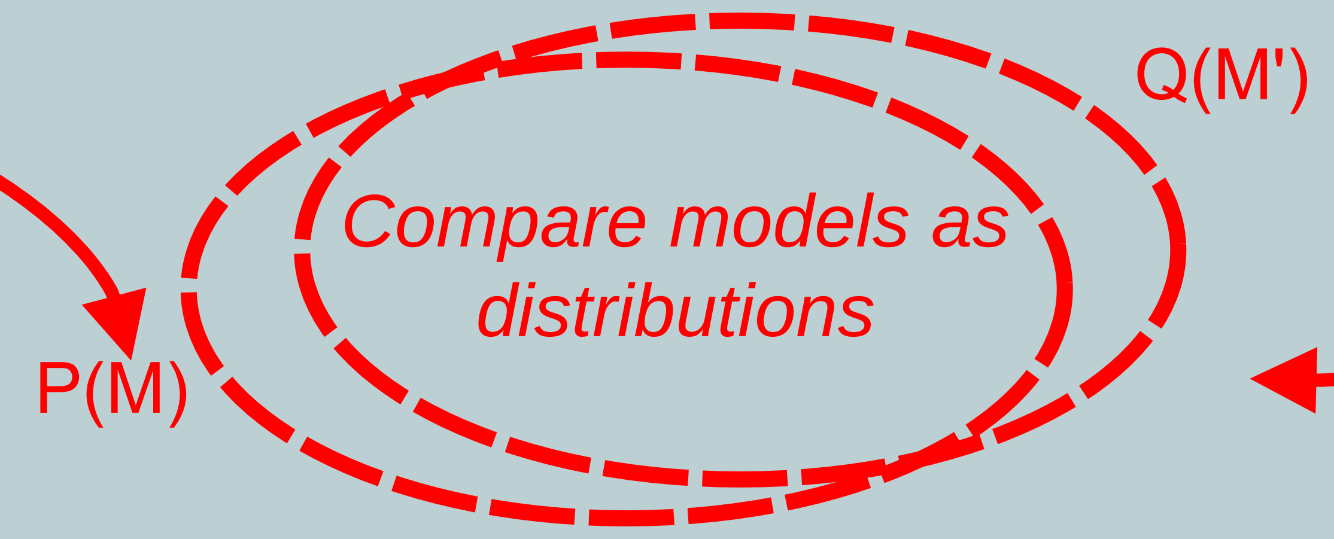
	Case ID	User	Task	Other Data...
20110901	0001	AB	Rec	orderno
20110901	0001	CD	Check	---
20110902	0001		Pick	---
20110903	0002	AB	Rec	orderno
...	0001	MN	Billing	BACSxxxx
...	0002	PQ	Check	fail
...	0003	AB	Rec	orderno
...	...	...	...	...

Many algorithms!
Many representations!

? *Which is best for my situation?*
? *How to compare?*

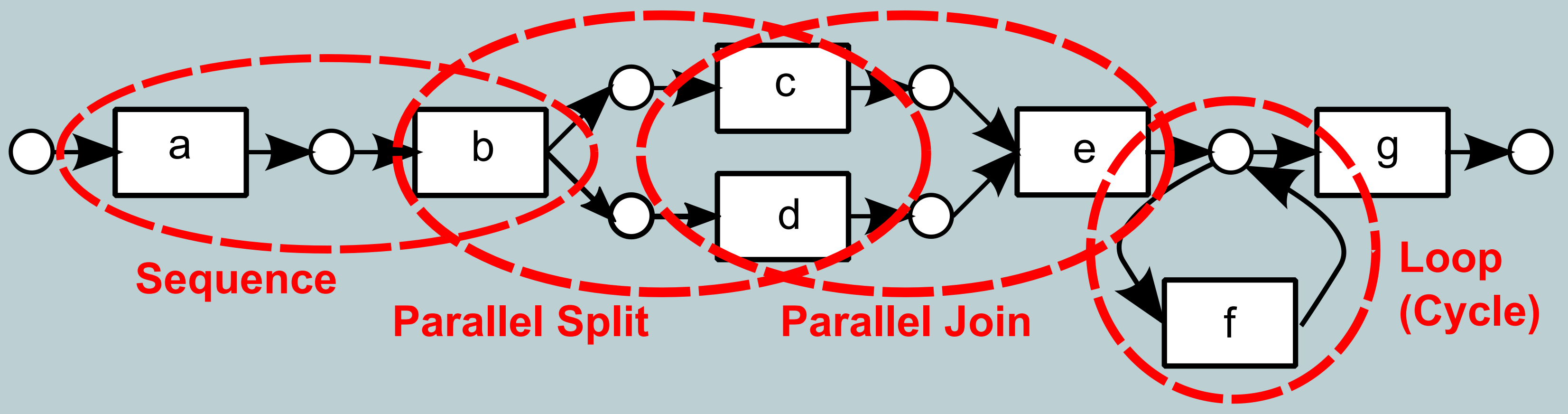
? *What are their learning properties?*
? *How much data do I need?*

A Probabilistic Framework



Abstract from the representation

1. business processes: probability distributions over strings of symbols.
2. underlying process *M* generates traces according to distribution *P(M)*,
3. algorithm learns model *M'* representing (different) distribution *Q(M')*,
4. distance e.g. *d[P(M),Q(M')]* between distributions measures convergence,
5. stochastic automata represent both process models and distributions.



