ICACT 2019 TUTORIAL: 16:00 ~ 18:00, 2019. 02. 17

# WORKFLOW AND BUSINESS PROCESS MINING


**PROF. KWANGHOON PIO KIM, Ph.D.**

Collaboration Technology Research Lab.
Division of Computer Science and Engineering
KYONGGI UNIVERSITY, REPUBLIC OF KOREA

# CONTENTS

# INTRODUCTION

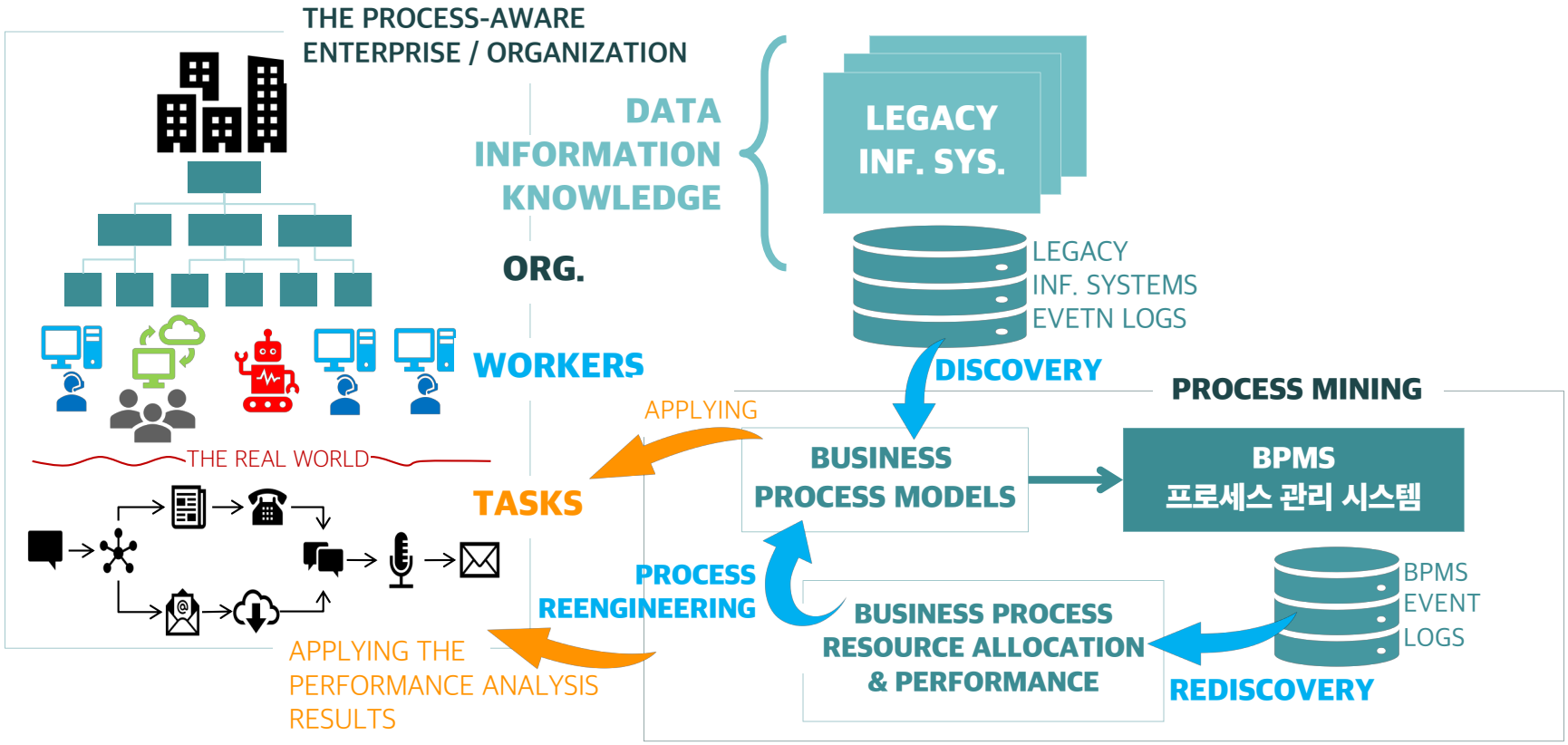
## 1) BACKGROUND

- ENTERPRISE-WIDE PROCESS AUTOMATION:

**BUSINESS PROCESS MANAGEMENT SYSTEMS**

  - BUSINESS PROCESS MODELING (BPM MODELING SYSTEMS) METHODOLOGIES
    - PROCCESS MODELING METHODOLOGIES: INFORMATION CONTROL NETS AND PETRI NETS
  - BUSINESS PROCESS ENACTMENT (BPM ENGINE) ARCHITECTURES
    - ENACTMENT EVENT LOG FORMATS: XWELL, BPAF, **IEEE XES** (Extensible Event Stream)

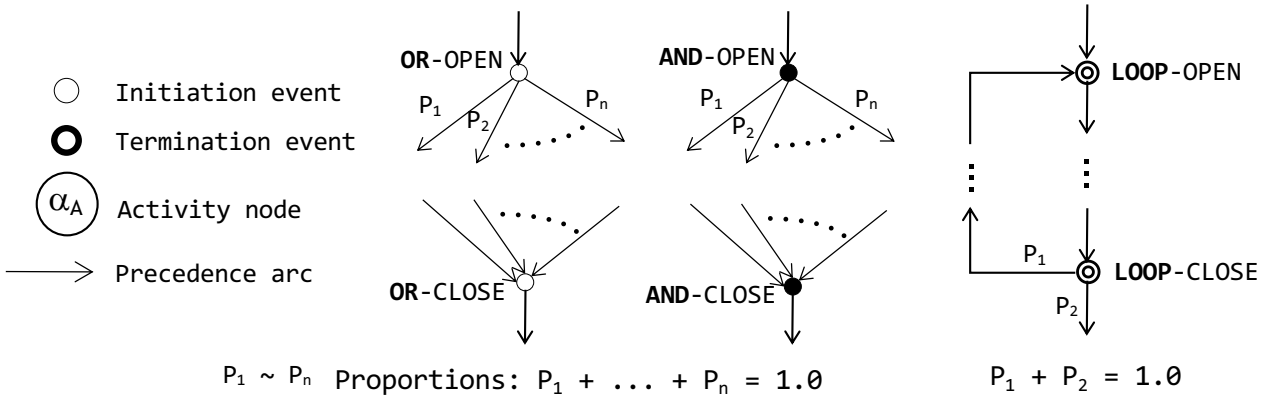
- THE RECENT RESEARCH ISSUE: **PROCESS MINING** - DISCOVERY & REDISCOVERY



# INTRODUCTION

## 2) MOTIVATION

- **THE ISSUE: WORKFLOW AND BUSINESS PROCESS MINING**
  - TECHNOLOGY: BIGDATA ANALYSIS BY USING THE PROCESS MINING TECHNIQUES
  - MANAGEMENT: PROCESS LIFECYCLE ( **발견**⋯›분석⋯›설계⋯›실행⋯›이력⋯›**재발견**⋯› ⋯ ) MANAGEMENT
- **IN THIS TUTORIAL, PROCESS MINING TECHNQUES**
  - PROCESS REDISCOVERY TECHNIQUES:
    - PETRINET-BASED: α-알고리즘(alpha-Algorithm: 순차적/선택적/병렬적 패턴 재발견)
    - ICN-BASED: σ-알고리즘(sigma-Algorithm: 순차적/선택적/병렬적 패턴 재발견)
    - THE OUR APPROACH: **정보제어넷기반 구조적 프로세스 재발견 기법**
      - **ρ-ALGORITHM(rho-Algorithm): DISCOVERING STRUCTURED PROCESSES**
  - PROCESS-AWARE KNOWLEDGE DISCOVERY:
    - **THE OUR APPROACH: PROCESS PROPORTIONS DISCOVERY**
      - **PROPORTIONAL ICN-BASED PROCESS PATTERN DISCOVERY**



