

第十六届 D2 前端技术论坛

# 行为链路的识别、挖掘及诊断

演讲者：徐昊（南冬）



# Contents

## 目录

### 案例篇

01 背景知识

02 案例分析

---

### 技术篇

03 链路识别

04 链路可视化

05 链路诊断

06 未来展望



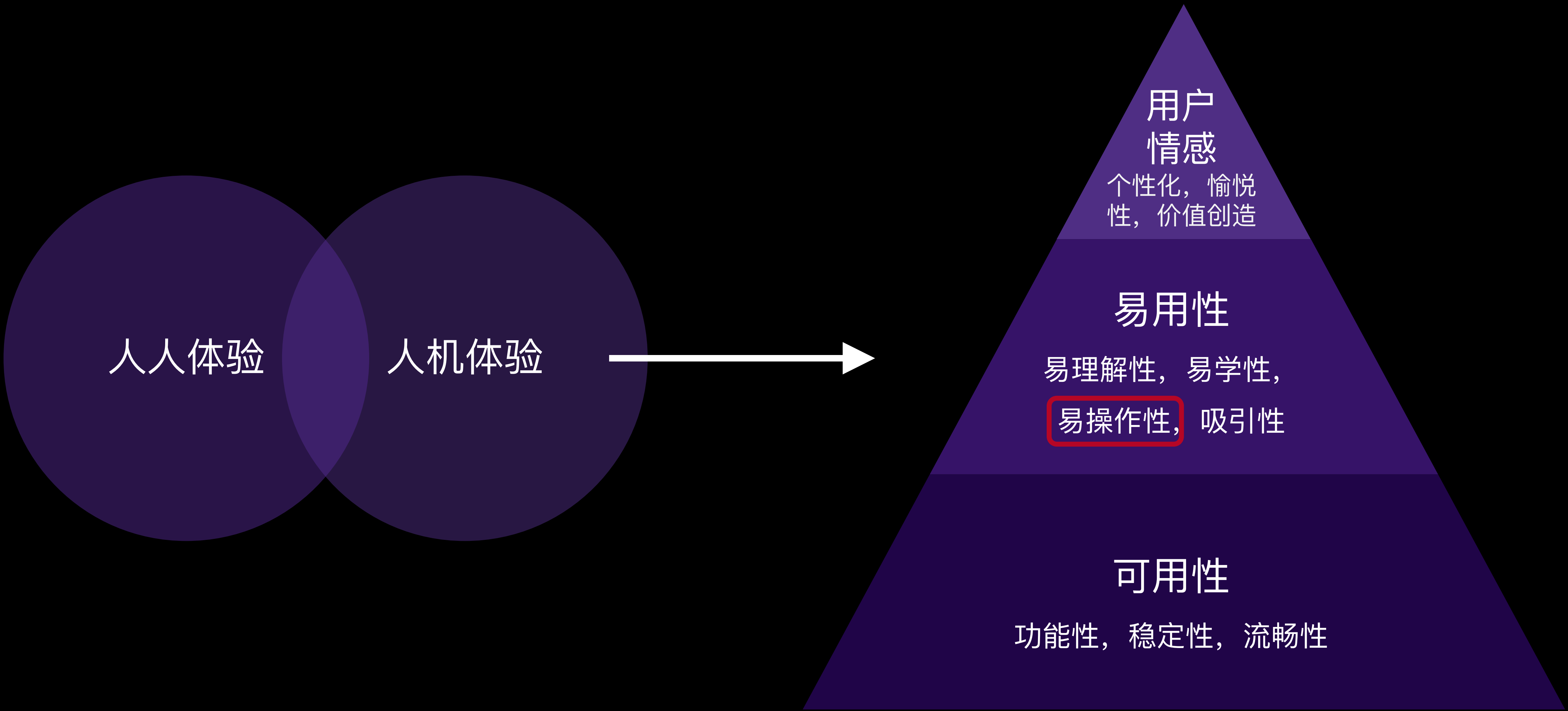


D2 前端技术论坛  
D2 FRONTEND TECHNOLOGY FORUM

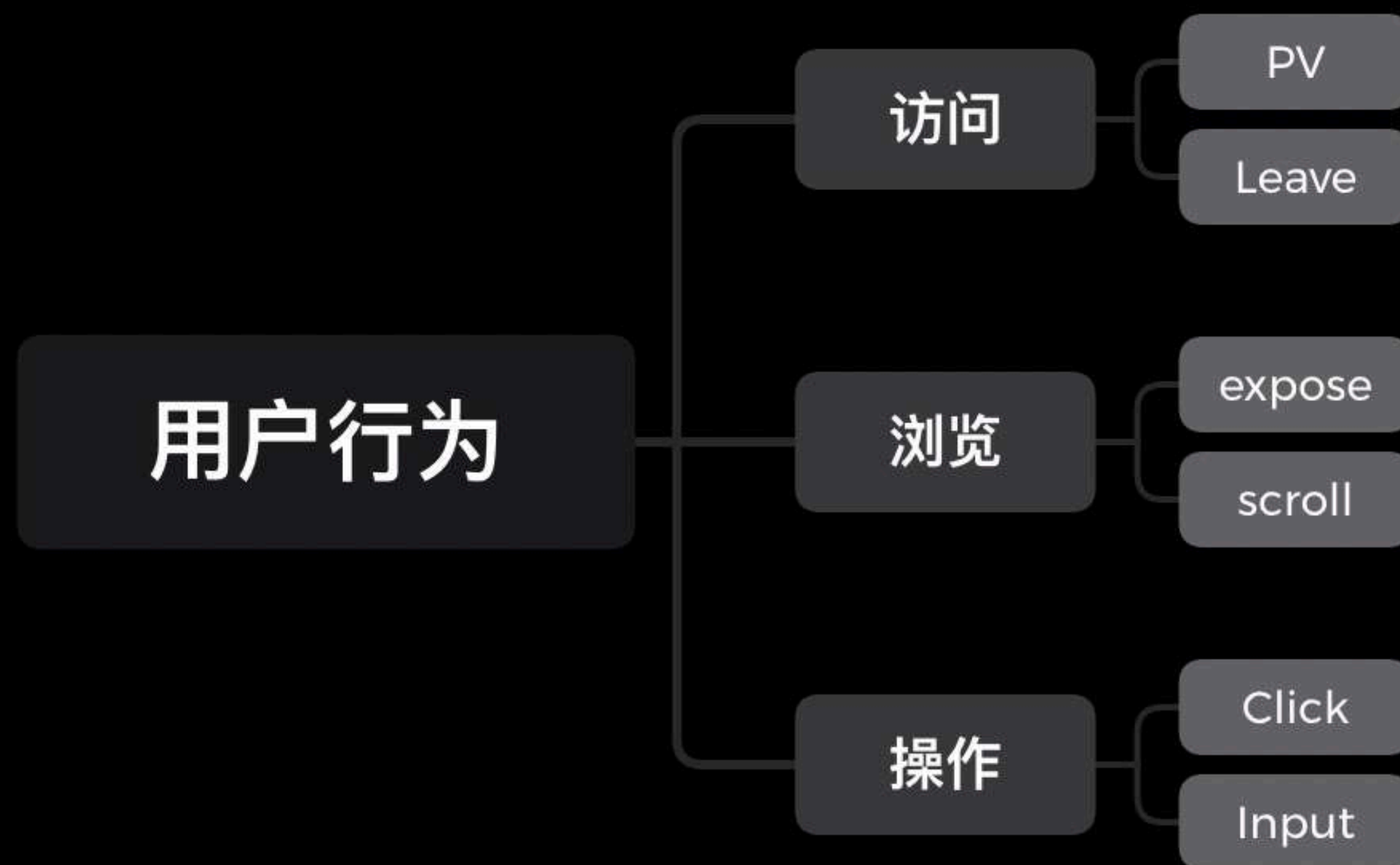
精心

01

# 背景知识







用户使用我们人机交互过程中产生的数据，称之为用户行为数据。

访问链路分析

➡

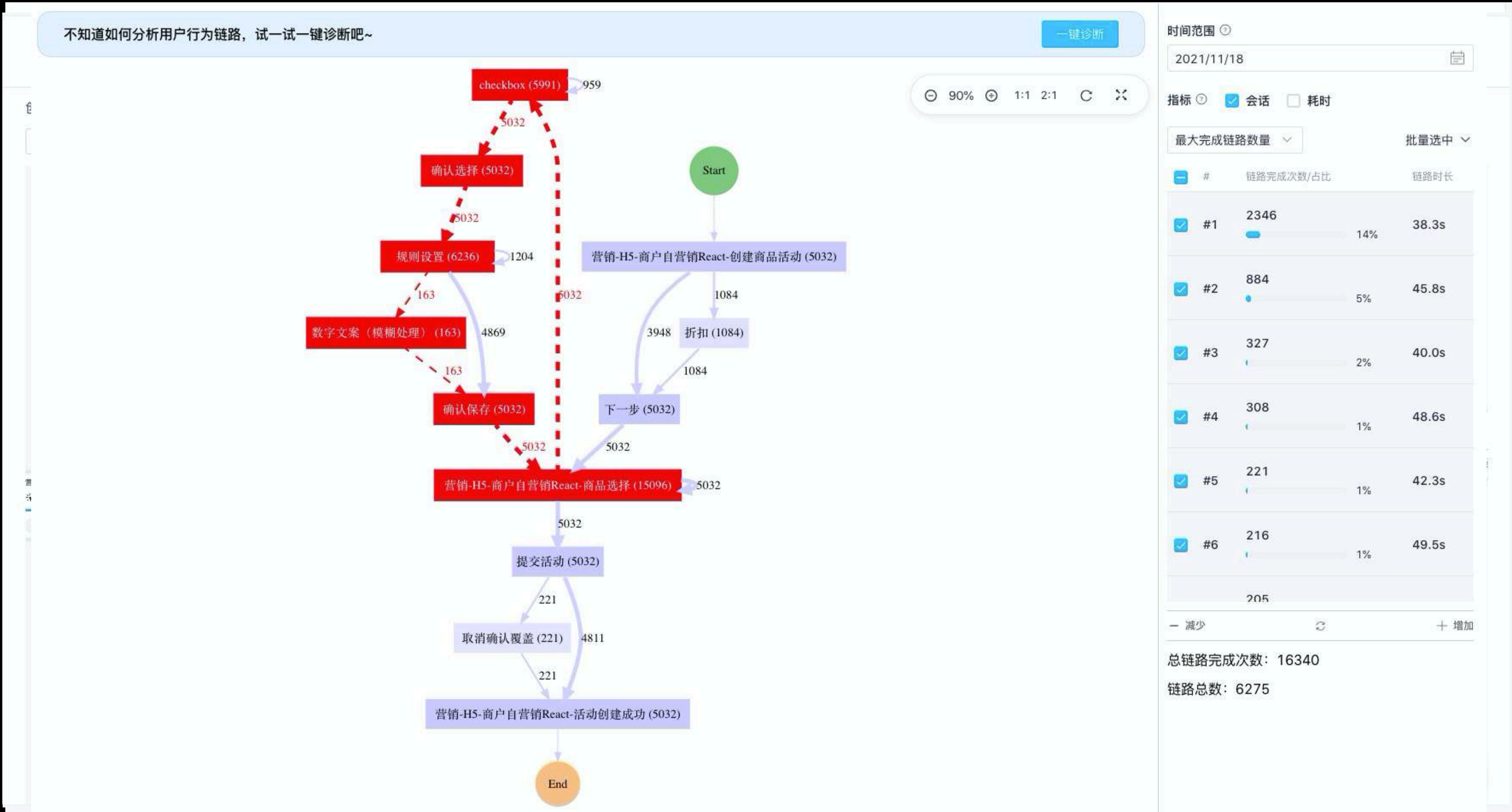
用户链路分析

➡

指定链路分析

➡

操作下钻分析







D2 前端技术论坛  
D2 FRONTEND TECHNOLOGY FORUM

精心

02

# 案例分析

晚上7:50

1.1M/s

71%

×

商品活动

创建必读

1 基础信息

2 创建活动

3 完成创建

选择活动人群

全部顾客  
全部订单来源

超级吃货  
高价值用户

门店新客  
潜力增量用户

促转化神器，大数据辅助配置，事半功倍不是梦

选择优惠类型

特价

折扣

基础信息设置

开始日期

