云南大学软件学院 实验报告

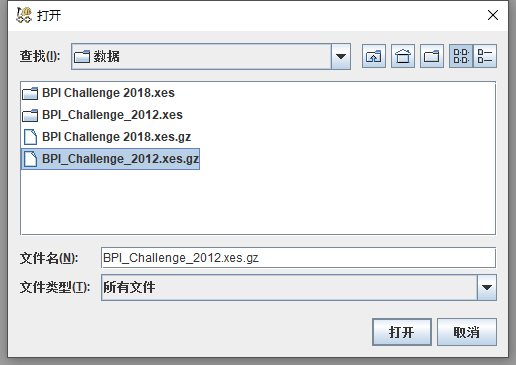
课程： 软件过程与管理 学期： 2024 - 2025学年 第一学期 任课教师： 莫启

专业： 软件工程 学号： 20221120044 姓名： 唐嘉骏 成绩：

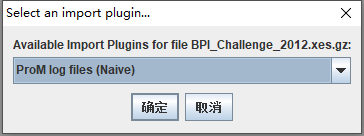
**实验4**

利用ProM对实验3所下载的事件日志进行挖掘，挖掘算法至少选择3种以上，分析挖掘出的结果间的差异性,并分析其原因。

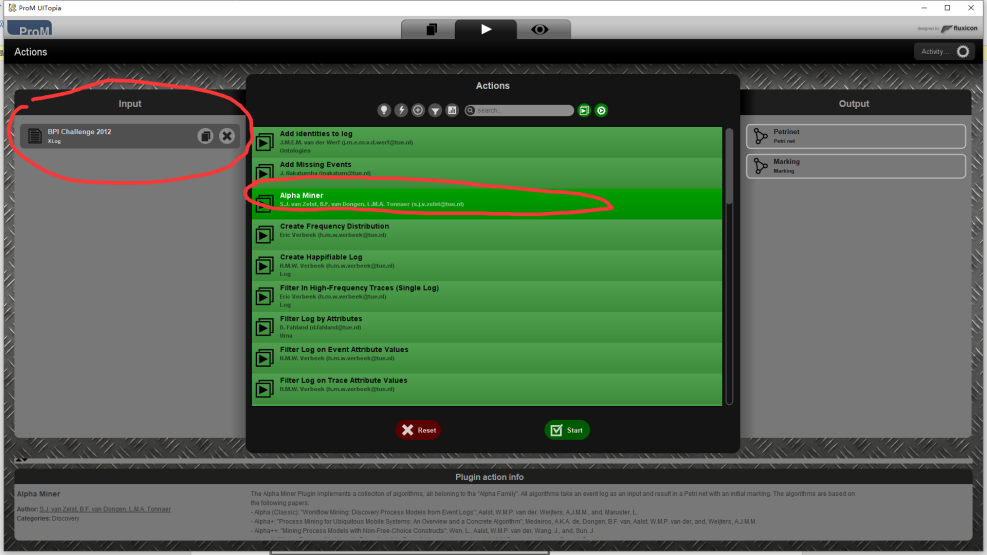
（1）导入数据，数据可选BPI\_Challenge任意版本



（2）选择Naive



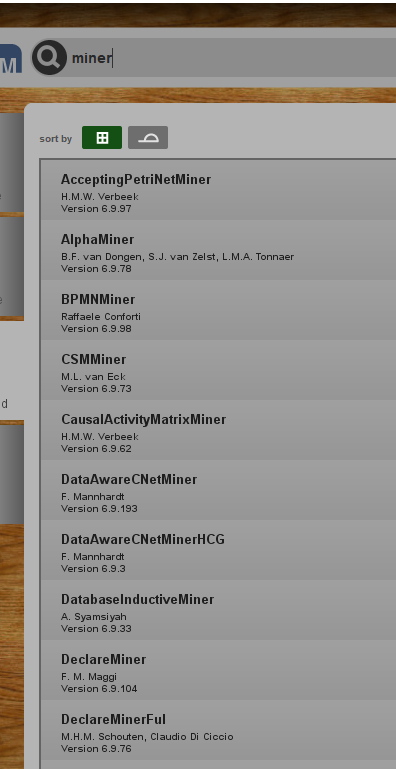
（3）选择已导入的源数据，再选择一种数据挖掘方法，输出结果（此处为Alpha算法）。