# INTRODUCTION

## 3) THE SCOPE

- **PROCESS DISCOVERY ALGORITHM**
  - PROPOSING A PROCESS PATTERN REDISCOVERY APPROACH
- **PROCESS-AWARE KNOWLEDGE DISCOVERY**
  - PROPOSING A PROPORTION DISCOVERY TECHNIQUE ON PROCESS PATTERNS
- **A SERIES OF EXPERIMENTS**
  - AN EXPERIMENTAL STUDY ON A GROUP OF PROCESS EVENT LOG DATASETS

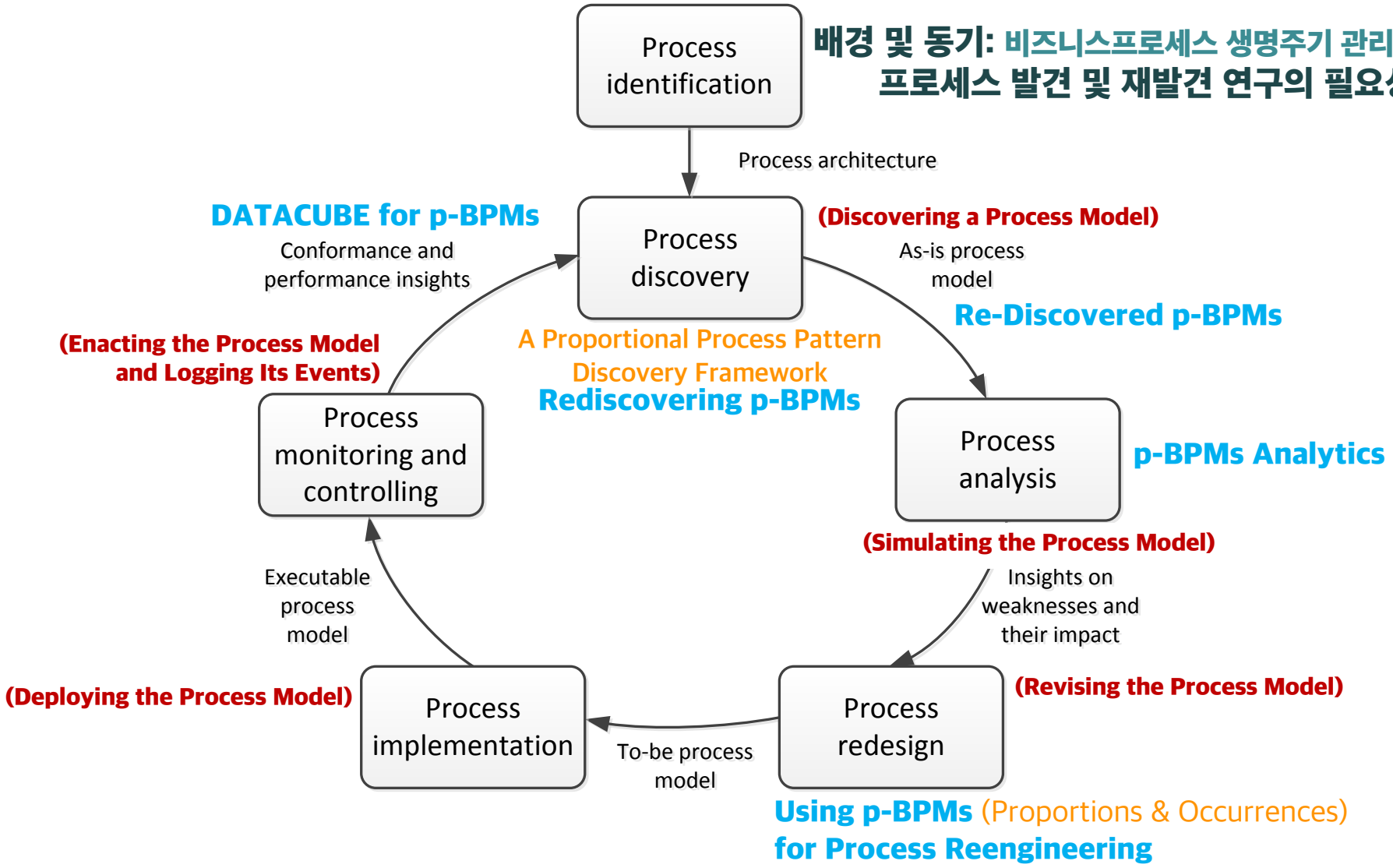
## 4) THE GOALS OF OUR PROCESS MINNG RESEARCH

- **A PROPORTIONAL PROCESS PATTERN DISCOVERY FRAMEWORK**
  - **REDISCOVERING A COMPLETE MODEL OF STRUCTURED INFORMATION CONTROL NETS**
  - **DISCOVERING PROPORTIONS ON PROCESS PATTERNS**
  - **IMPLEMENTING THE PROPOSED FRAMEWORK**
- **AN EXPERIMENTAL STUDY ON AN IEEE XES-FORMATTED DATASET**
  - **ACQUIRING PROCESS EVENT LOG DATASETS**
  - **CARRYING OUT A SERIES OF EXPERIMENTAL ANALYSES**