2021-07-14

>

结束日期

2022-07-13

>

自动延期

自动延期开启后，将在活动到期后自动进行延期，每次延期30天，累计时长最多1年

生效时段

整周 全天

>

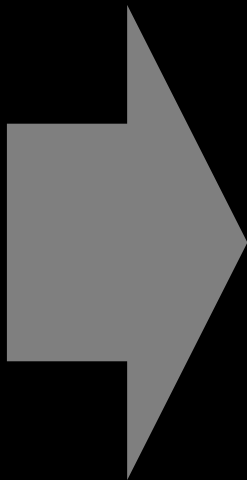
适用渠道

全渠道

>

核销须知

下一步



体验不好

繁琐

麻烦

太慢

不好用

不方便

感觉还行

变差

勉强

不流畅

负向用户反馈





将选商品、创建规则两个步骤合并，选择商品并填写规则。



那多个商品可以分开，一次选2个商品填8折，第二次2个填5折。

不行**假设**有4个商品一起创建，2个8折，2个5折这样不方便。

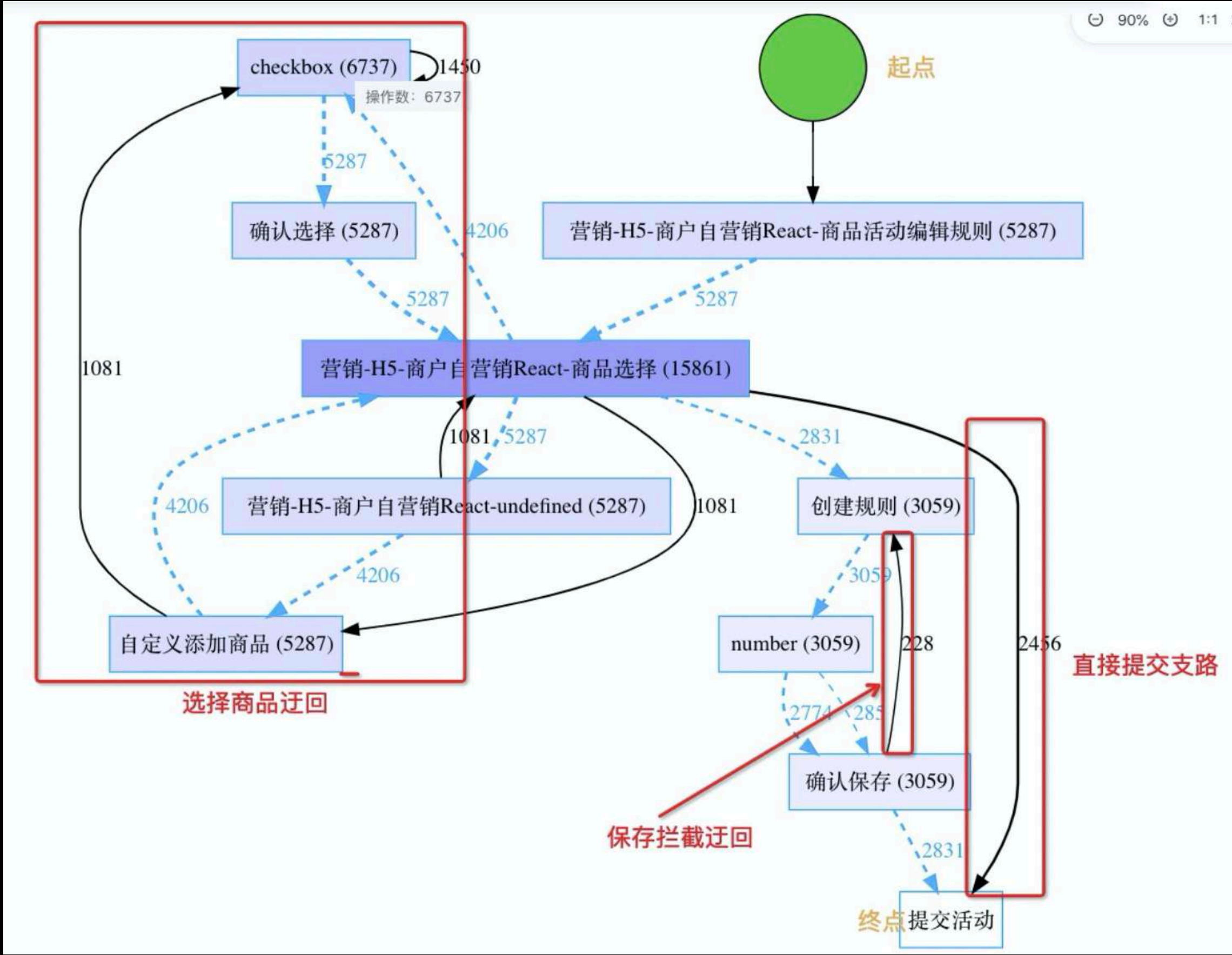


**可能**商家希望一次性选菜再分别填写规则，**我觉得**这样方便。



从场景组合和个人主观认知出发，很难得到结论。  
不清楚用户真实的操作习惯，迭代没有数据支撑。

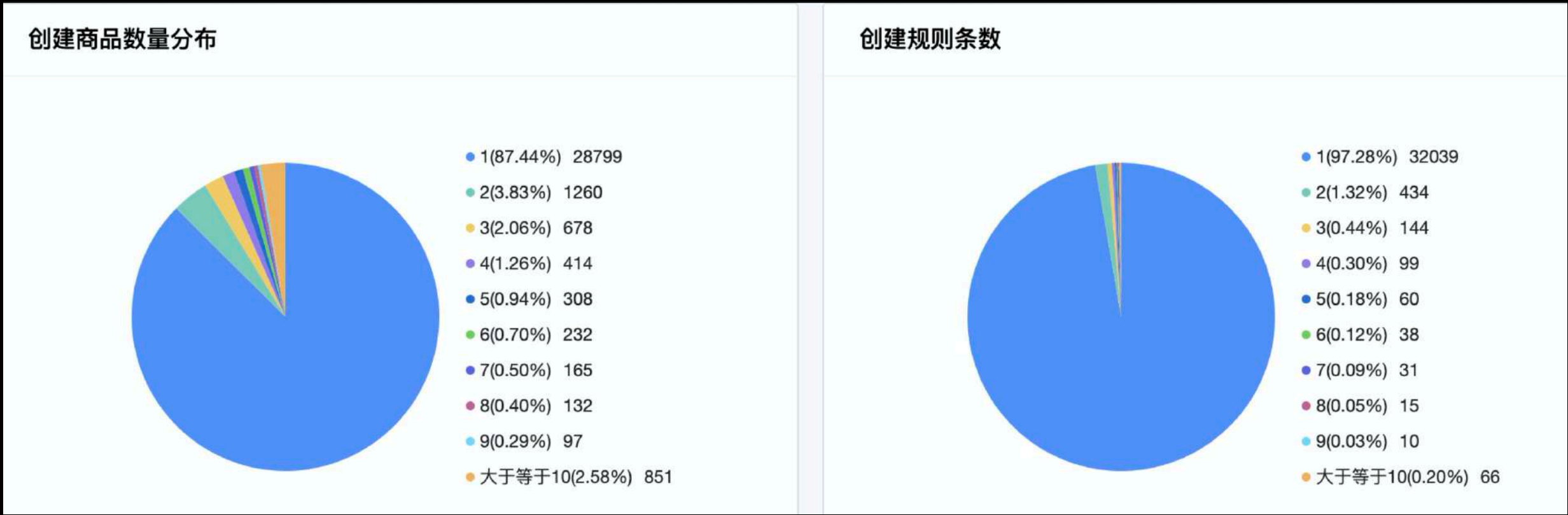




用户操作链路



提交链路分析



下钻分析



晚上7:23

45.3K/s

77%

商品活动

创建必读

1 基础信息

2 创建活动

3 完成创建

选择活动人群

全部顾客

超级吃货

门店新客

全部订单来源 高价值用户 潜力增量用户

促转化神器，大数据辅助配置，事半功倍不是梦