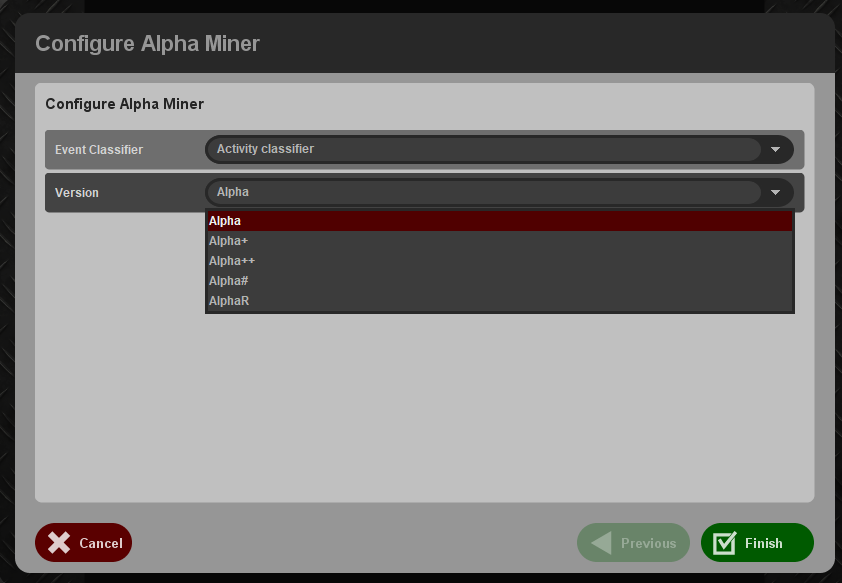
（4）注意，如果在此处查询不到要使用的挖掘算法，先进入包管理器，搜索后下载。



包管理器