## 연구의 배경 및 동기: THE PROPORTIONAL BUSINESS PROCESS MODEL

# ON THE BUSINESS PROCESS LIFE-CYCLE

**AFTER COMPLETING THE CYCLE ONCE, WE REDISCOVER A PROPORTIONAL BUSINESS PROCESS MODEL**



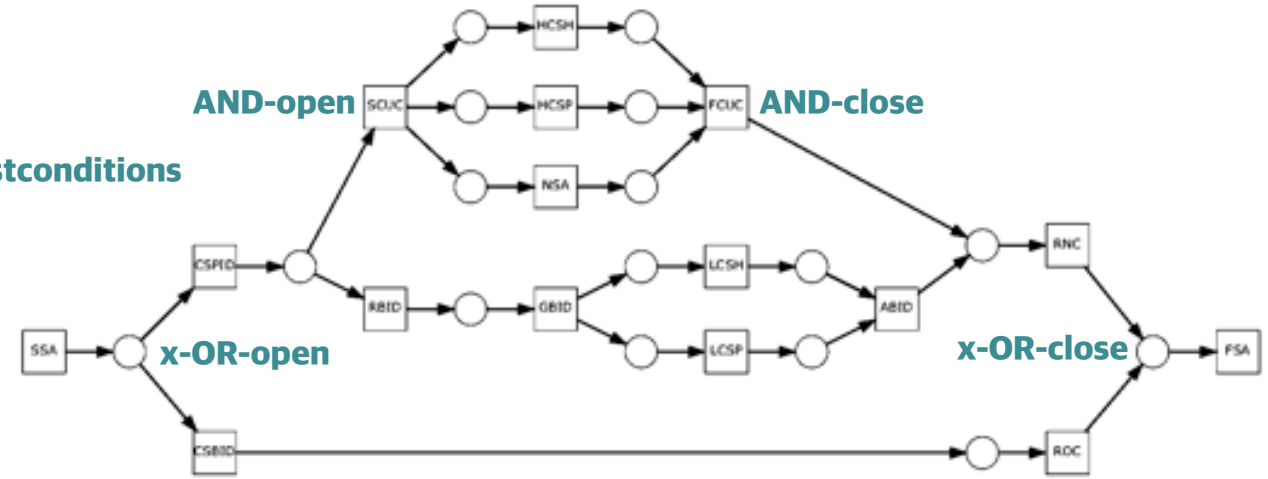
# 연구의 범위: 이론적 측면의 재발견 프로세스 모델

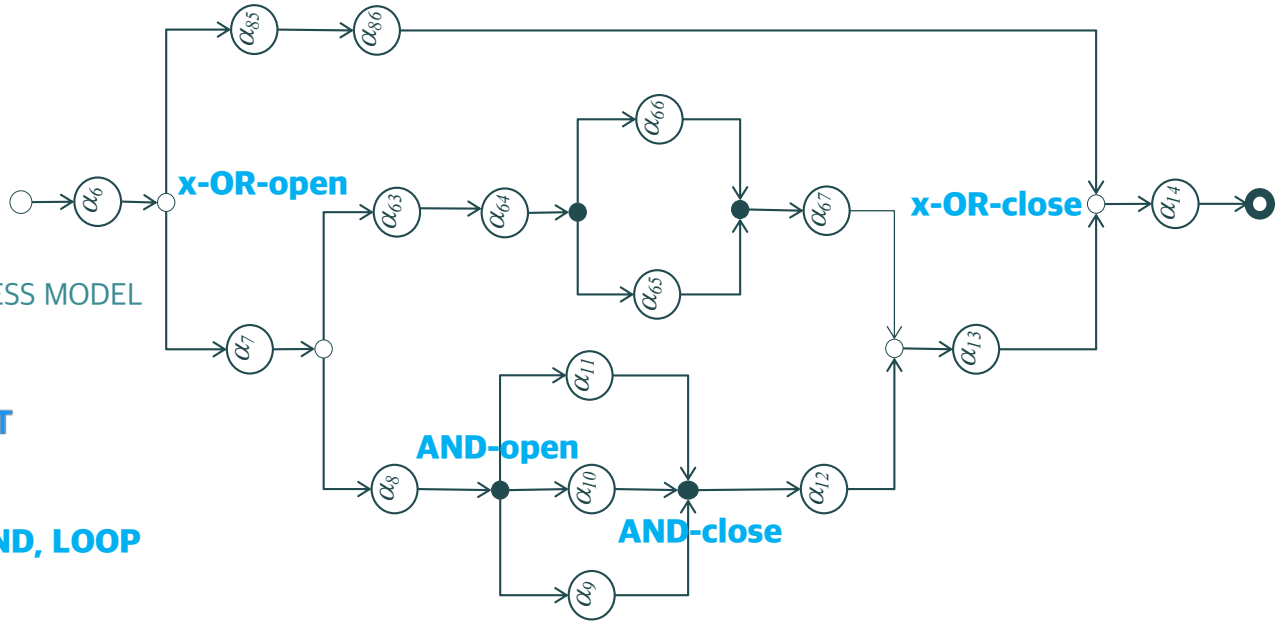
## STRUCTURED INFORMATION CONTROL NET MODELS

# STRUCTURED INFORMATION CONTROL NETS

**PETRINET(WF-NET)**

- **Place : Preconditions/Postconditions**
- **Transition: Activity**




A STRUCTURED ICN-BASED PROCESS MODEL

- **Matched Pairing** Property
- **Proper Nesting** Property

**INFORMATION CONTROL NET**

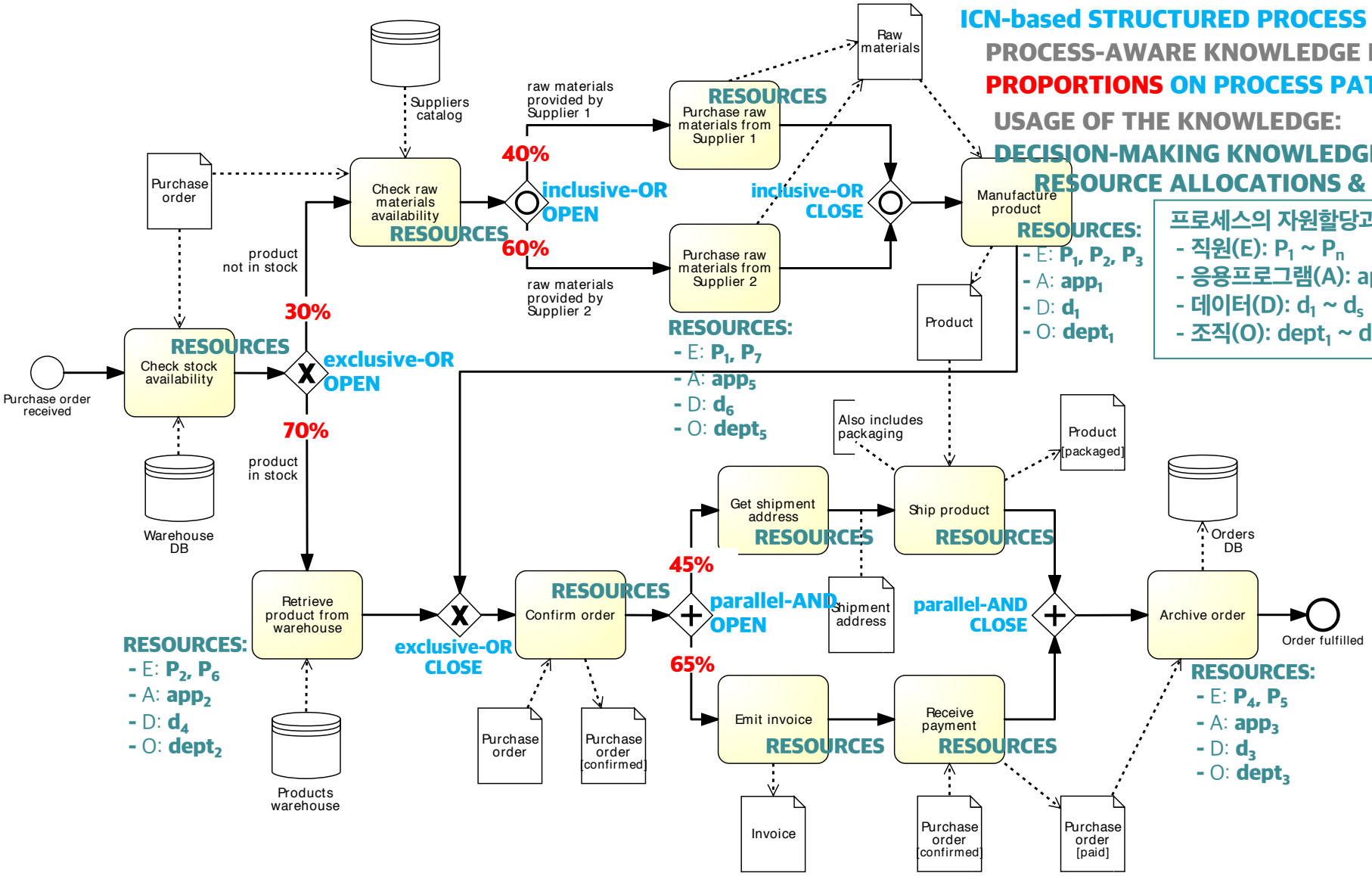
- **Event Activity: start, end**
- **Performer Activity**
- **Gateway Activity: XOR, AND, LOOP**

# 연구의 범위: 프로세스 패턴상의 실행비율 지식 발견

# A PROPORTIONAL BUSINESS PROCESS MODEL

**PROCESS REDISCOVERY:**
**ICN-based STRUCTURED PROCESS PATTERNS**
**PROCESS-AWARE KNOWLEDGE DISCOVERY:**
**PROPORTIONS ON PROCESS PATTERNS**
**USAGE OF THE KNOWLEDGE:**
**DECISION-MAKING KNOWLEDGE for RESOURCE ALLOCATIONS & PLANNINGS**