选择优惠类型

特价

折扣

基础信息设置

开始日期

2021-07-14

结束日期

2022-07-13

自动延期

自动延期开启后，将在活动到期后自动进行延期，每次延期30天，最多延12次，累计时长最多1年

生效时段

整周 全天

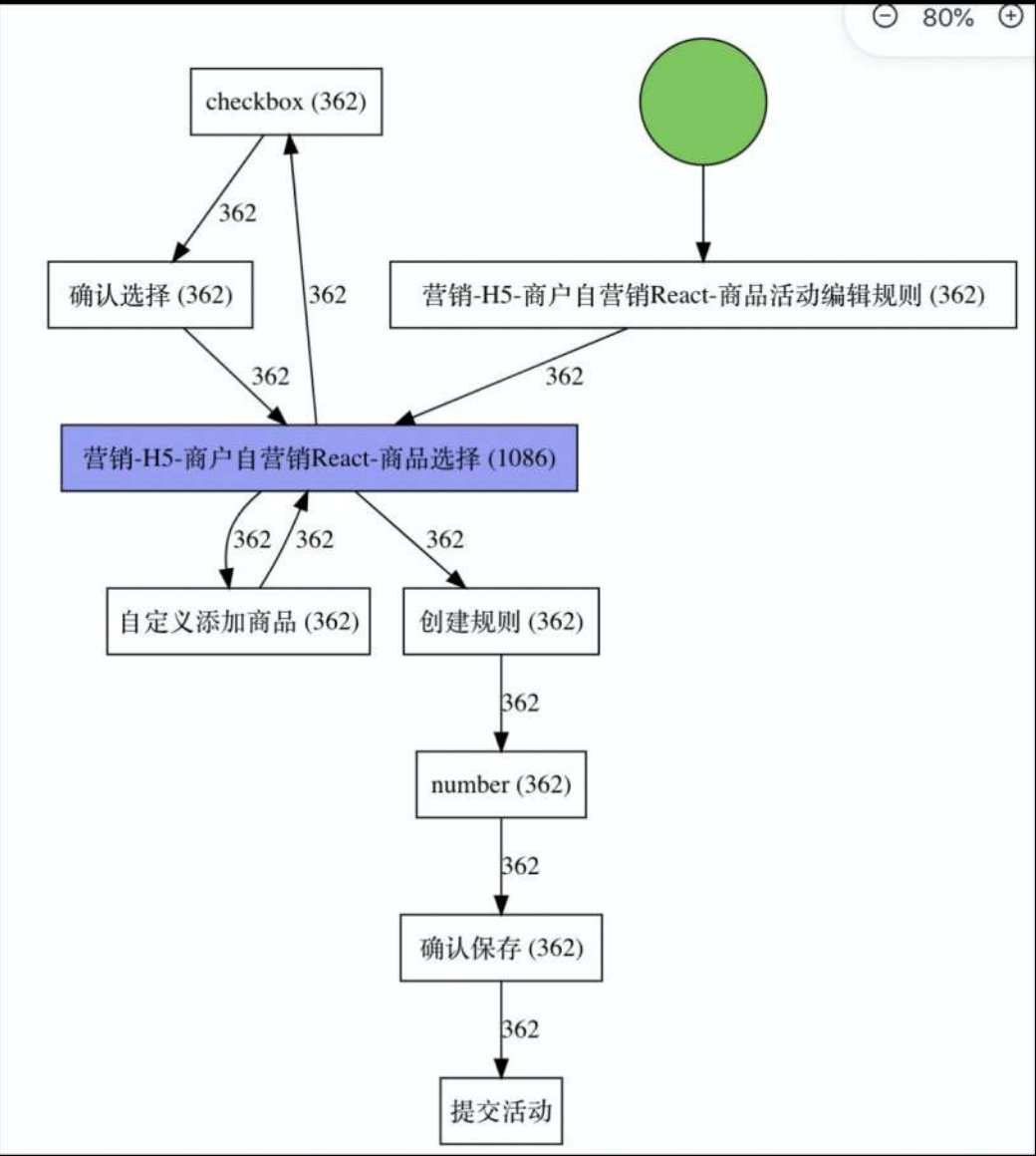
适用渠道

全渠道

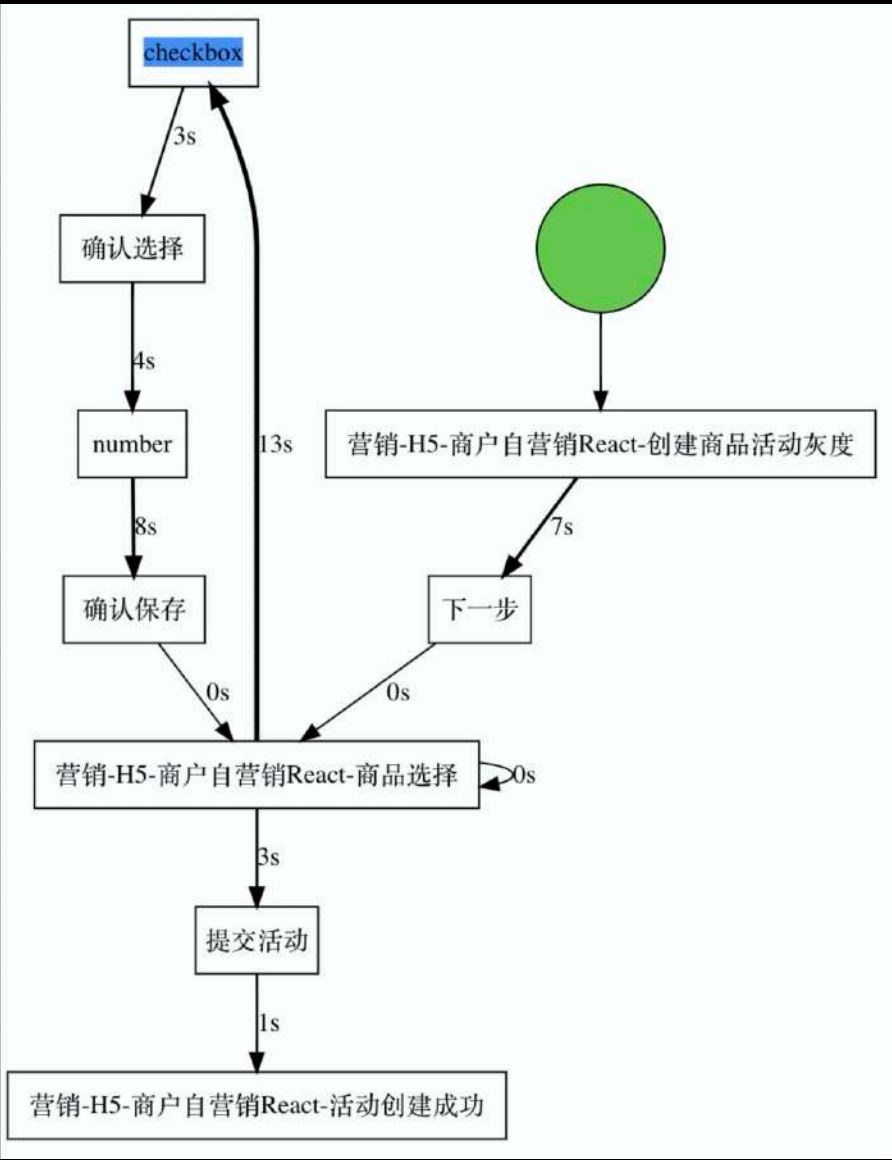
核销须知

下一步

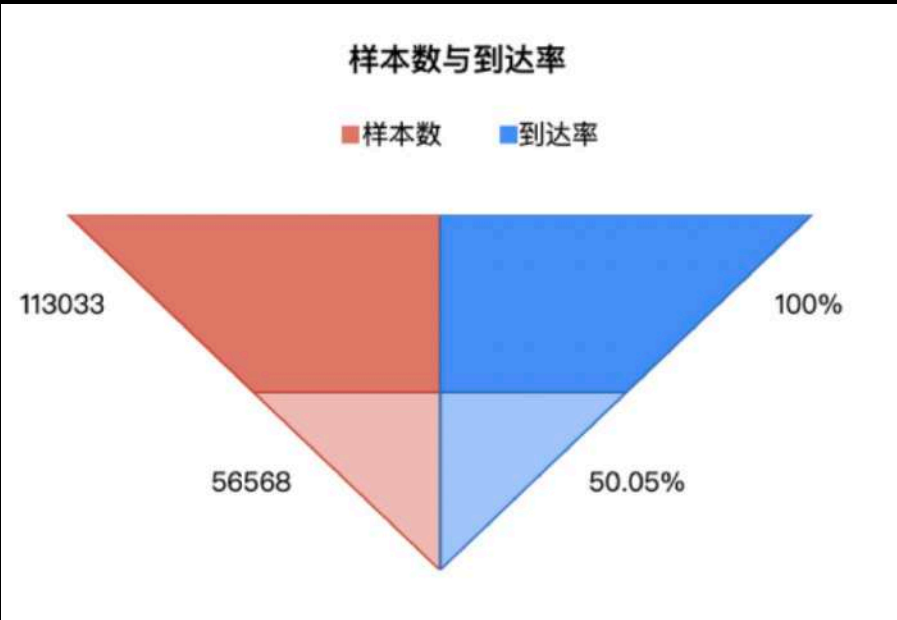
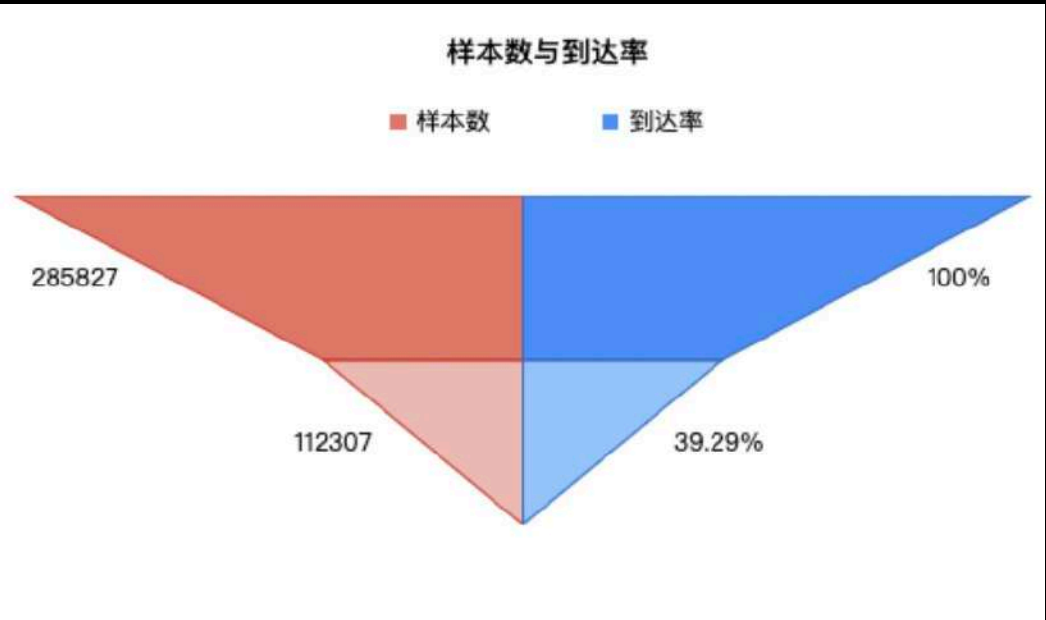
主链路占比 46%



主链路占比 91%



创建时长减少 43s，完成率提升 10%





## P2 技术篇

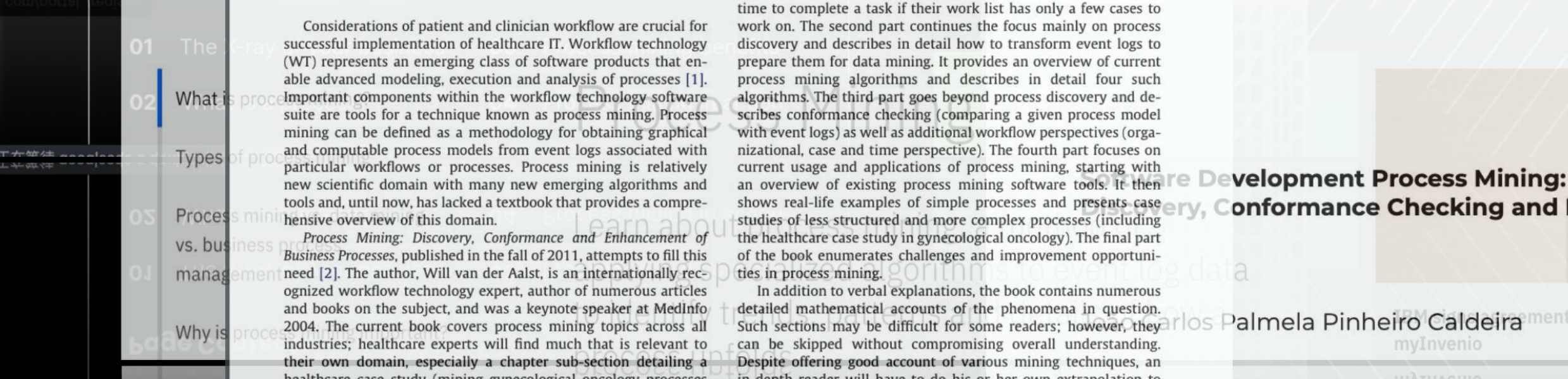
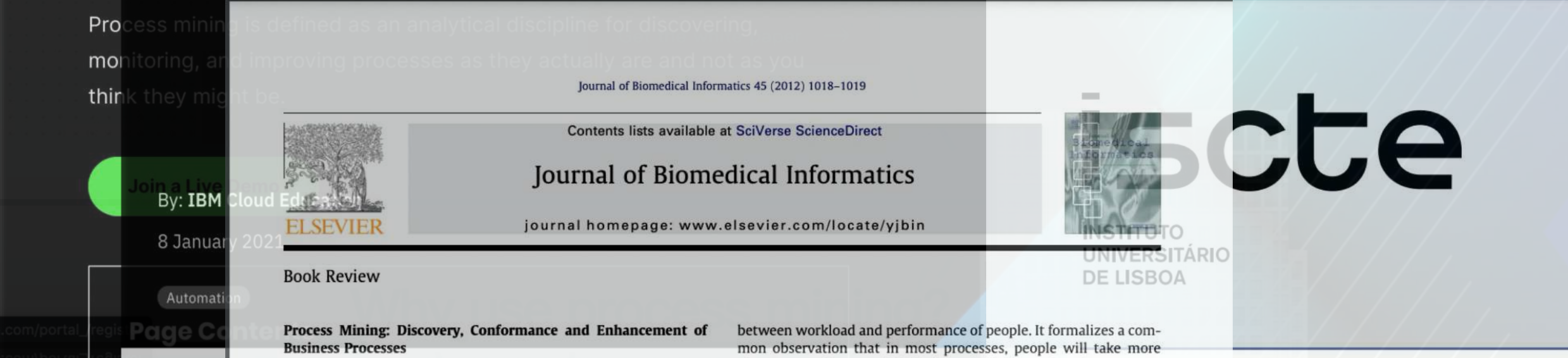
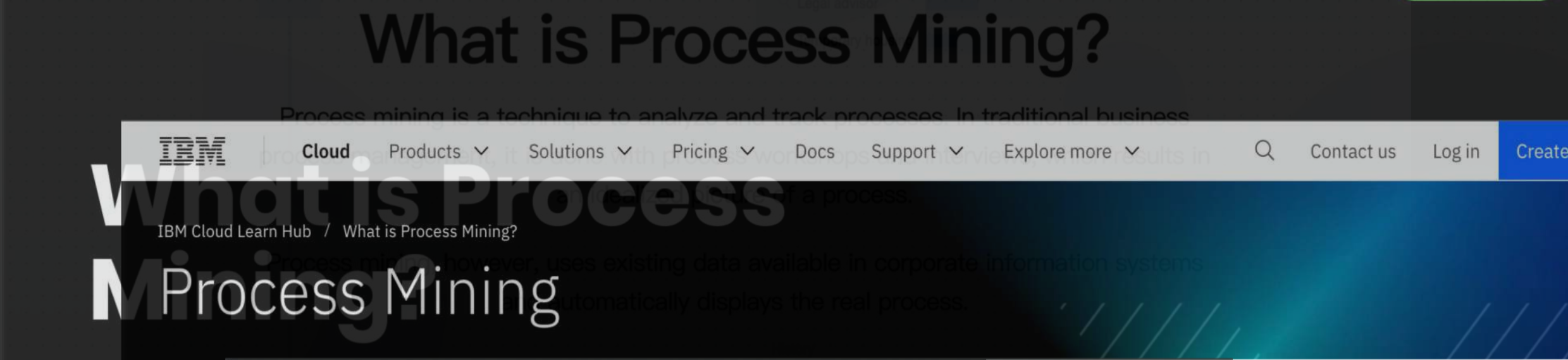
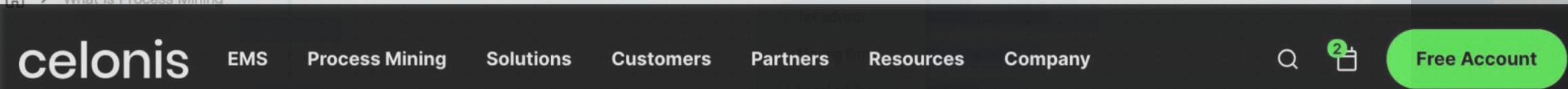
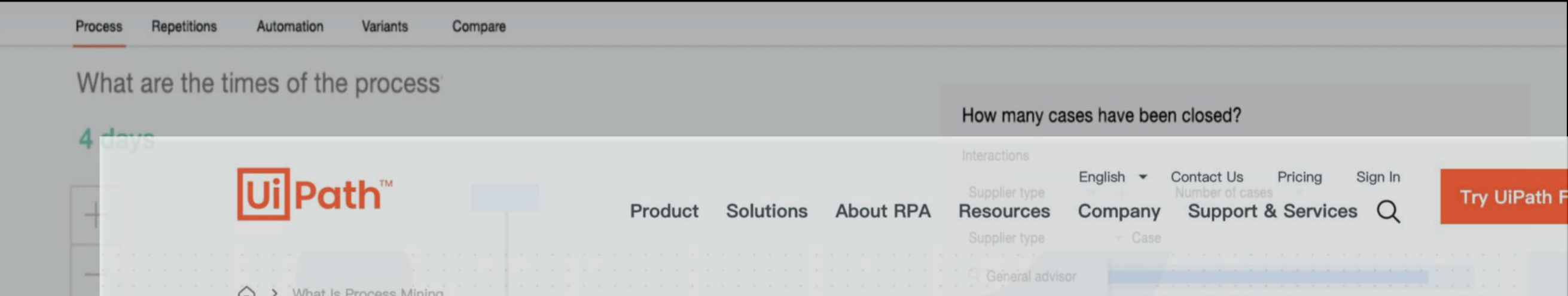
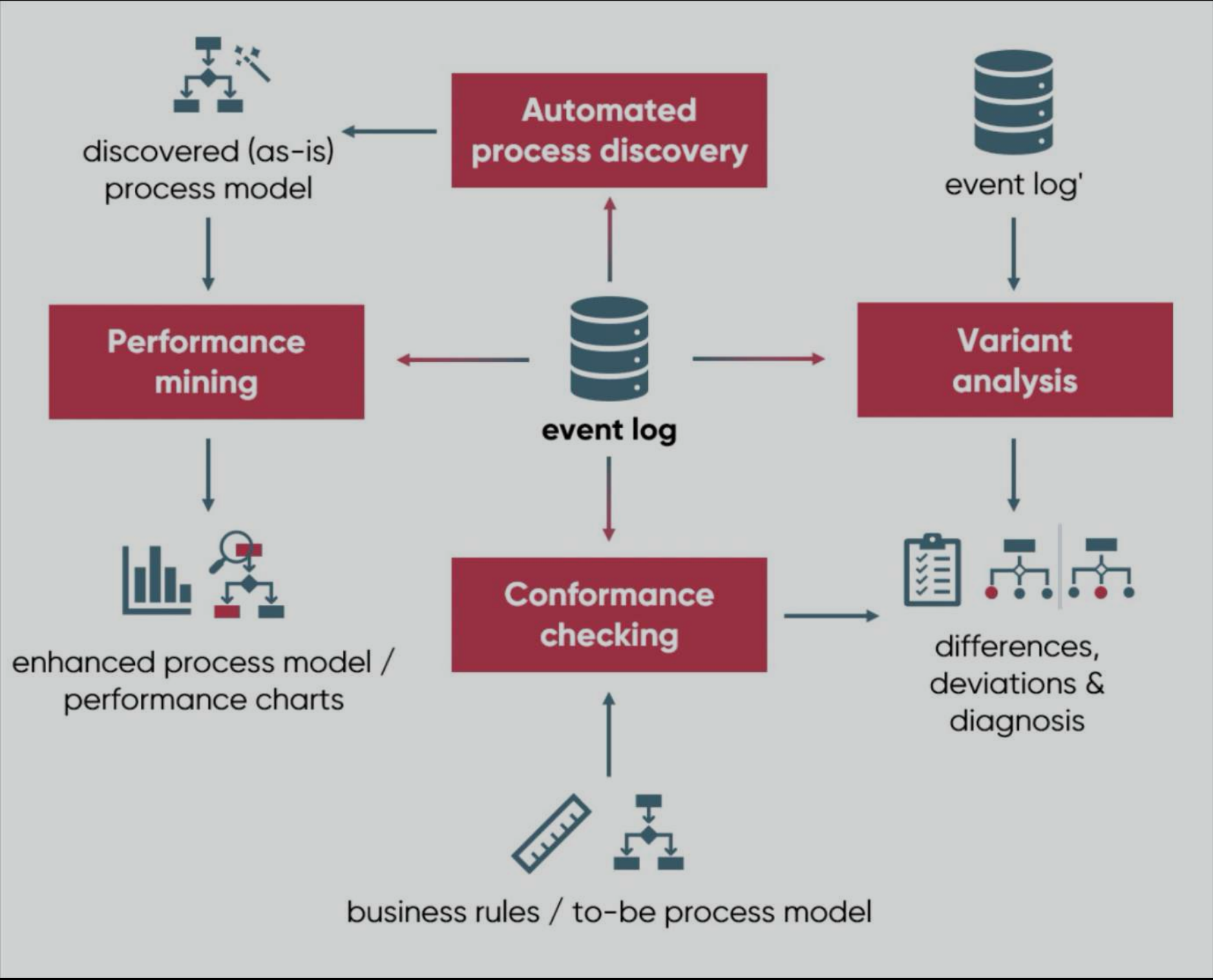
# 行为链路的识别、挖掘及诊断