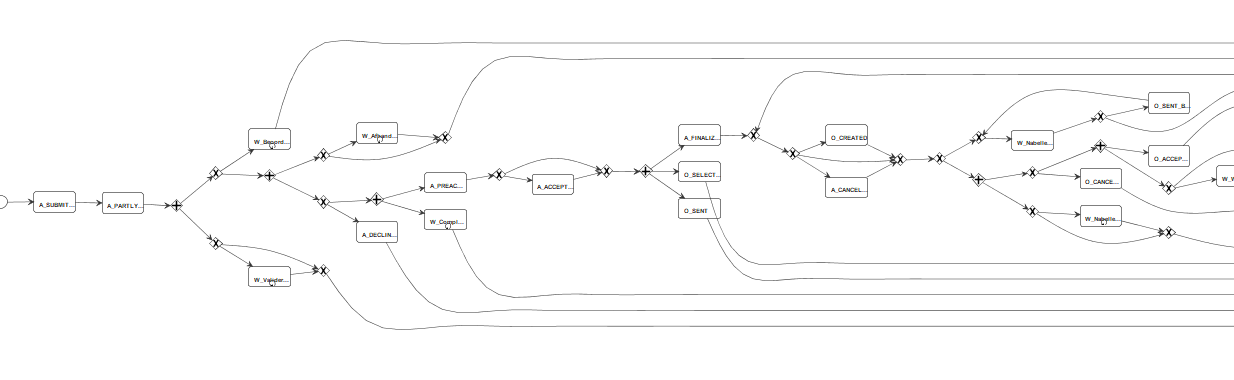
（5）选择生成的配置选项。



（5）可选择导出到文件。



α算法

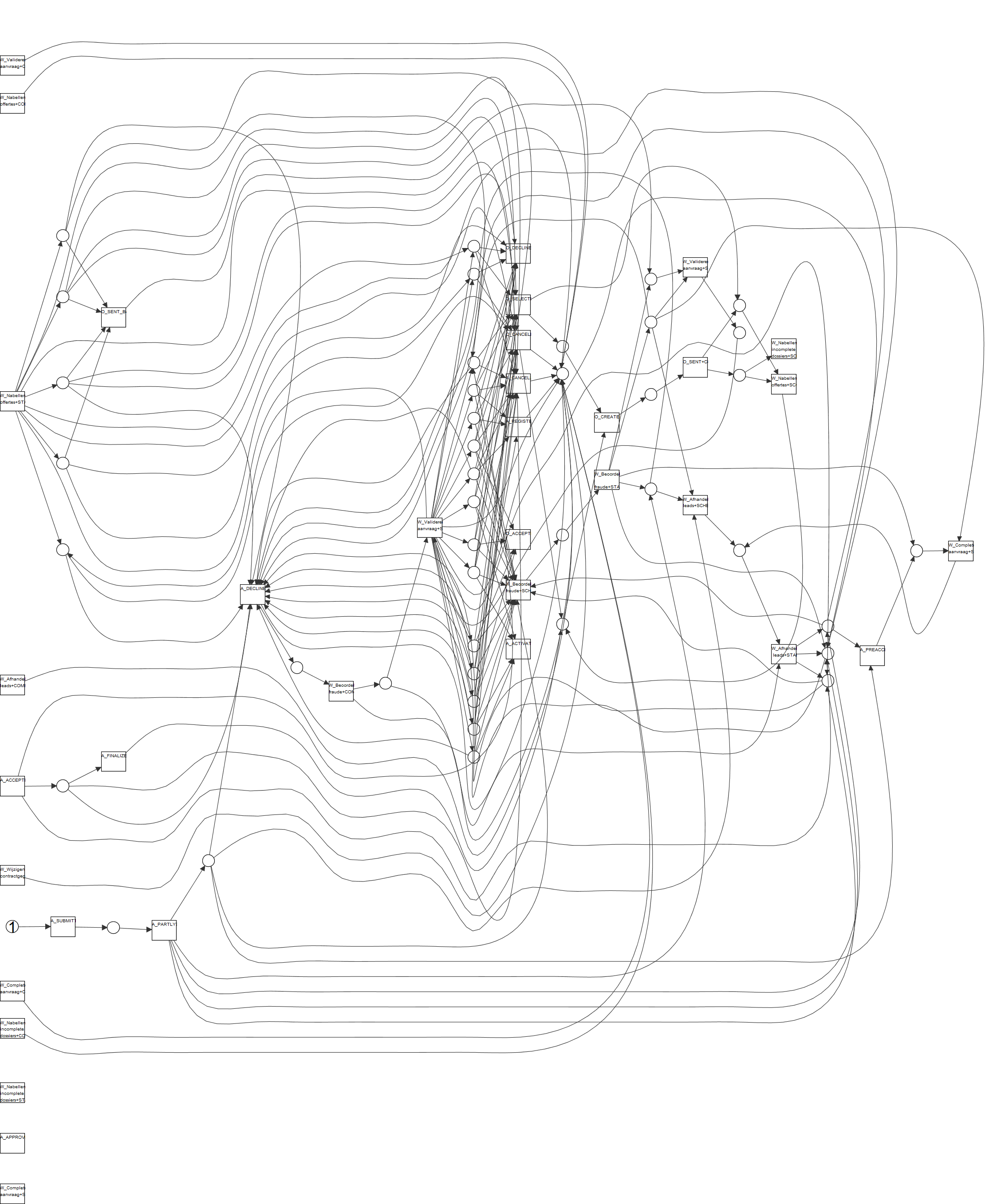


BPMN