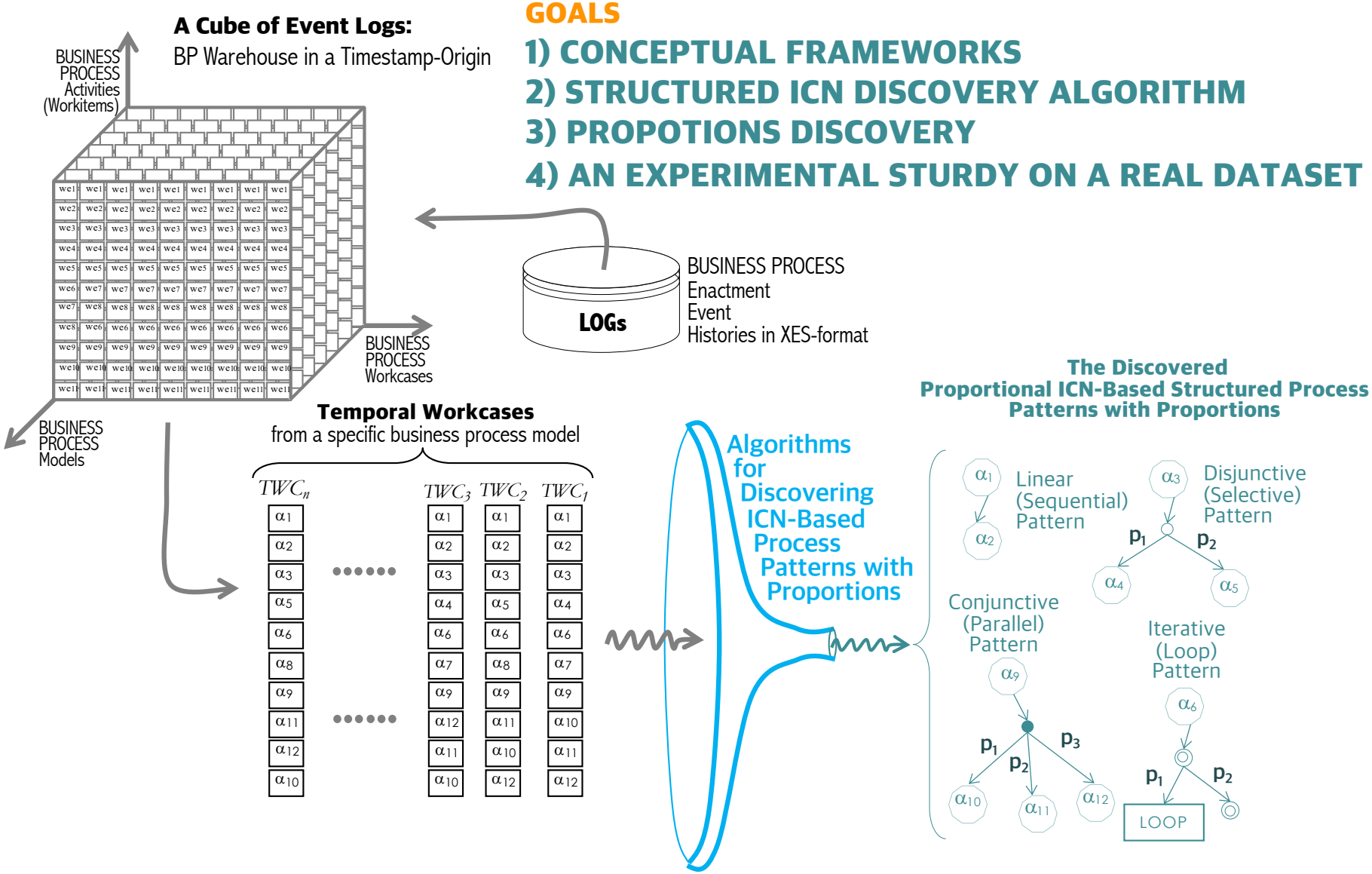
프로세스의 자원할당과 계획:
- 직원(E): $P_1 \sim P_n$
- 응용프로그램(A): $app_1 \sim app_m$
- 데이터(D): $d_1 \sim d_s$
- 조직(O): $dept_1 \sim dept_t$




BPMN: BUSINESS PROCESS MODEL AND NOTATION

**A PURCHASE ORDER BUSINESS PROCESS MODEL**

# GOAL: A FRAMEWORK FROM CONCEPTUAL TO CONCRETE
# PROPORTIONAL PROCESS PATTERN DISCOVERY AND ANLYTICS



# RELATED WORKS

# THE RELATED WORKS

## ❖ PROCESS DISCOVERY AND REDISCOVERY ALGORITHMS

- ✓ 페트리넷기반 프로세스 재발견 알고리즘
  - α-ALGORITHM: 페트리넷기반 프로세스 재발견 알고리즘
- ✓ 정보제어넷기반 프로세스 재발견 알고리즘
  - σ-ALGORITHM: 정보제어넷기반 프로세스 재발견 알고리즘
- ✓ IN THIS RESEARCH, PROPOSING A PROESS MINING ALGORITHM
  - ρ-ALGORITHM[주1)]: 구조적 정보제어넷기반 프로세스 패턴 재발견 알고리즘

## ❖ PROCESS-AWARE KNOWLEDGE DISCOVERY APPROACHES

- ✓ STOCHASTIC APPROACHES
  - 시뮬레이션 접근방법에 의한 프로세스 구조분석 및 실행 비율 예측/추정 분석
  - 이론: **STOCASTIC BUSINESS PROCESS MODEL**
- ✓ CONTROL-PATH DISCOVERY APPROACHES
  - 비즈니스 프로세스 모델의 제어경로 발견과 실행로그로부터의 제어경로 재발견
- ✓ PROCESS-AWARE SOCIAL NETWORK DISCOVERY APPROCHES
  - 프로세스 실행이력으로부터 엔터프라이즈 소셜네트워크 발견
- ✓ IN THIS RESEARCH, PROCESS-AWARE KNOWLEDGE DISCOVERY APPROACH
  - DISCOVERING PROCESS ENACTMENT PROPORTIONS

주1) - **ρX** *gives the number of elements in* **X**.

ρ(Rho) 심볼의 의미가 APL(A Programming Language, 1960s)에서 원소의 개 수를 나타내는 것에 착안하였습니다. 본 학위논문에서 제안하는 프로세스 패턴 재발견 알고리즘의 핵심 아이디어가 실행 개 수를 기반으로 패턴의 유형을 결정하는 방법이고, 그 의미가 동일하여 알고리즘의 이름으로 정한 것입니다.

# α(α⁺)-알고리즘: 페트리넷기반 프로세스 재발견

W. van der Aalst, T. Weijters, L. Maruster, "Workflow mining: discovering process models from event logs," IEEE Transactions on Knowledge and Data Engineering, Volume: 16 , Issue: 9 , pp. 1128 - 1142, Sept. 2004

## ❖ 페트리넷기반의 세가지 프로세스 패턴

1) 순차적 패턴: SEQUENCIAL PATTERN
2) 병렬적 패턴: PARALLEL-SPLIT PATTERN and PARALLEL-JOIN PATTERN
3) 선택적 패턴: SELECTIVE-SPLIT PATTERN and SELECTIVE-JOIN PATTERN

## ❖ 기본적인 시간적 우선순위에 따른 한 쌍의 액티버티들간의 관계기호 정의

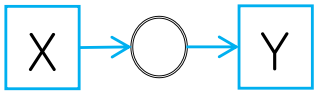
**1) Direct succession: x>y**
**iff** for some case **x** is directly followed by **y**

**2) Causality(Sequential): x→y**
**iff x>y** and **not y>x**

**3) Parallel: x||y**
**iff x>y** and **y>x**

**4) Choice(Selective): x#y**
**iff not x>y** and **not y>x.**

## 1) 순차적 프로세스 패턴

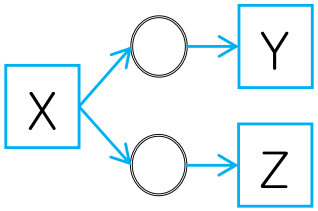



: X → Y

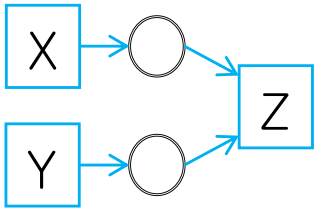
시간적 우선순위를 갖는 두 단위업무의 순차적 액티버티들
➔ 순차적 프로세스 패턴의 페트리넷 모델