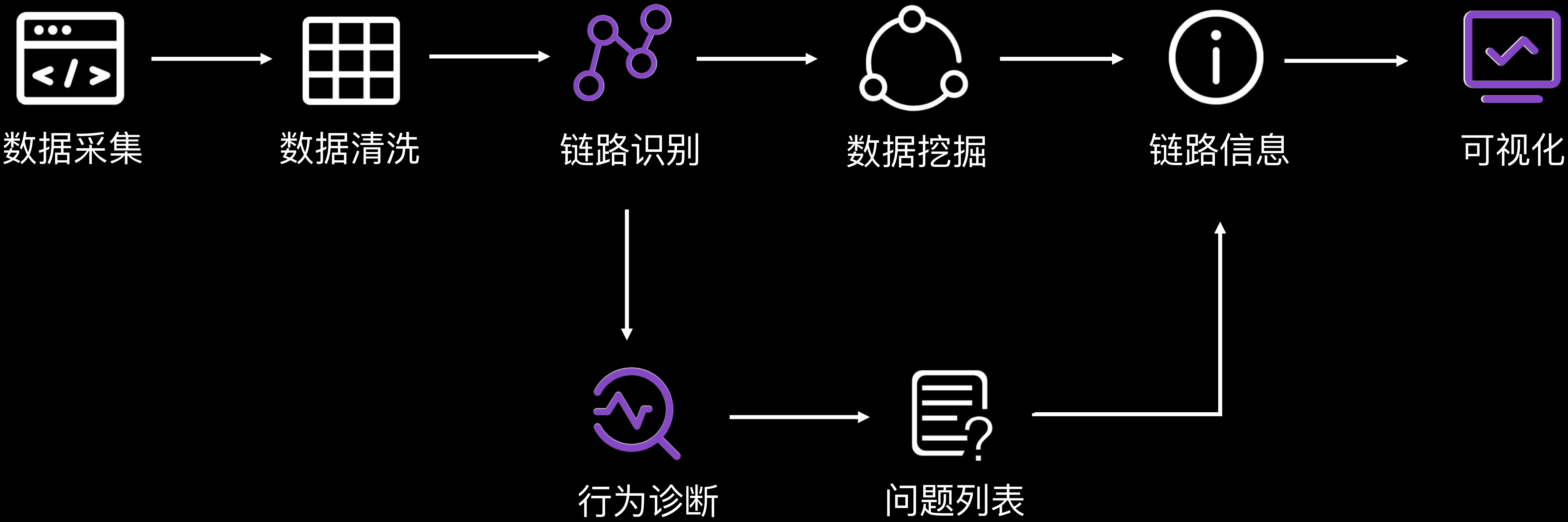
# Process Mining

[Wikipedia's explanation](#) of process mining:

During process mining, specialized data mining algorithms are applied to event log data to identify trends, patterns, and details contained in event logs recorded by an information system.











D2 前端技术论坛  
D2 FRONTEND TECHNOLOGY FORUM

精心

03

# 链路识别

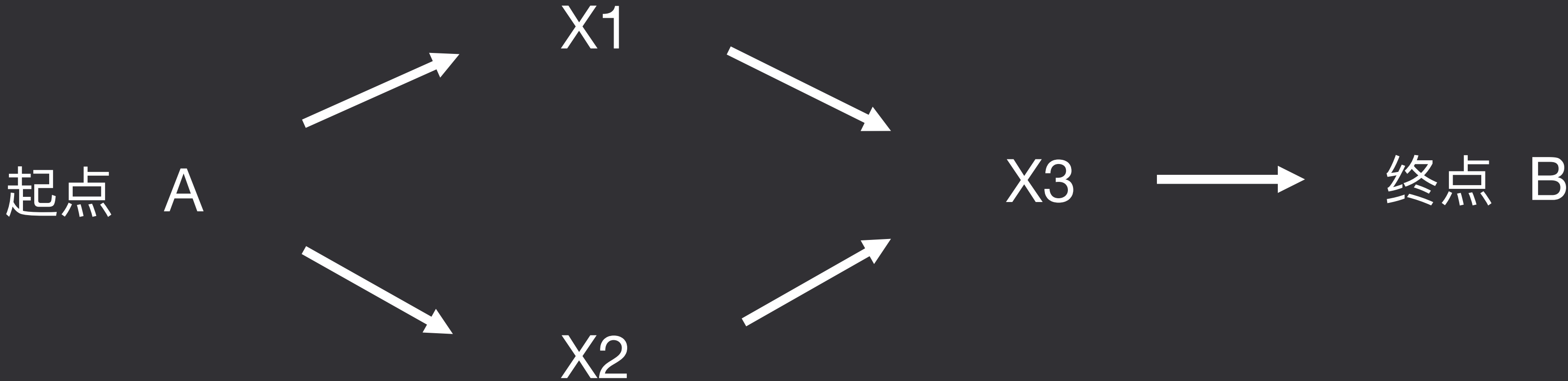


用户的操作数据

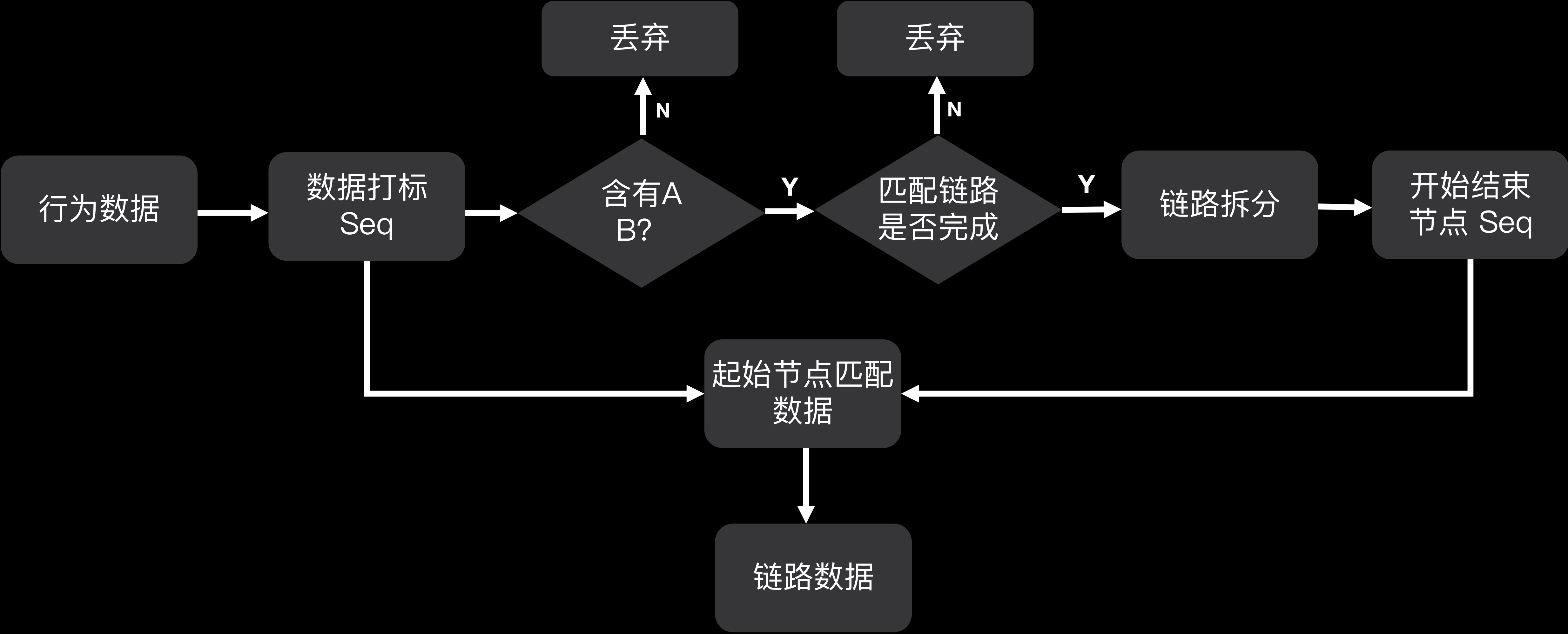
User1:	S	→	A	→	X <sub>1</sub>	→	X <sub>3</sub>	→	B	→	C
User2:	S	→	A	→	X <sub>2</sub>	→	X <sub>3</sub>	→	B	→	C
User1:	S	→	A	→	X <sub>2</sub>	→	X <sub>3</sub>	→	C	→	C
User2:	S	→	Y	→	X <sub>3</sub>	→	X <sub>4</sub>	→	B	→	C

↓ A为起始点 B为终止点

行为链路图



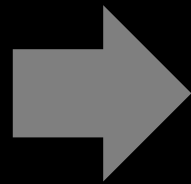






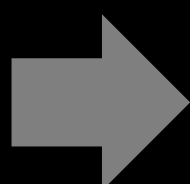
1. 数据打标

sid	node	seq
1	S	1
1	A	2
1	X <sub>1</sub>	3
1	X <sub>3</sub>	4
1	B	5
1	C	6
2	S	1
2	A	2
2	X <sub>2</sub>	3
2	X <sub>3</sub>	4
2	B	5
...	...	...