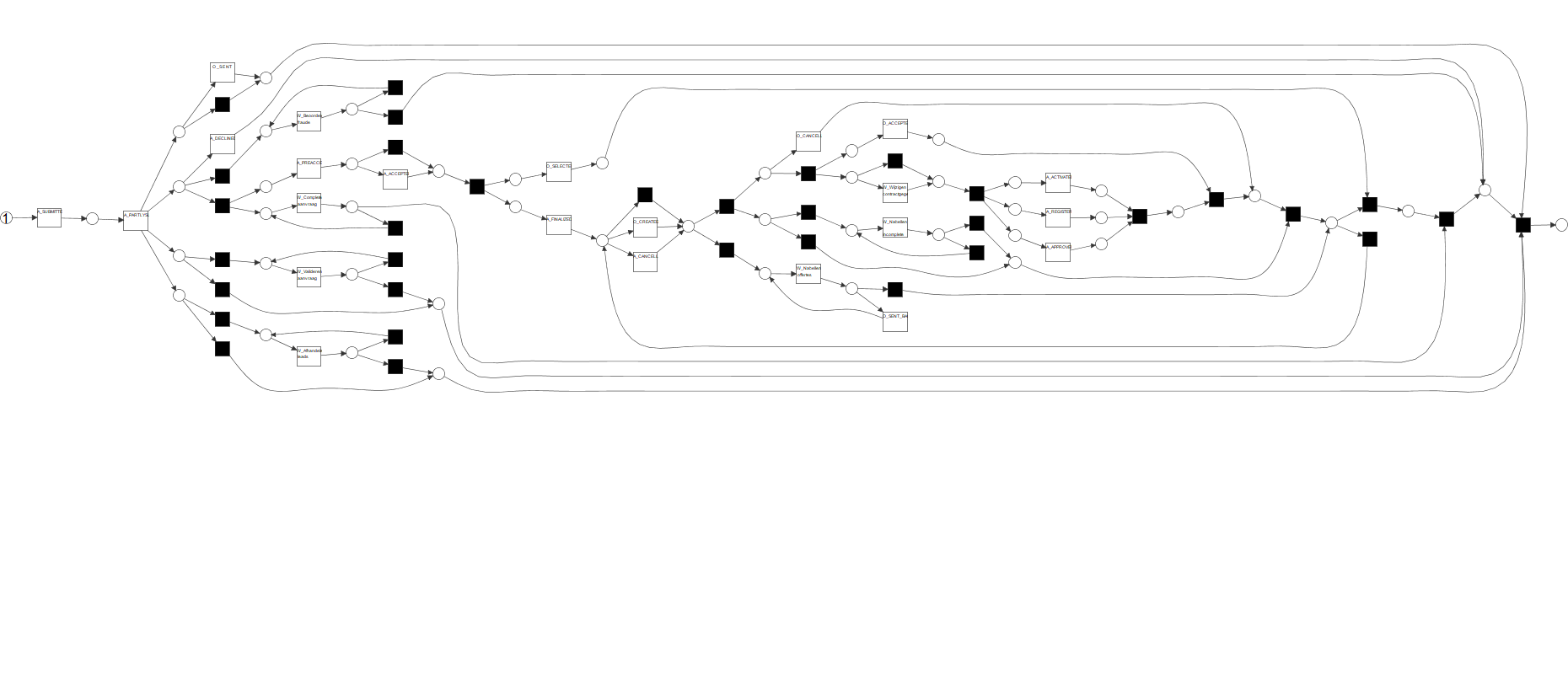
1. 重复以上步骤，选择至少三种挖掘算法生成结果，根据挖掘算法的特征，分析三种挖掘算法之间的差异，以及为什么会造成挖掘结果的差异。