## 2) 병렬적 프로세스 패턴

트랜지션의
다중 출력-플레이스
➔ 병렬적 패턴




: X → Y, X→ Z, AND Y||Z

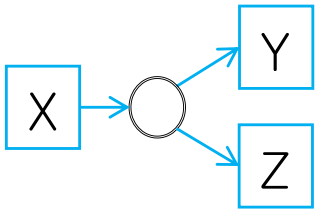



: X → Z, Y→ Z, AND X||Y

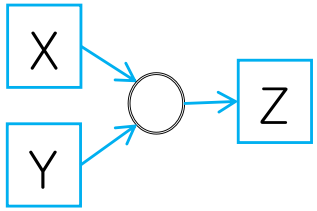
같은 source액티버티를 갖는 두 쌍의 액티버티들이 존재하고, 그의 destination액티버티들이 결정적 시간적 우선순위를 갖지 못한다면,
➔ 병렬적 프로세스 패턴의 페트리넷 모델 (OPEN-parallel), Close-parallel은 이의 반대

## 3) 선택적 프로세스 패턴

플레이스의
다중 출력-트랜지션
➔ 선택적 패턴




: X → Y, X→ Z, AND Y#Z




: X → Z, Y→ Z, AND X#Y

같은 source액티버티를 갖는 두 쌍의 액티버티들이 존재하고, 그의 destination액티버티들이 동시에 존재하지 않는다면,
➔ 선택적 프로세스 패턴의 페트리넷 모델 (OPEN-selective), Close-selective는 이의 반대

# The α-Algorithm

**Let L be an event log over T. α(L) is defined as follows:**

1. $T_L = \{ t \in T \mid \exists_{\sigma \in L}\ t \in \sigma \}$,
   //로그-트레이스상의 각 액티버티는 곧 모델상의 트랜지션에 해당한다.
2. $T_I = \{ t \in T \mid \exists_{\sigma \in L}\ t = first(\sigma) \}$,
   //로그-트레이스상의 첫번째 액티버티를 정의한다.
3. $T_O = \{ t \in T \mid \exists_{\sigma \in L}\ t = last(\sigma) \}$,
   //로그-트레이스상의 마지막번째 액티버티를 정의한다.
4. $X_L = \{(A,B) \mid A \subseteq T_L \wedge A \neq \emptyset \wedge B \subseteq T_L \wedge B \neq \emptyset \wedge \forall_{a \in A}\ \forall_{b \in B}\ a \rightarrow_L b \wedge \forall_{a1,a2 \in A}\ a_1 \#_L a_2 \wedge \forall_{b1,b2 \in B}\ b_1 \#_L b_2 \}$,
   //로그-트레이스상의 모든 액티버티 쌍을 구한다.
5. $Y_L = \{(A,B) \in X_L \mid \forall_{(A',B') \in XL}\ A \subseteq A' \wedge B \subseteq B' \Rightarrow (A,B) = (A',B') \}$,
   //로그-트레이스상의 모든 액티버티 쌍에서 중복되는 쌍을 제거한다.
6. $P_L = \{ p_{(A,B)} \mid (A,B) \in Y_L \} \cup \{i_L, o_L\}$,
   //페트리넷기반 프로세스 구조를 형성하기 위해 시작-플레이스와 종료-플레이스를 추가한다.
7. $F_L = \{(a, p_{(A,B)}) \mid (A,B) \in Y_L \wedge a \in A \} \cup \{(p_{(A,B)}, b) \mid (A,B) \in Y_L \wedge b \in B \} \cup \{(i_L, t) \mid t \in T_I \} \cup \{(t, o_L) \mid t \in T_O \}$, and
   //페트리넷기반 프로세스 모델을 구성하는 트랜지션들간의 시간적 우선순위에 따른 제어흐름을 결정한다.
8. $\alpha(L) = (P_L, T_L, F_L)$.
   //프로세스 모델(T)의 실행이밴트로그(L)로부터 재발견한 페트리넷기반 프로세스 모델을 정의한다.

# An Example of the α-Algorithm

**EVENT LOGS (L)**
**of PROCESS MODEL (T)**

**L:**

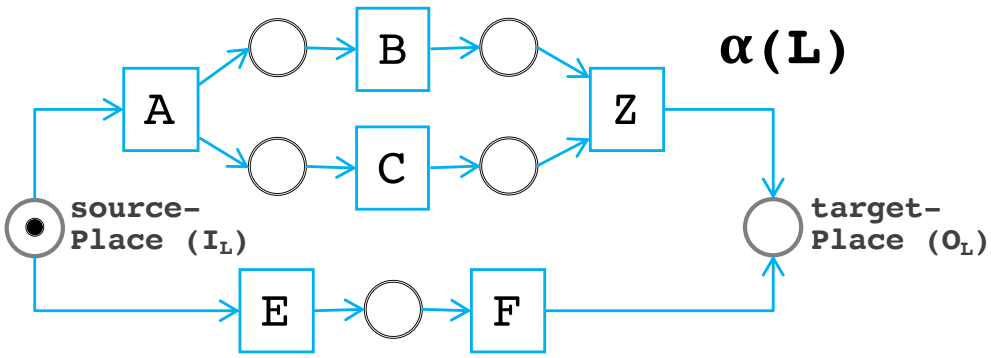
```
case 1 : task A
case 2 : task A
case 3 : task A
case 3 : task B
case 1 : task B
case 1 : task C
case 2 : task C
case 4 : task A
case 2 : task B
case 2 : task D
case 5 : task E
case 4 : task C
case 1 : task D
case 3 : task C
case 3 : task D
case 4 : task B
case 5 : task F
case 4 : task D
```

A>B
A>C
**B>C**
B>D
**C>B**
C>D
E>F

A→B
A→C
B→D
C→D
E→F

**B||C**
**C||B**

**A REDISCOVERED PETRINET-BASED PROCESS MODEL**




액티버티 B와 C는
시간적 우선순위를 결정할 수 없음.
왜냐하면, 각기 다른 case에서의 우선순위가 동일하게 나타나지 않기 때문이다.
그러므로 액티버티 B와 C는 병렬적 관계이다.

**Direct succession:**
**x>y iff for some case**
**x is directly followed by y**

- Causality: x→y iff x>y and not y>x
- **Parallel: x||y iff x>y and y>x**
- Choice: x#y iff not x>y and not y>x.

# Summary of the α-Algorithm

## ❖ Properties of the α-Algorithm

- 페트리넷 모델 기반 프로세스 마이닝 알고리즘
- 재발견 프로세스 패턴: 순차적 패턴, 선택적 패턴, 병렬적 패턴
  - 반복적 패턴 (LOOP PROCESS PATTERN): 선택적 패턴에 포함되어 재발견 되고 있음.
- 프로세스 재발견 핵심 원리
  - 한 쌍의 액티버티의 실행로그시간의 선후행 관계(→)와 로그의 동시존재여부(|| or #) 관계를 기반으로 패턴의 유형(순차적, 선택적, 병렬적)을 결정하는 방법

## ❖ α-Algorithm과 ρ-Algorithm의 차별성

- 정보제어넷 기반 프로세스 마이닝 알고리즘
- 재발견 프로세스 패턴: 순차적 패턴, 선택적 패턴, 병렬적 패턴, 반복적 패턴
  - 구조적 프로세스 패턴을 재발견: 선택적 패턴과 반복적 패턴을 분리하여 재발견하고 있음.
- 프로세스 재발견 핵심 원리
  - 한 쌍의 액티버티의 실행로그시간의 선후행 관계와 로그의 개 수 크기비교(<, ==, >) 관계를 기반으로 패턴의 유형(순차적, 선택적, 병렬적, 반복적)을 결정하는 방법