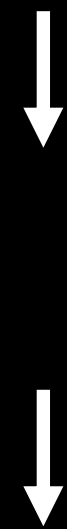
2. 保留A, B, 并标记

sid	node	seq	mark
1	A	2	1
1	B	5	2
1	A	10	1
2	A	2	1
2	B	5	2
2	B	9	2



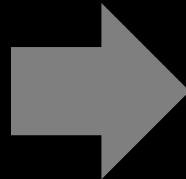
3. 行匹配

sid	node	seq	mark
1	A	2	1
1	B,	5,	2,
1	A	10	1
2	A	2	1
2	B,	5,	2,
2	B	9	2

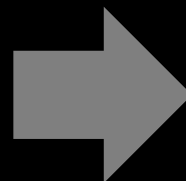




3. 多行转一行



sid	node	seq	mark
1	A,B,,A	2,5,,10	1,2,,1
2	A,B,,B	2,5,,9	1,2,,2



4. 一行转多行

sid	node	seq	mark	chainId
1	A,B	2,5	1,2	1
1	A	10	1	2
2	A,B	2,5	1,2	3
2	B	9	1	4

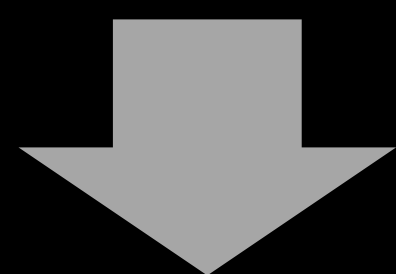


海量数据

数十亿/日

识别算法复杂

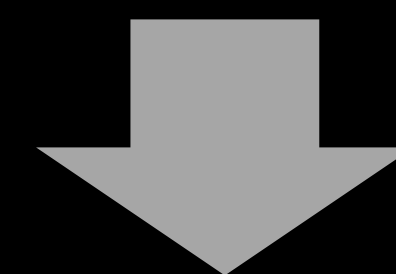
十余步骤



1. 数据产出慢

行为路径繁杂

动作不可预知



2. 链路过于分散

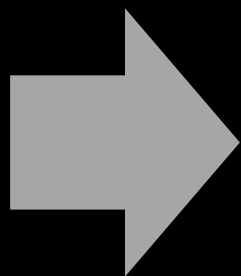


# 1. 性能解决方案

增加数据链路关系，优化中间处理节点，快速构造用户链路树

批处理

ct	node
1	A
2	B
3	C
4	D



流处理

ct	node	father	child
1	A		B
2	B	A	C
3	C	B	D
4	D	C	



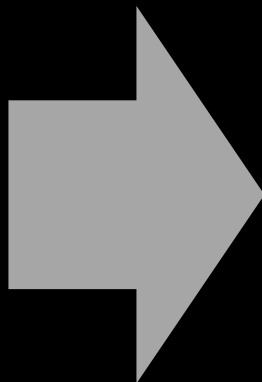
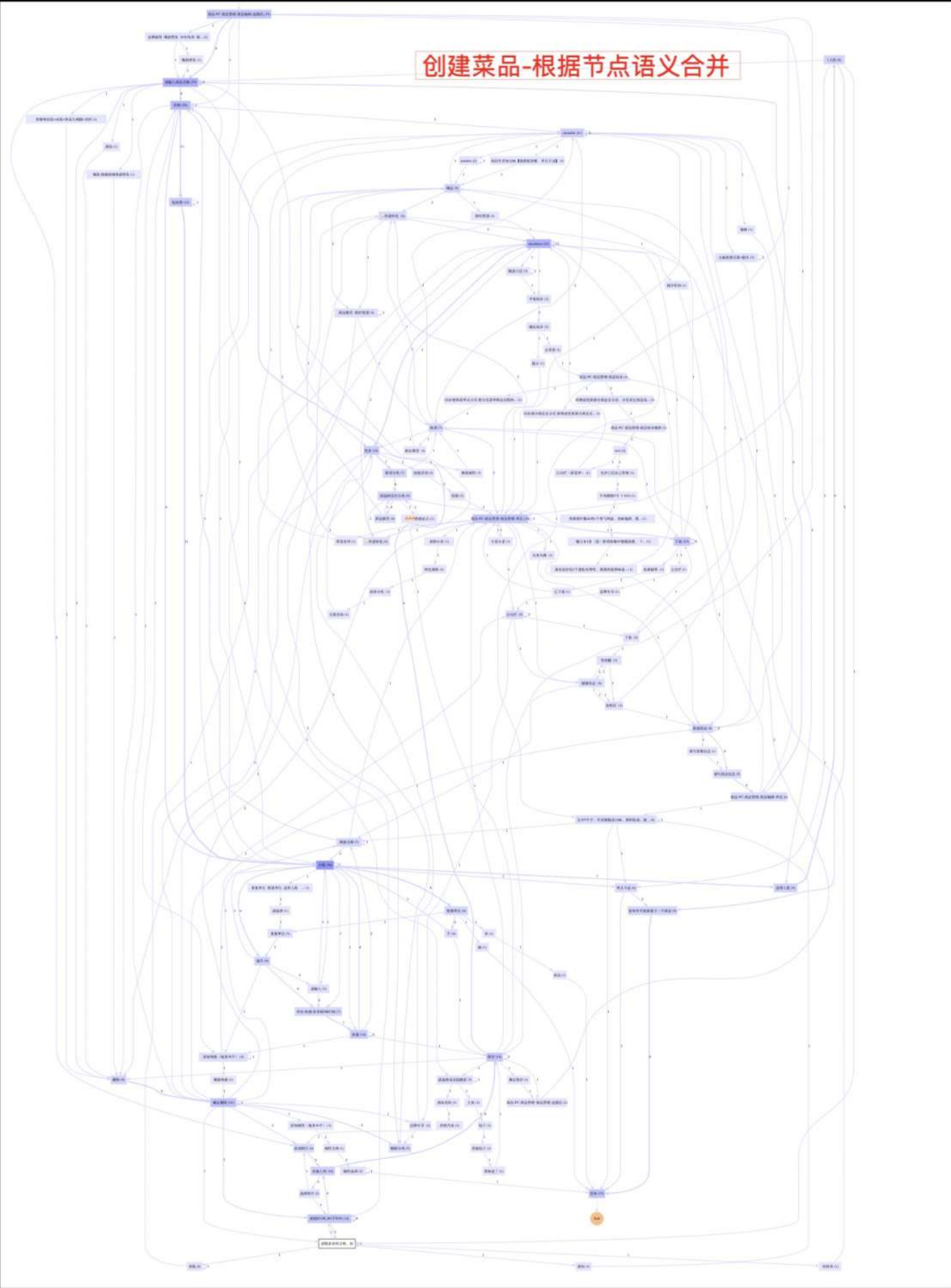
# 2. 链路分散解决方案

当节点的文案带有个性化语义，或链路中过多节点时，会导致链路过于分散，失去趋势观测意义。

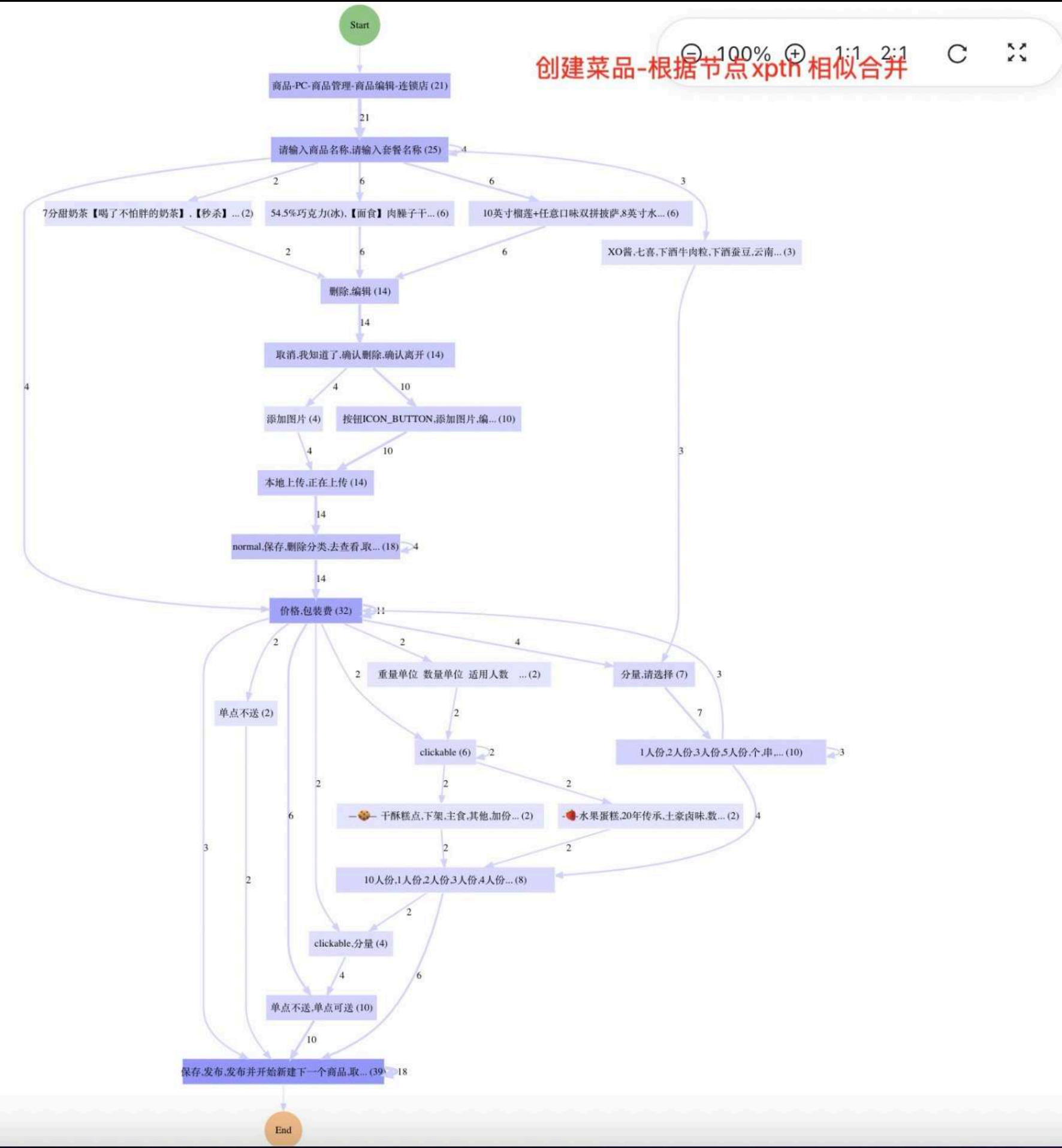




# 节点相似合并



相同链路数据下对比







D2 前端技术论坛  
D2 FRONTEND TECHNOLOGY FORUM

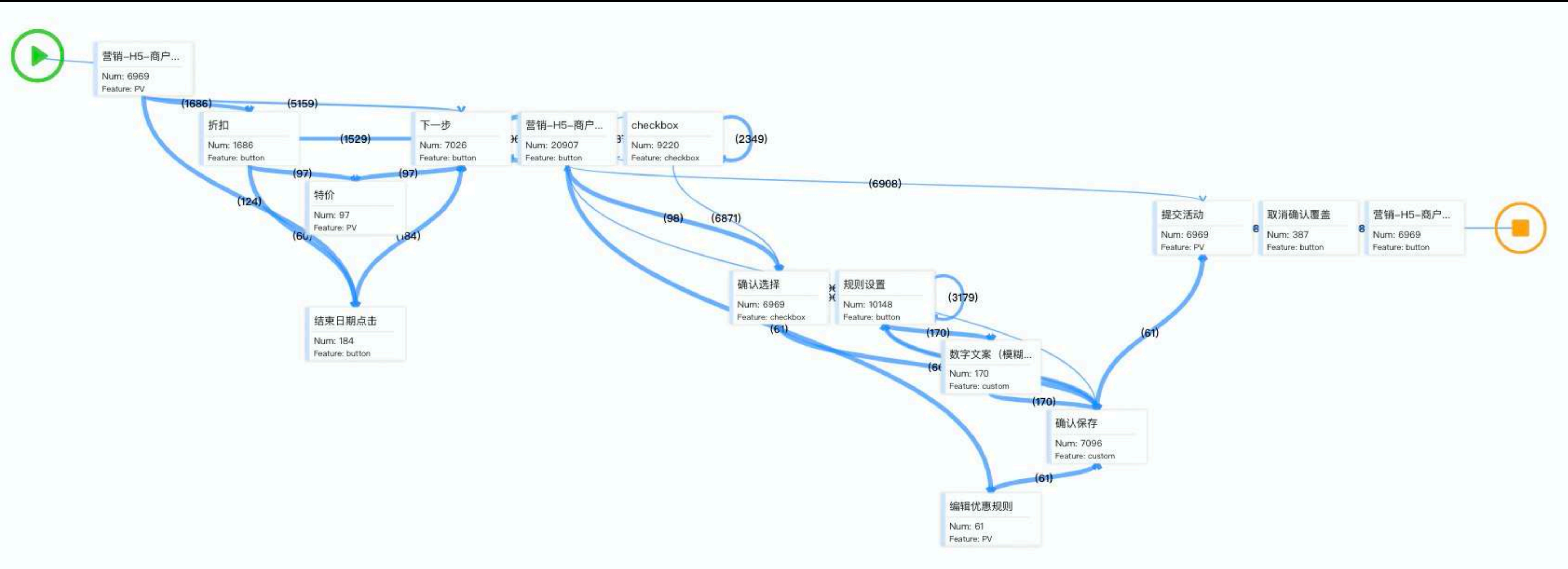
精心

04

# 链路可视化



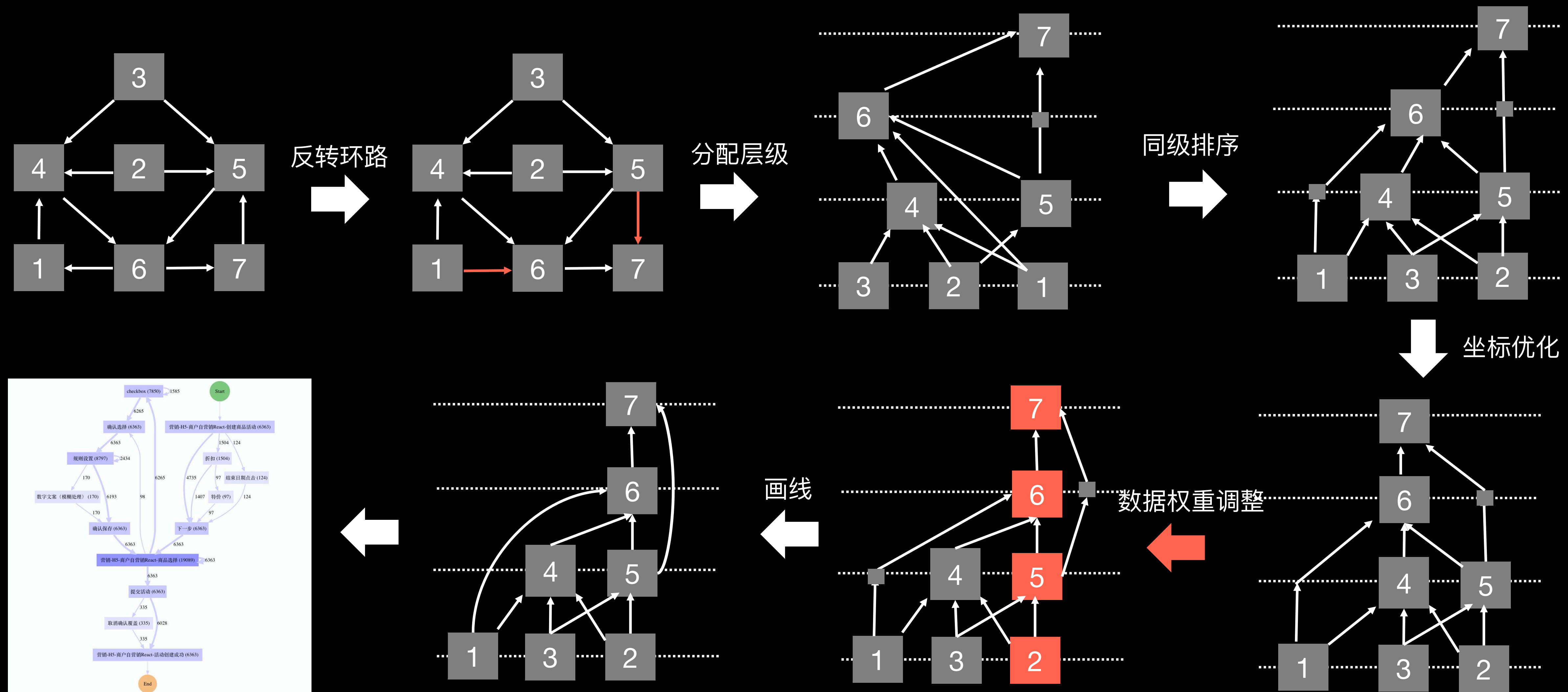
Drage 布局是很多可视化库选择的方案，但是在用户操作链路的场景中，效果并不理想。