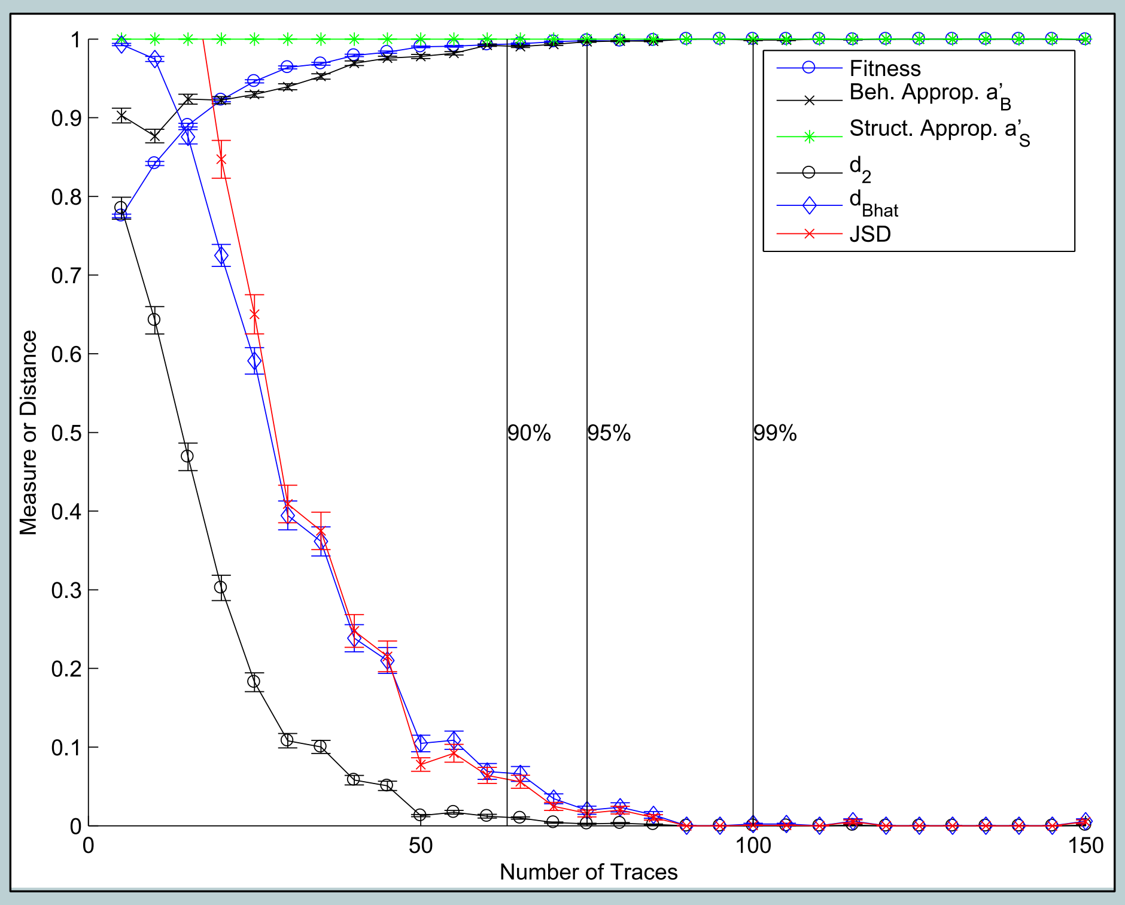
Analysis

1. identify basic structures in model (above),
2. model algorithms' behaviour in terms of
   - requirements for mining basic structures (above)
   - probability of achieving requirements – in terms of
   - probabilities of substrings and amount of data.

So What?

Rigorous foundation for generalising/abstracting models, modelling noise, comparing processes, ...
Practical applications include detecting change, online or real-time process mining. anomaly or error detection.

For Example



The **Alpha algorithm** uses relations between task pairs (string ‘***ab***’) to build a Petri net.

We determine probability of mining structures and model from the probability of ‘seeing’ these strings in the log.

The **Heuristics Miner** uses a metric, the ratio of how often a pair of tasks occurs in either order, to construct a directed graph.

A simple metric, following a complex distribution. By modelling it we can give bounds for how many traces are needed for the ‘correct’ metrics to be maximised, so mining the correct model.

Next question: how to compare these algorithms?

www.bham.ac.uk

References
P. Weber, B. Bordbar and P. Tiňo: *A Framework for the Analysis of Process Mining Algorithms. Accepted: IEEE Trans. Systems, Man and Cybernetics, 2012.*