# σ-알고리즘: 정보제어넷기반 프로세스 재발견

Kim K., Ellis C.A., "σ-Algorithm: Structured Workflow Process Mining Through Amalgamating Temporal Workcases," In: Zhou ZH., Li H., Yang Q. (eds) Advances in Knowledge Discovery and Data Mining. PAKDD 2007. Lecture Notes in Computer Science, vol. 4426, pp. 119-130, Springer, Berlin, Heidelberg, 2007

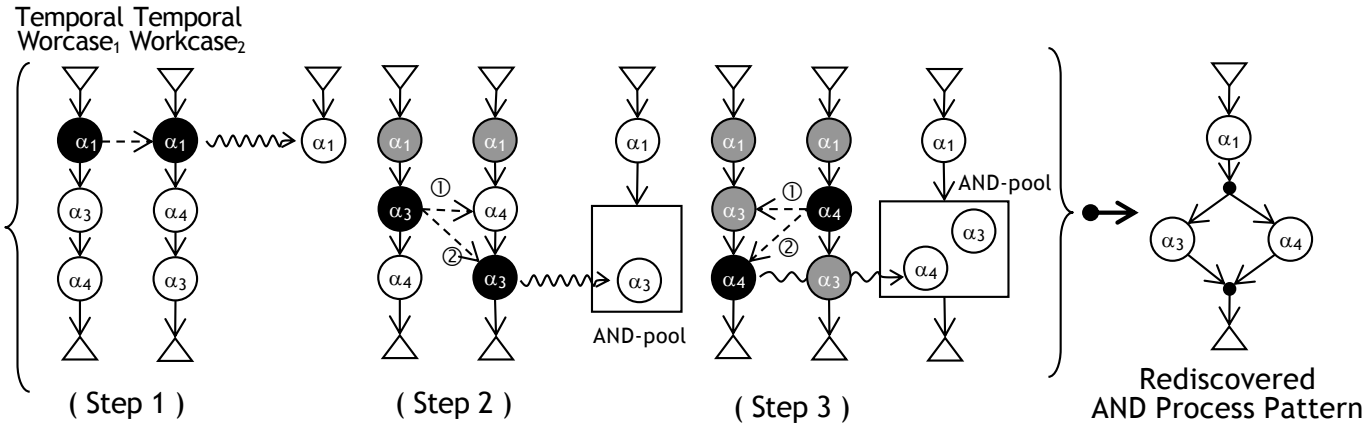
## ❖ 3 BASIC ICN-BASED PROCESS PATTERNS

### 1) SEQUENCIAL PATTERN

### 2) PARALLEL-SPLIT PATTERN and PARALLEL-JOIN PATTERN

병합할 두 개의 다른 워크케이스에 속한 인접-액티버티들의 시간적 우선순위가 서로 다를 때

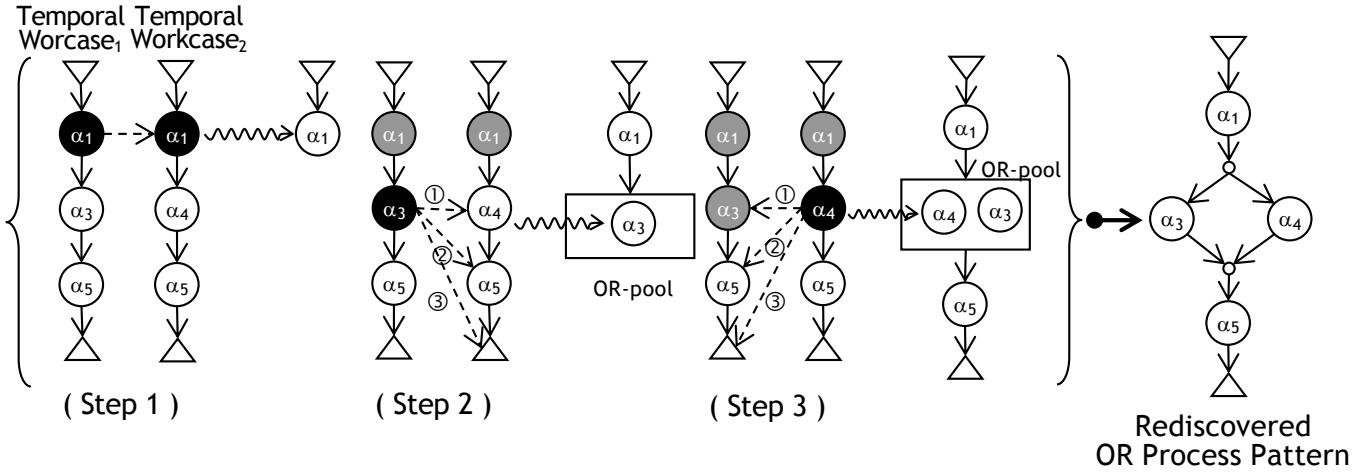
**AND-POOL**




### 3) SELECTIVE-SPLIT PATTERN and SELECTIVE-JOIN PATTERN

병합할 두 개의 다른 워크케이스에 속한 인접-액티버티들의 존재여부가 서로 다를 때

**OR-POOL**



# An Example of the σ-Algorithm

**PROCESS LOGS**

**L:**

| | |
|---|---|
| case 1 | task A |
| case 2 | task A |
| case 3 | task A |
| case 3 | task B |
| case 1 | task B |
| case 1 | task C |
| case 2 | task C |
| case 4 | task A |
| case 2 | task B |
| case 2 | task D |
| case 5 | task E |
| case 4 | task C |
| case 1 | task D |
| case 3 | task C |
| case 3 | task D |
| case 4 | task B |
| case 5 | task F |
| case 4 | task D |

**TEMPORAL WORKCASES**

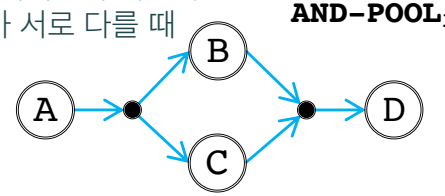
$temporal\_workcase_1$: A → B → C → D
$temporal\_workcase_2$: A → C → B → D
$temporal\_workcase_3$: A → B → C → D
$temporal\_workcase_4$: A → C → B → D
$temporal\_workcase_5$: E → F

Temporal Workcase들에 대한 클러스터링을 수행한 후에 각 그룹을 병합하는 Sigma-Algorithm을 적용하면, 알고리즘의 효율성을 크게 향상시킬 수 있음.

**AMALGAMATING TEMPORAL WORKCASES**

1. $temporal\_workcase_1$: A → [B → C] → D ⊗ $temporal\_workcase_2$: A → [C → B] → D
2. A → [B C] → D (**AND-POOL$_1$**) ⊗ $temporal\_workcase_3$: A → B → C → D
3. A → [B C] → D (**AND-POOL$_1$**) ⊗ $temporal\_workcase_4$: A → C → B → D
4. A → [B C] → D (**AND-POOL$_1$**) ⊗ $temporal\_workcase_5$: E → F

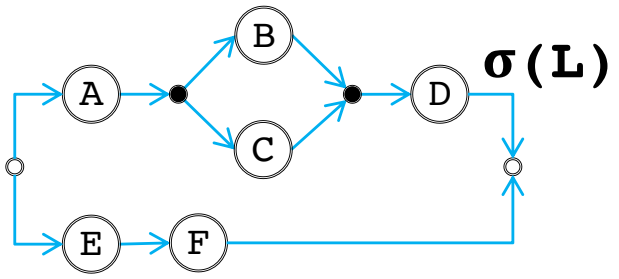
병합(⊗)할 두 개의 다른 워크케이스에 속한 인접-액티버티들의 시간적 우선순위가 서로 다를 때 → 병렬적 패턴 **AND-POOL$_1$**