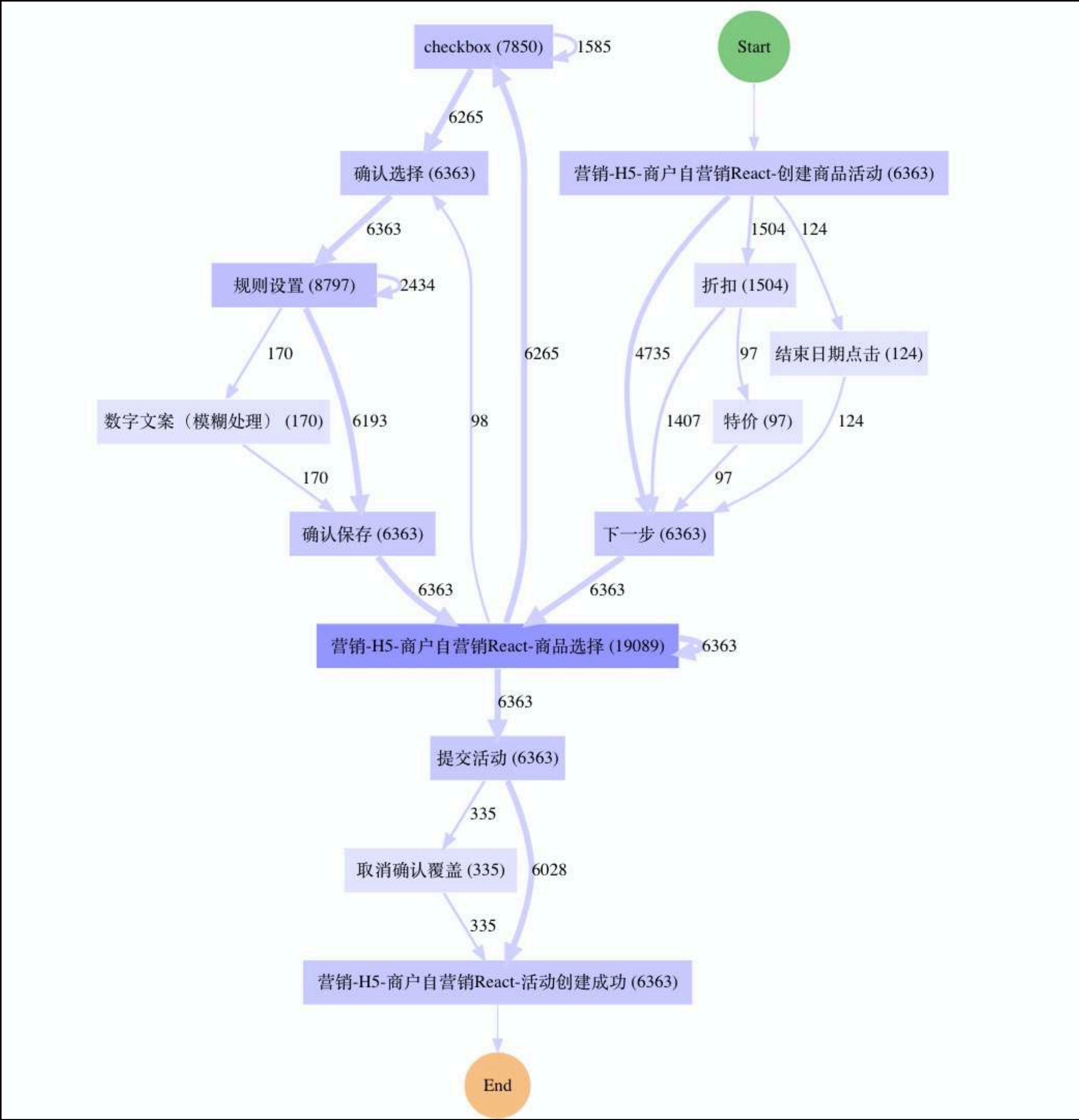
Top30 链路图（Dagre布局）



在 Dot Layout Engine 基础上，引入节点数据关系权重









# 用户链路的复杂性

12:04

吃货红包

创建必读

补贴翻倍

多场景增单

提升实收

根据您的近期的经营情况，为您推荐：

吃货红包

自领取日起7天内有效

红包金额

4.0

元

提高红包金额，有助于提升曝光转化哦

有效期

自领取日起 7天 内

投放规则

发放渠道

店外发放

获得更多曝光

店内领取

提高下单转化

每日发放

25

起始日期

2021-10-23

结束日期

2021-10-26

自动延期

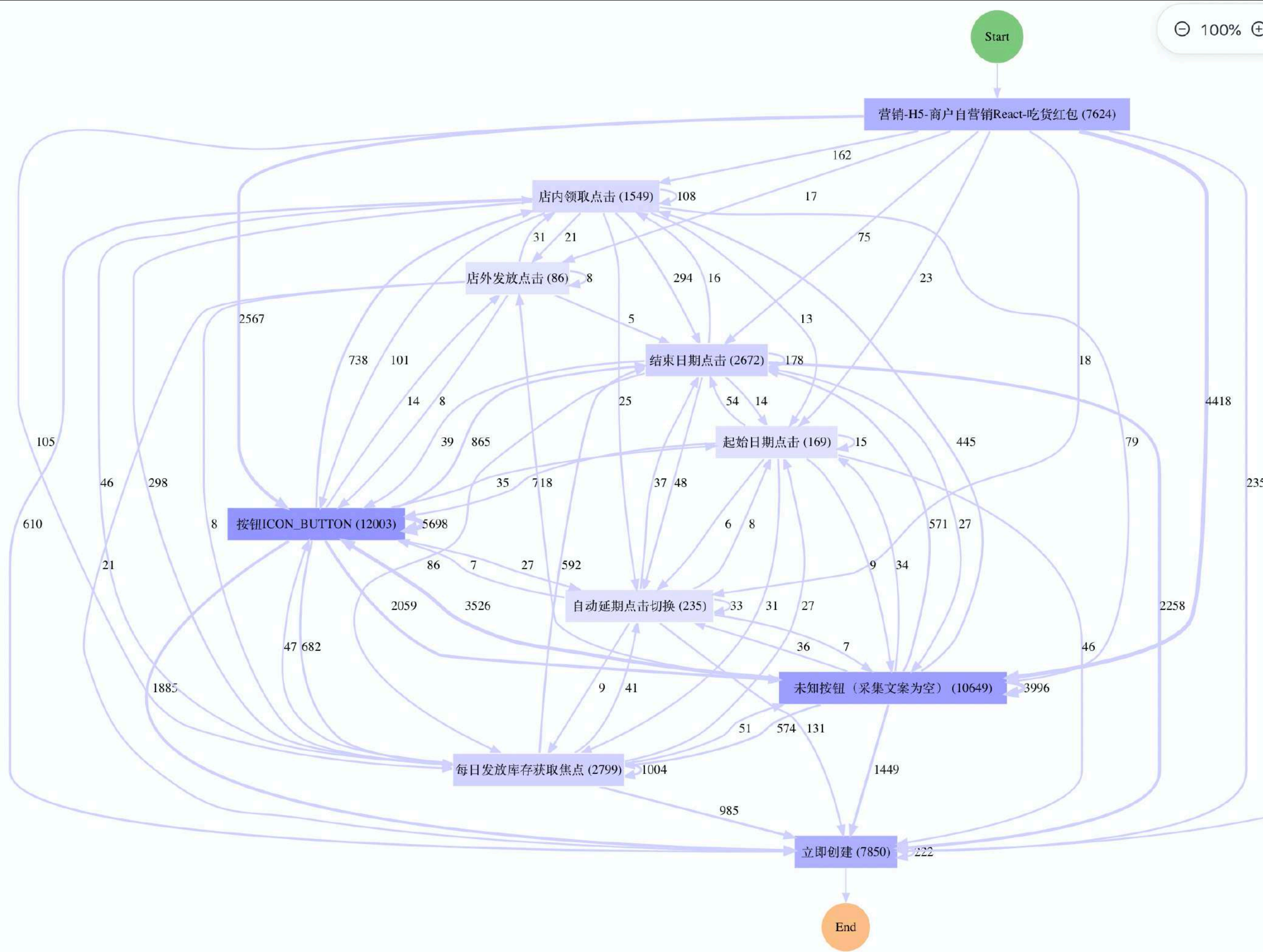
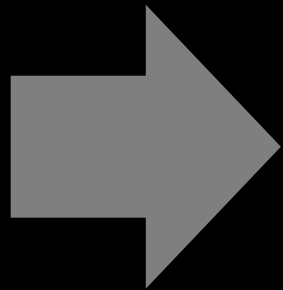
自动延期开启后，将在活动到期后自动进行延期，每次延期30天，累计时长最多1年

同享规则

默认与满减、配送费等店铺活动同

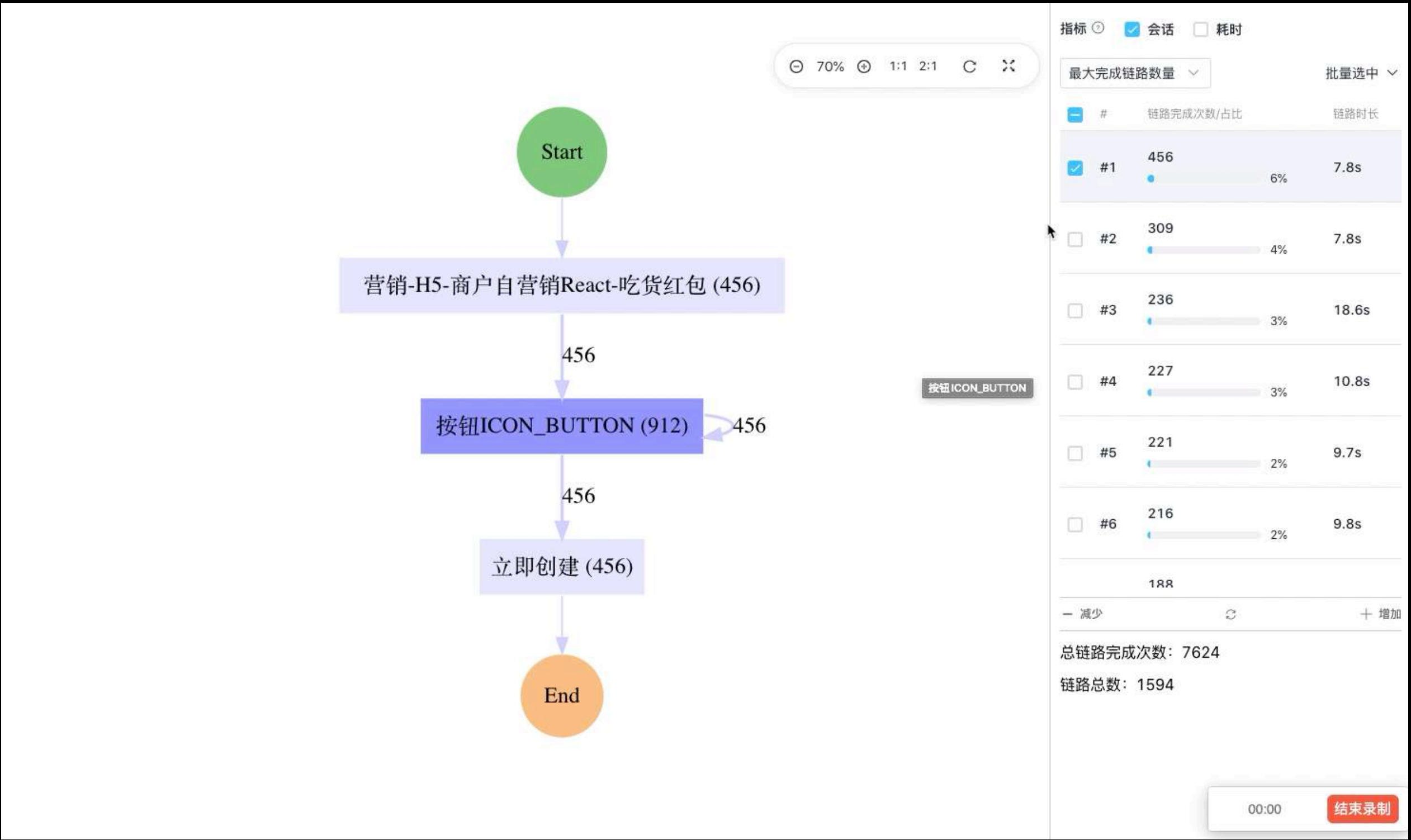
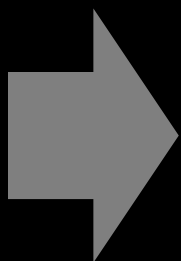
立即创建