**三种算法：**

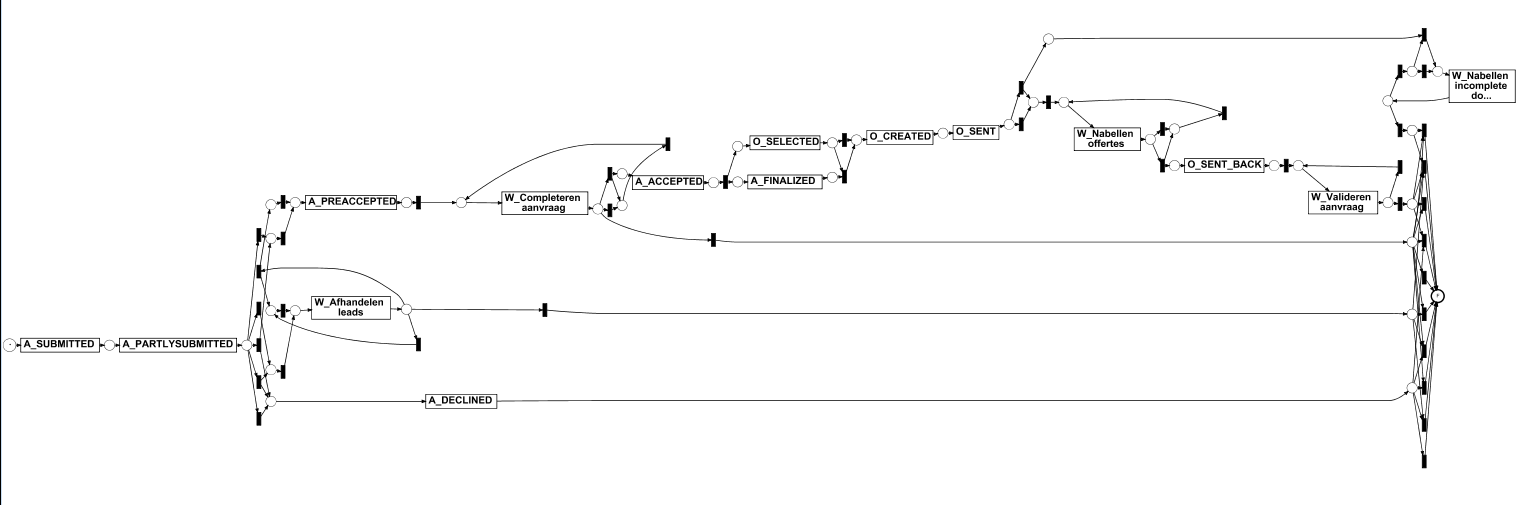
1. α算法：



1. Inductive Miner-InFrequency算法



1. Heuristic Miner算法



**特征：**

1. α 算法的核心思想是通过分析事件日志中的直接因果关系、并行关系和选择关系，生成一个过程模型。利用频繁出现的顺序关系构建流程。其优点是理论基础清晰，能够准确地挖掘出流程中的结构性关系；对于简单的行为日志数据比如无噪声的理想数据，可以很好地生成模型。但是对噪声和复杂流程结构较为敏感，仅适用于理想、完整的事件日志，可能无法处理实际中不完全的数据。
2. Inductive Miner-InFrequency 算法基于递归的方式，将事件日志分解为更小的子日志，生成符合行为的数据模型。在 InFrequency 模式下，允许不常见的行为模式被忽略，以提高模型的通用性和可理解性。

优点是通过分解策略，使算法对噪声具有更好的鲁棒性且生成的模型通常是结构化和无陷阱的。但是对于频率非常低但具有业务意义的行为模式可能会被忽略并且对复杂日志数据的分解过程需要精细的调优，可能影响效率。

1. Heuristic Miner 算法基于频率的启发式规则，通过挖掘活动之间的直接依赖关系图生成过程模型。优点是对噪声和不完整事件日志的鲁棒性较高，能够挖掘出带有概率性、频率信息的流程模型，更贴近实际业务需求。但是模型的精确度可能较低，可能遗漏或简化低频路径。并且过多依赖事件日志的频率特征，可能导致对偶发行为的处理能力不足。

**差异：**

1. 第一个图可以看出模型中有大量的交叉和并发路径，线条密集，结构复杂。这和α算法倾向于严格遵循日志中直接因果关系，但不区分噪声和实际行为，容易生成过度复杂的模型的特征相符。这种模型不适合直接用于业务分析。
2. 第二个图可以看出模型呈现明显的分层结构，并对分支进行了较好处理。这是Inductive Miner 通过递归分解事件日志，强调结构化和清晰的模型，但允许忽略低频行为导致的。这种模型保留了较多分支和结构化信息，能够反映出实际流程的复杂性，但部分细节可能被抽象化。
3. 第三个图可以看出模型看起来更加清晰，分支较少，流程图更贴近实际业务。这是因为Heuristic Miner 根据频率信息构建依赖关系图，忽略了低频或噪声数据，从而生成更简洁的模型。Heuristic Miner挖掘出来的模型更简洁，分支较少，非常适合用于高层次的业务概览，但可能忽略了一些关键行为或低频路径。

**为什么会产生差异：**

1. 算法对噪声的处理方式不同：

α 算法严格依赖日志中的直接因果关系，无法适应异常和噪声数据。

Inductive Miner-InFrequency通过递归分解和频率过滤，忽略低频噪声行为。

Heuristic Miner基于频率规则，能够有效处理噪声数据，但可能忽略低频行为。

1. 数据完整性的要求不同：

α 算法需要完整的事件日志，少量的缺失数据可能会导致模型无法生成。

Inductive Miner-InFrequency和Heuristic Miner允许数据不完整，但处理方式不同：前者通过分解策略处理，后者通过依赖图和频率信息补充。

1. 模型生成策略不同：

α 算法基于因果关系严格定义模型。

Inductive Miner-InFrequency递归分解日志以生成更可解释的模型。

Heuristic Miner利用依赖关系图构建贴近实际业务流程的模型。

1. 对频率信息的利用不同：

α 算法不使用频率信息，仅依赖结构。

Inductive Miner-InFrequency结合了结构与低频行为的过滤。

Heuristic Miner完全基于频率信息，生成的模型强调概率性。