병합할 두 개의 다른 워크케이스에 속한 인접-액티버티들의 존재여부가 서로 다를 때 → 선택적 패턴 **OR-POOL$_1$**

A → [B C] → D (**AND-POOL$_1$**)
E → F
**OR-POOL$_1$**

**A REDISCOVERED ICN-BASE PROCESS MODEL**



# σ-알고리즘과 SUMMARY

```
PROCEDURE SWPMRediscovery():

1:  Input : A Set of Temporal Workcases, ∀(wc[i], i = 1..m);
2:              where, wc[1] == START(▽), wc[m] == END(△);
3:  Output : (1) A Rediscovered Structured Workflow Process Model (SWPM), R = (δ, κ, I, O);
4:              - The Activity Set of SWPM, A = {α_1 .. α_n}, (wc[i], i = 1..m) ∈ A;
5:           (2) A Set of Workcase Models (WCMs), ∀ W = (ω, P, S);
6:
7:  Initialize : δ_i(START(▽)) ← {NULL};
8:               δ_o(END(△)) ← {NULL};
9:  PROCEDURE WPMRediscovery()
10: BEGIN
11:     WHILE ((wc[] ← readOneWorkcase()) ≠ EOF) DO
12:         i ← 1;
13:         WHILE (wc[i] ≠ END(△)) DO
14:             ω_o(wc[i]) ← wc[i + 1]; i ← i + 1; ω_i(wc[i]) ← wc[i − 1];
15:         END WHILE
16:         /* Rediscovering the temporary RWPM from the current WCM */
17:         FOR (i = 1; i < m; i + +;) DO
18:             IF (Is δ_o(wc[i]) an empty-set?) THEN
19:                 δ_o(wc[i]) ← ω_o(wc[i]); continue;
20:             END IF
21:             IF (isANDTransition(wc[i], ω_o(wc[i])) == TRUE) THEN
22:                 continue;
23:             END IF
24:             FOR (each set, a, of sets in δ_o(wc[i])) DO
25:                 SWITCH (checkupTransition(a, ω_o(wc[i])) DO
26:                     Case 'fixed transition':
27:                     Case 'sequential relationship':
28:                         δ_o(wc[i]) ← ω_o(wc[i]);
29:                         break;
30:                     Case 'conjunctive transition (AND-split)':
31:                         ANDset ← makeANDTransition(a, ω_o(wc[i]));
32:                         δ_o(wc[i]) ← δ_o(wc[i]) ∪ ANDset;
33:                         eliminatePreviousTransition(a, ω_o(wc[i]));
34:                         break;
35:                     Case 'disjunctive transition (OR-split)':
36:                         ORset ← makeORTransition(a, ω_o(wc[i]));
37:                         δ_o(wc[i]) ← δ_o(wc[i]) ∪ ORset;
38:                         eliminatePreviousTransition(a, ω_o(wc[i]));
39:                         break;
40:                     Default: /* Exceptions */
41:                         printErrorMessage();
42:                         break;
43:                 END SWITCH
44:             END FOR
45:         END FOR
46:     END WHILE
47:     finishupSWPM(); /* with its input-activity sets, (δ_i(wc[i]), i = 1..n)
48:                                                 and its transition-conditions */
49:     δ_i(α_1) ← {START(▽)}; δ_o(α_n) ← {END(△)};
50:     PRINTOUT
51:         (1) The Rediscovered Structured Workflow Process Model, SWPM, R = (δ, κ, I, O);
52:         (2) A Set of the Workcase Models, WCMs, ∀W = (ω, P, S);
53: END PROCEDURE
```

## ❖ Properties of the σ-Algorithm

- 정보제어넷 모델 기반 프로세스 마이닝 알고리즘
- 재발견 프로세스 패턴: 순차적 패턴, 선택적 패턴, 병렬적 패턴
  - 반복적 패턴 (LOOP PROCESS PATTERN): 별도의 LOOPCASES 개념 적용
  - 구조적 프로세스 모델의 재발견
- 프로세스 재발견 핵심 원리
  - 프로세스 인스턴스의 실행이력모델인 TEMPORAL WORKCASE들을 하나씩 합병시키면서 패턴의 유형을 결정하는 방법
  - TEMPORAL WORKCASE를 클러스터링시킴으로써 알고리즘의 효율성을 크게 향상시킬 수 있음.

## ❖ σ-Algorithm과 ρ-Algorithm의 차별성

- 동일성: 정보제어넷 기반 프로세스 마이닝 알고리즘
- 재발견 프로세스 패턴: 순차적 패턴, 선택적 패턴, 병렬적 패턴, 반복적 패턴
  - 구조적 프로세스 패턴을 재발견: 선택적 패턴과 반복적 패턴을 분리하여 재발견하고 있음.
- 프로세스 재발견 핵심 원리
  - 한 쌍의 액티버티의 실행로그시간의 선후행 관계와 로그의 개 수 크기비교 (<, ==, >) 관계를 기반으로 패턴의 유형(순차적, 선택적, 병렬적, 반복적)을 결정하는 방법

# 비즈니스 프로세스 모델 실행비율 추정확률 분석

Clarence A. Ellis, Kwanghoon Kim, Aubrey Rembert, Jacques Wainer, "Investigations on Stochastic Information Control Nets," Information Sciences, Volume 194, pp. 120-137, 1 July 2012

## ❖ A CONCEPT OF STOCHASTIC BUSINESS PROCESS MODEL

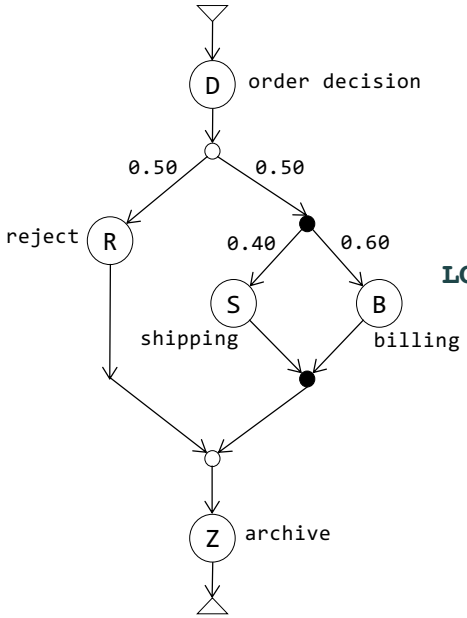
- 추정확률 프로세스 패턴 정의: 추정확률 정보제어넷 모델
  - PROPERLY NESTED: 모든 제어구조물들은 다른 제어구조물 내에 완전히 포함되어야 한다.
  - MATCHED PAIRED: BLOCK INITIATION NODES AND BLOCK TERMINATION NODES
    - Block Initiation nodes($n_i$): OR-SPLIT, AND-SPLIT, LOOP-SPLIT
    - Block Termination nodes($n_t$): OR-JOIN, AND-JOIN, LOOP-JOIN

      (1) $n_i \ll n_t$,

      (2) $d^{out}(n_i) + d^{in}(n_i) = d^{out}(n_t) + d^{in}(n_t)$,

      (3) $(n_i \ll n^*_i) \Rightarrow (n^*_t \ll n_t)$ where $(n_i, n_t)$ and $(n^*_i \ll n^*_t)$ are matched pairings,

      (4) Every execution sequence (path) that reaches $n_i$ must eventually reach $n_t$.
- PROBABILISTIC LANGUAGES AND MODEL-LOG COMPARISON(FIDELITY)




- Probabilistic Language generates DRZ, DSBZ, DBSZ.
  - MODEL: {[DRZ; 0.5],[DSBZ; 0.2],[DBSZ; 0.3]}
  - LOG: {[DRZ; 0.5],[DSBZ; 0.25],[DBSZ; 0.25]}
- **FIDELITY:** 모델에 의해 생성된 p-Language와 the p-Language 로그로부터 유도된 p-Language간의 거리(distance)를 수학적으로 계산한 것을 해당 모델의 충실도라고 정의한다. 거리가 0에 가까울수록 충실도가 좋은 것임.