十个节点，简单？





1. 多维度拆分链路（ 链路数量， 链路时间）
2. 关注热点 TOP 路径
3. 链路图数据展示（节点热度， 路径流量）



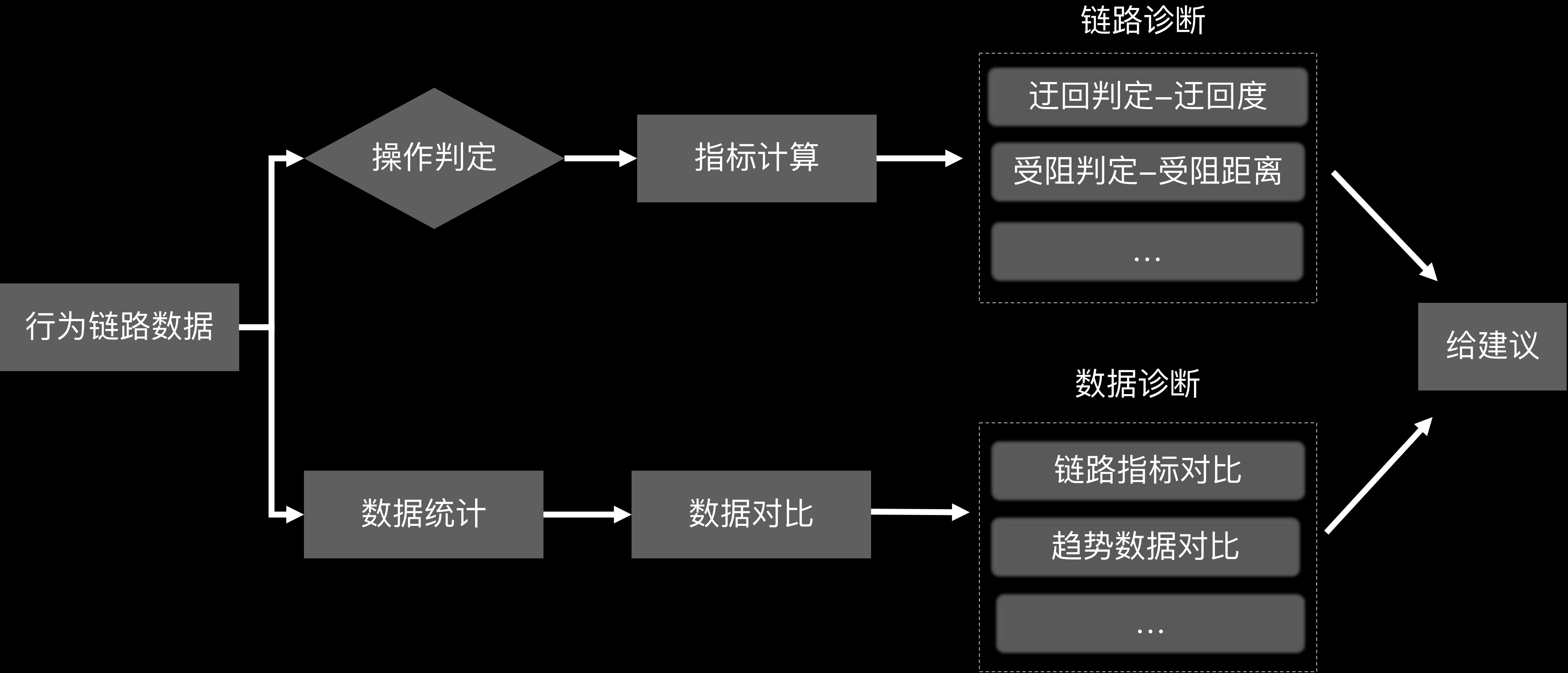




05

# 链路诊断





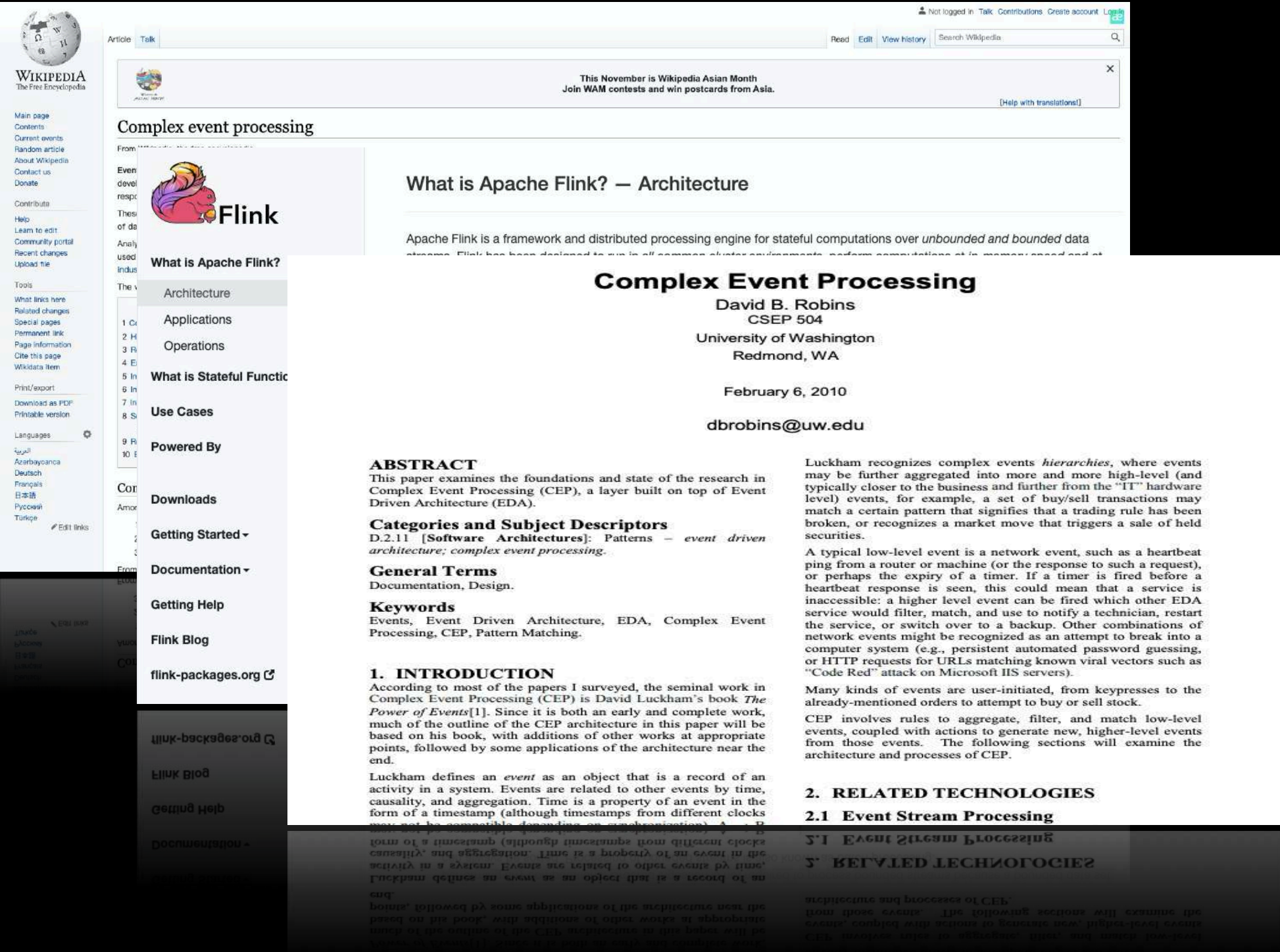


# 操作判定 1.事件流匹配

复杂事件处理 CEP（Complex event processing）

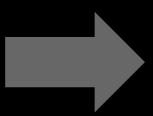
基于事件流的模式匹配（Pattern matching over event streams）

应用：数据处理（Fink cep，Esper等），端智能等场景

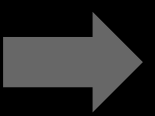


预设行为模式列表  
Preset Behavior Pattern List

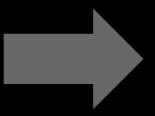
行为链路数据  
Behavior Data



事件流  
Event streams



模式匹配 JS实现  
JS-CEP Pattern match



匹配结果  
Match result



## Pattern

通过定义时间序列的模式来  
描述事件间的衔接关系。

next

notNext

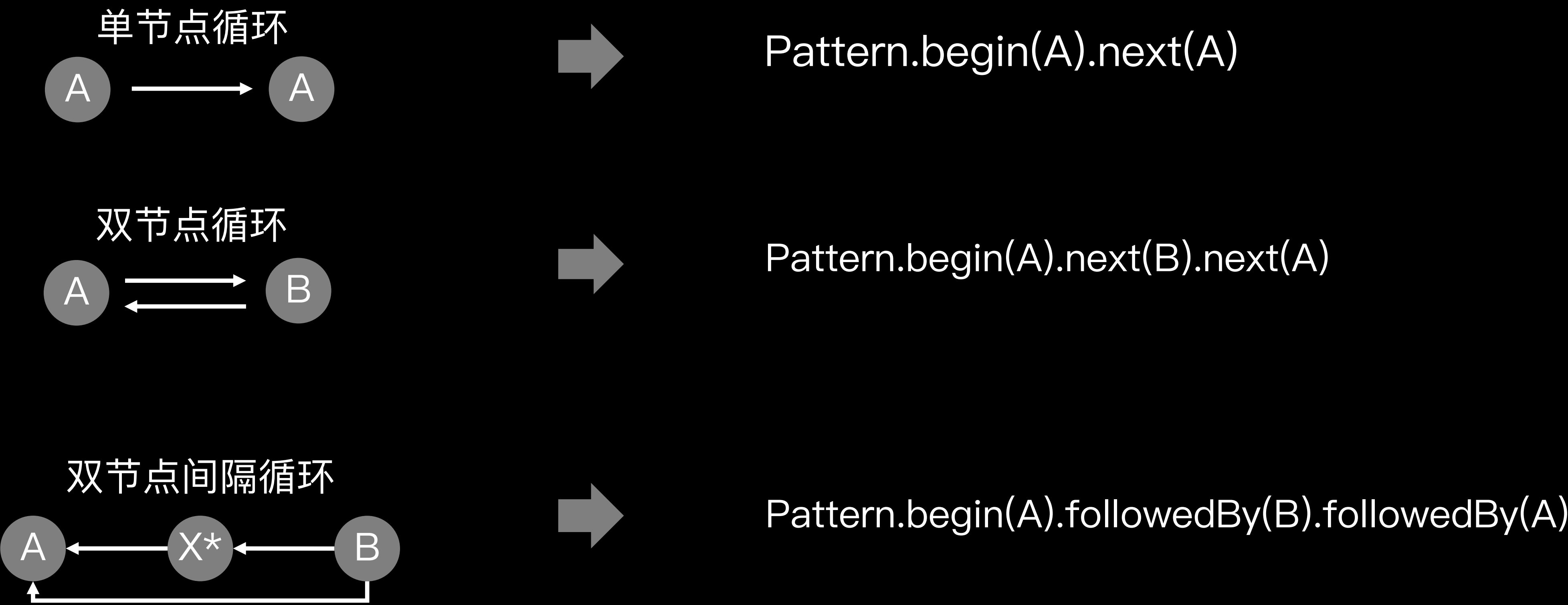
followedBy

followedByAny

notFollowedBy



例如：用户 循环 操作场景



其他场景依次类推，支路判定，操作受阻判定等场景进行规则描述。



Pattern 描述

Pattern.begin(A).next(A)

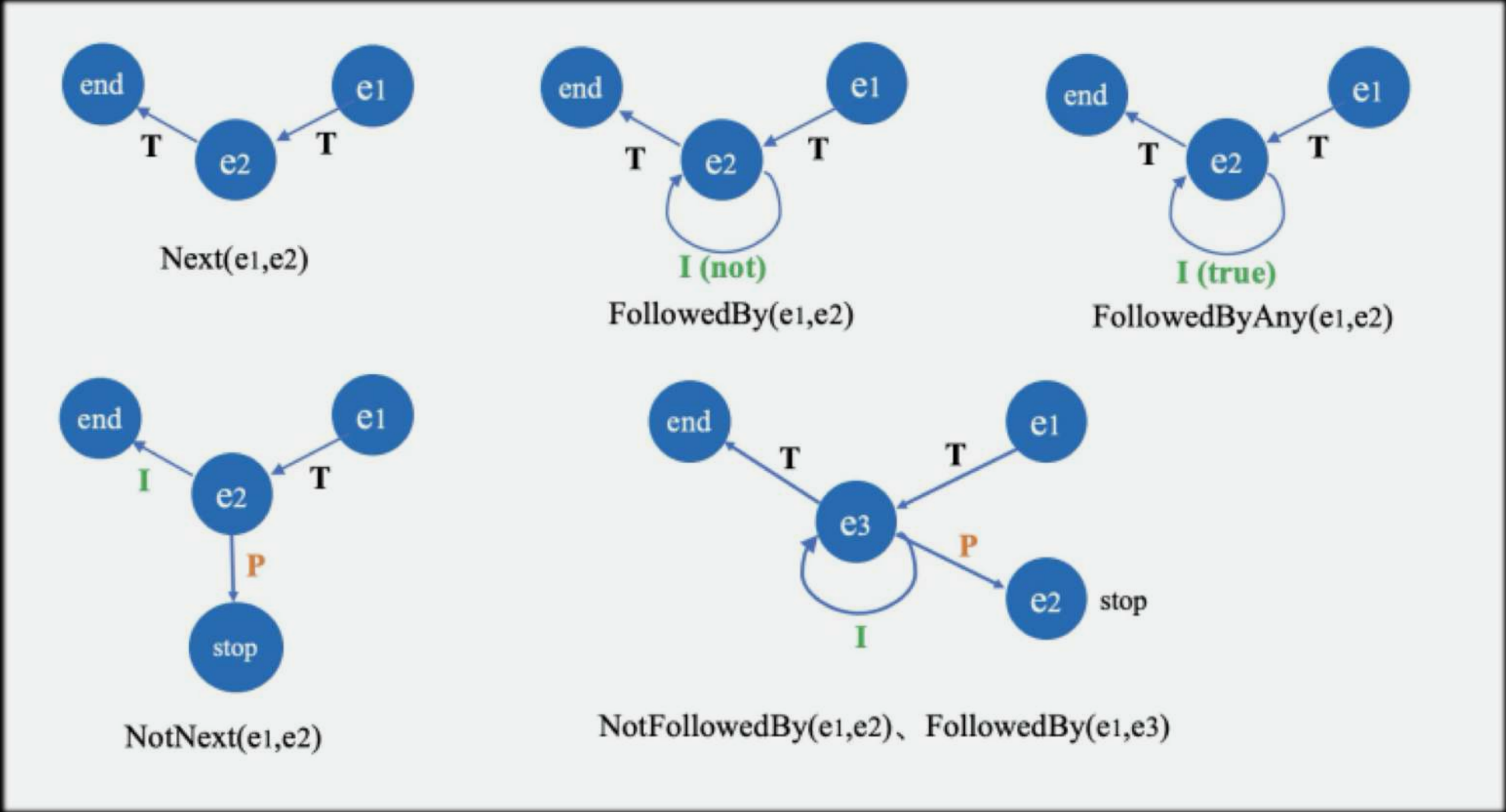
Pattern.begin(A).next(B).next(A)

Pattern.begin(A).followedBy(B).followedBy(A)

编译

NFA

(非确定的有限状态机)



当有事件触发时，就会进行 事件和规则 的匹配，根据判断事件和规则的匹配关系，决定是否 前进、忽略、废弃，假设最终走到 Final 时，表示定义的序列模式匹配成功，触发成功回调；如果走到 Stop，则会放弃已经匹配的序列模式。



图匹配（Graph Matching）是图同构（Graph Isomorphism）问题的进一步求解，复杂度未知，粗略估计属于 NP Hard 。

操作判定场景属于有向图的多次图匹配问题。

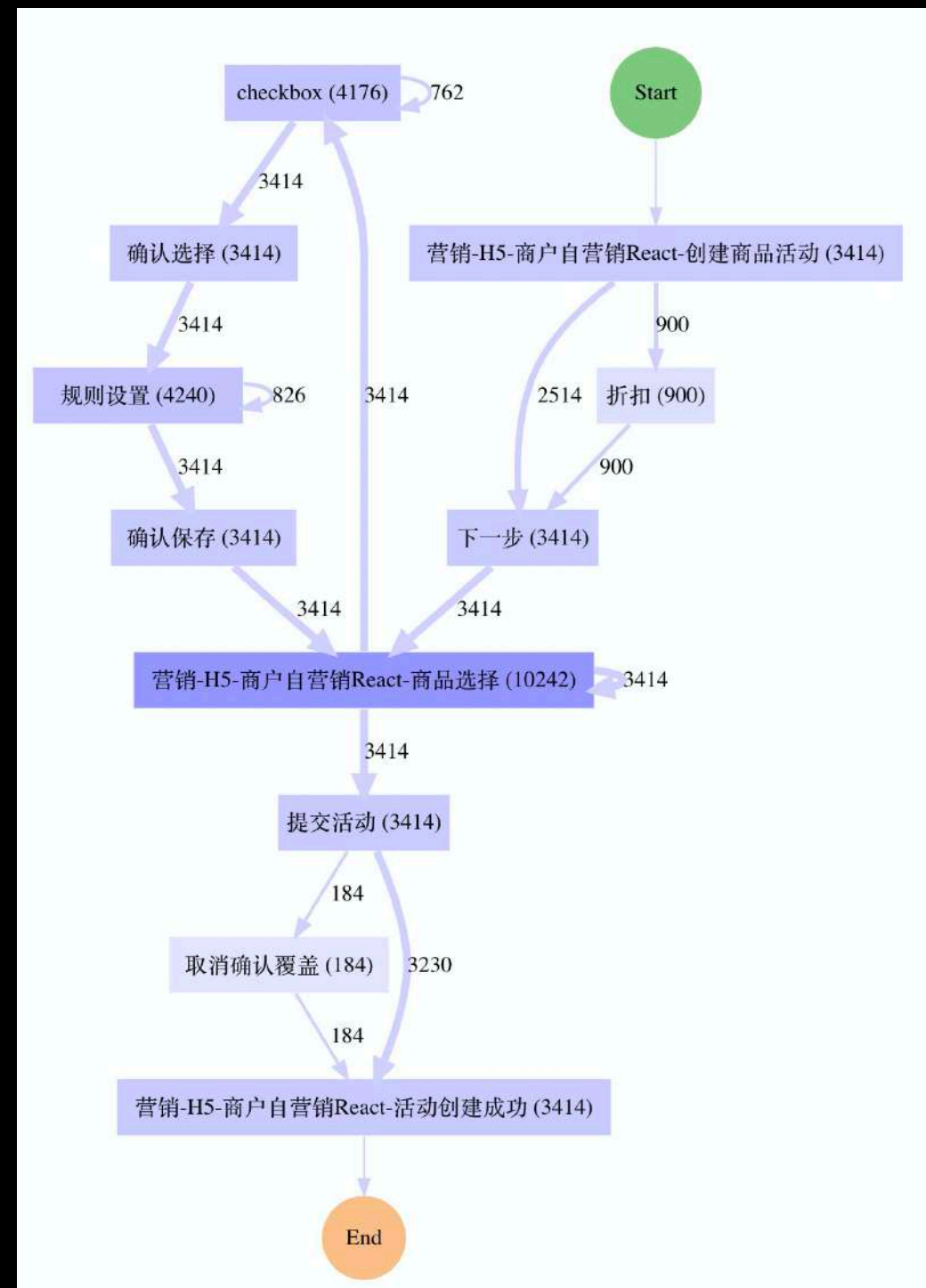
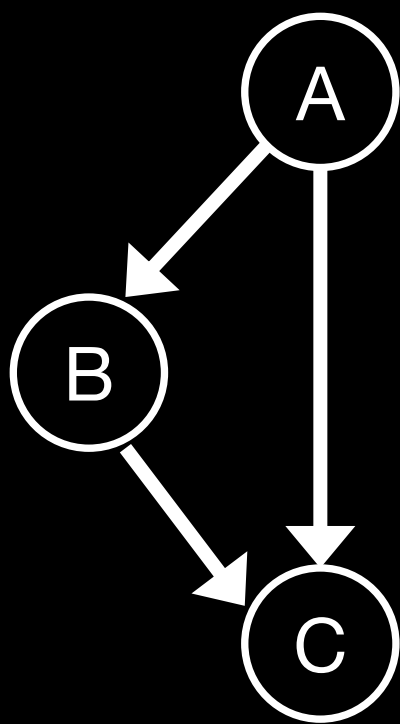
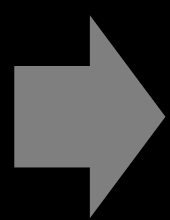
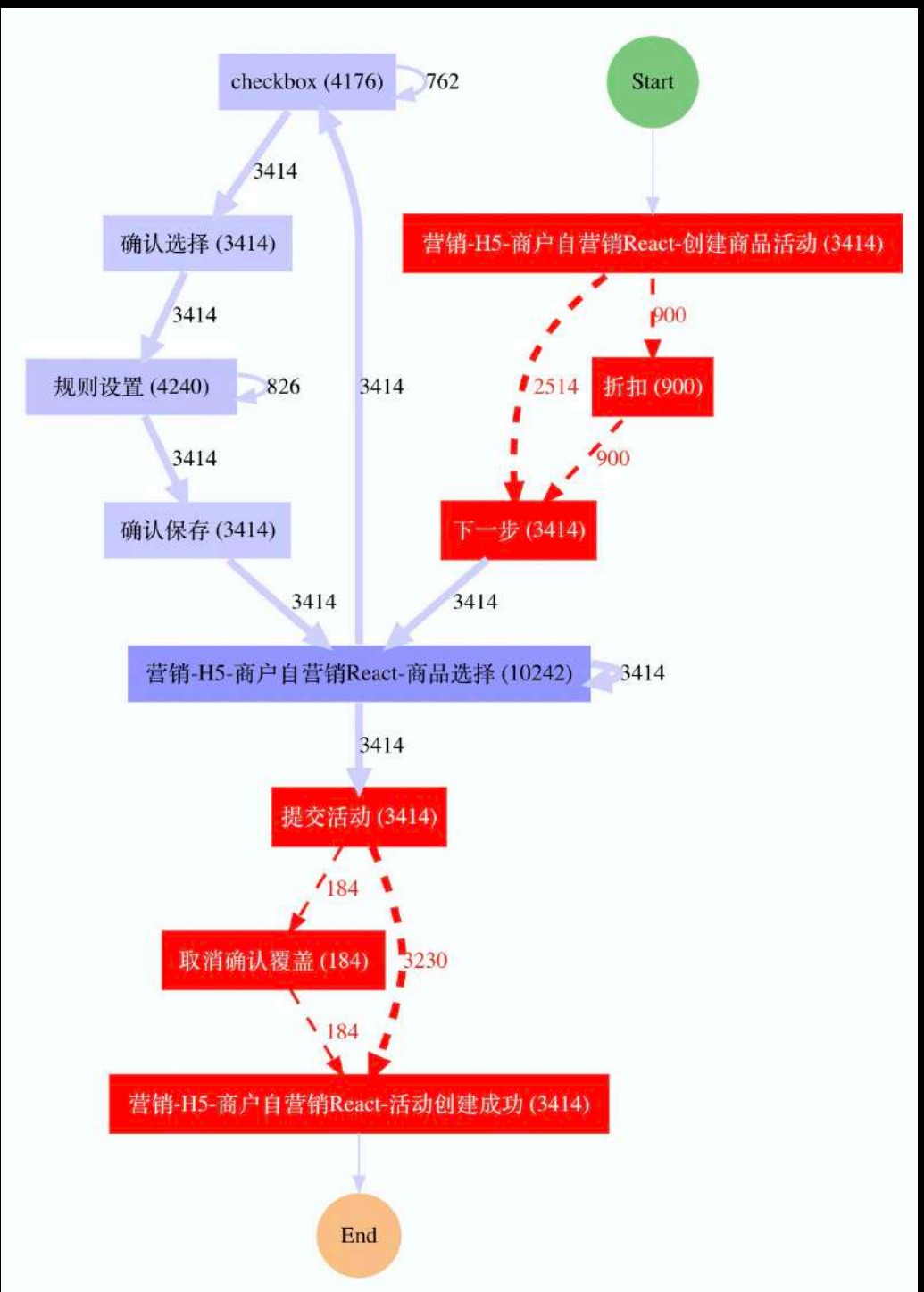
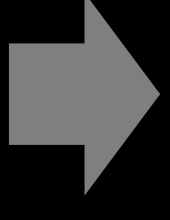


图 G



```
{
  nodes: [
    { id: "A" },
    { id: "B" },
    { id: "C" }
  ],
  edges: [
    { source: "A", target: "B" },
    { source: "B", target: "C" },
    { source: "A", target: "C" }
  ],
}
```



找到 G 中所有图 Q 的匹配





D2 前端技术论坛  
D2 FRONTEND TECHNOLOGY FORUM

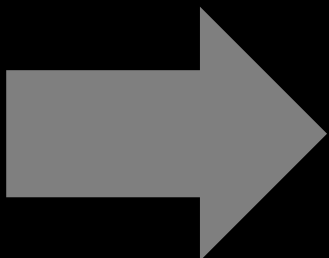
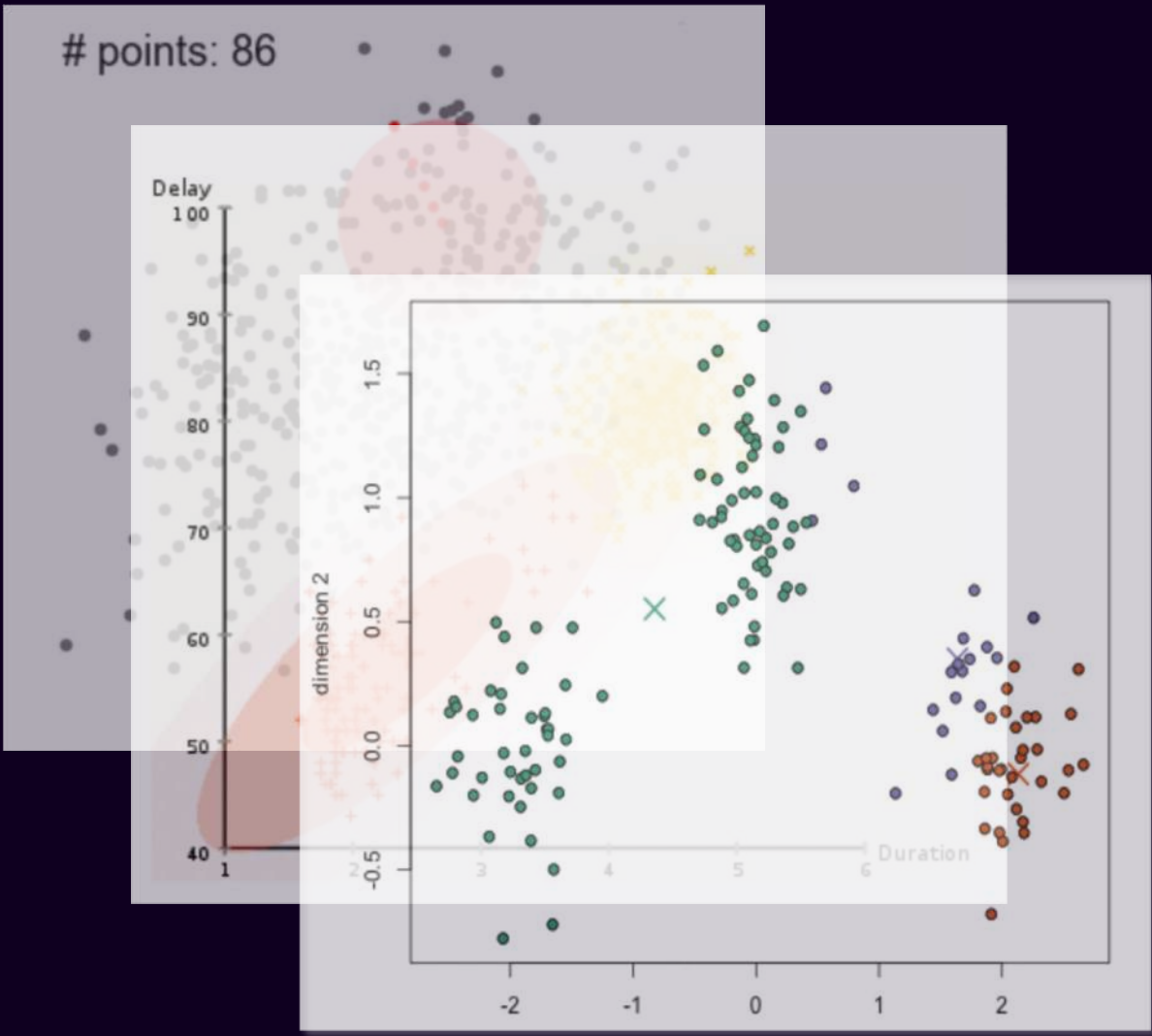
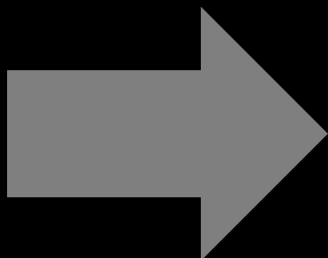
精心

06

未来展望



聚类



- 场景发现
- 动机分析
- 异常侦测



Thanks





Wx



Dingding