**LOG OF CUSTOMER SEQUENCES**

| |
|---|
| CUSTOMER 1: DBSZ |
| CUSTOMER 2: DRZ |
| CUSTOMER 3: DSBZ |
| CUSTOMER 4: DRZ |

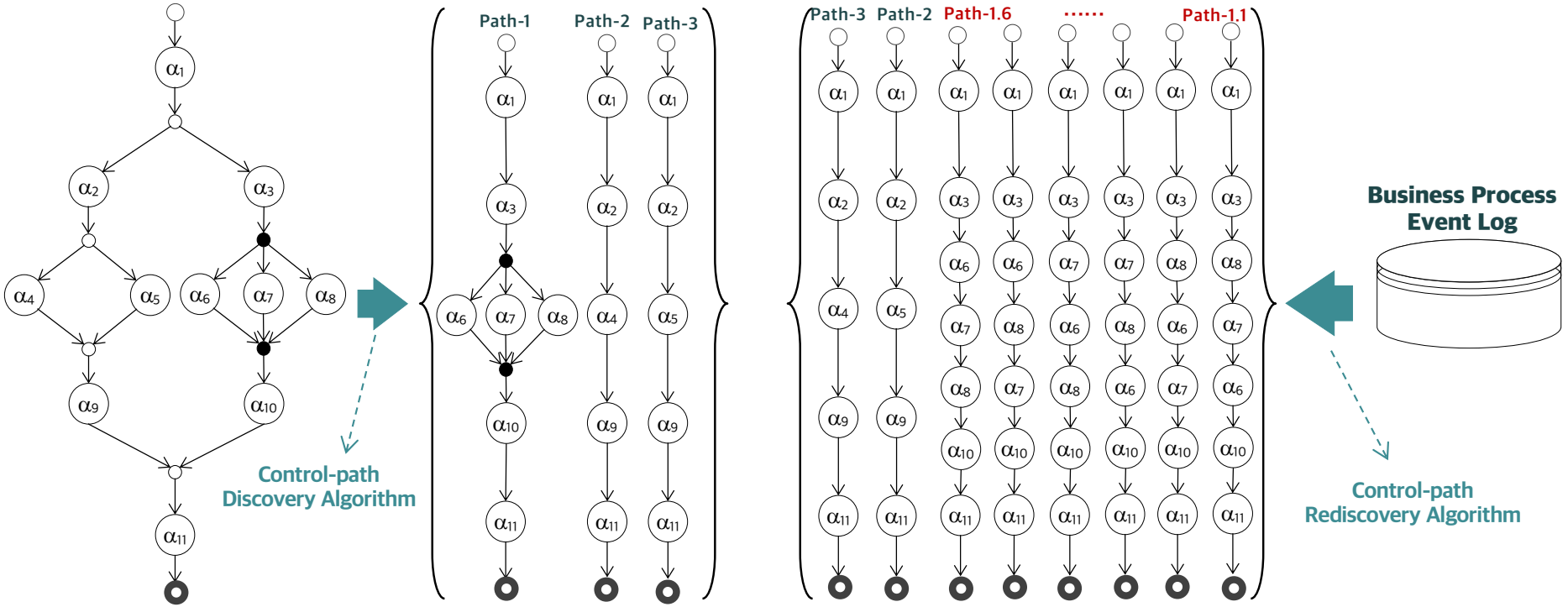
## 프로세스기반 비즈니스 지식 발견

# 프로세스 제어-경로 발견 및 재발견 접근방법

Minjae Park, Kwanghoon Kim, “Control-path Oriented Workflow Intelligence Analyses,” Journal of Information Science and Engineering, Volume 24, pp. 343-359, 2008

## ❖ A CONCEPT OF CONTROL-PATH KNOWLEDGE REDISCOVERY

- ○ 프로세스 제어경로 생성 알고리즘
- ○ 프로세스 제어경로 결정 트리 생성 알고리즘을 이용한 제어경로 재발견 알고리즘



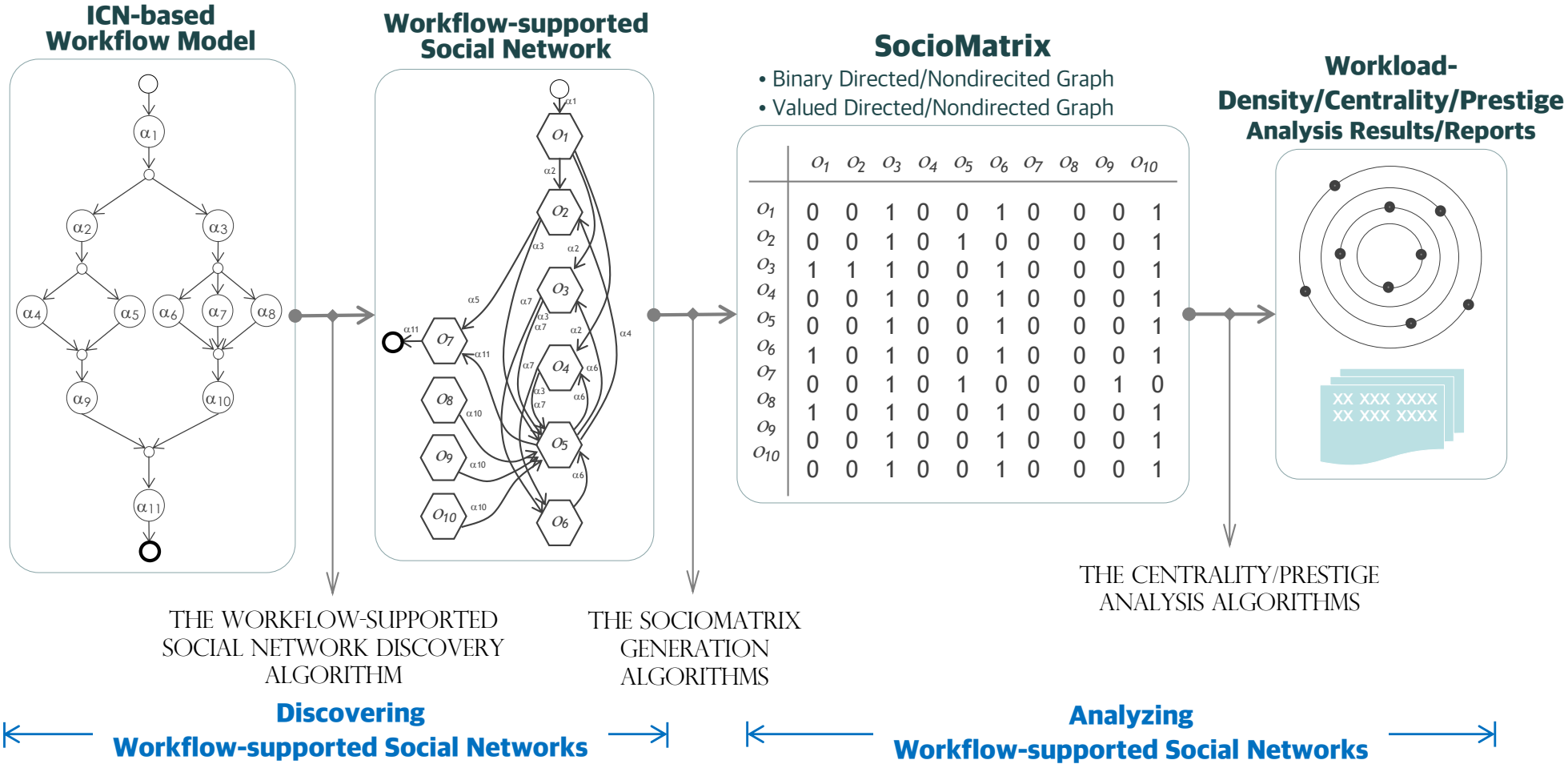
## 프로세스기반 비즈니스 지식 발견

# 프로세스기반 엔터프라이즈 소셜네트워크 발견 접근방법

Minjae Park, Hyun Ahn, Kwanghoon Kim, "Workflow-supported social networks: Discovery, analyses, and system," Journal of Network and Computer Applications, Volume 75, pp. 355-373, November 2016

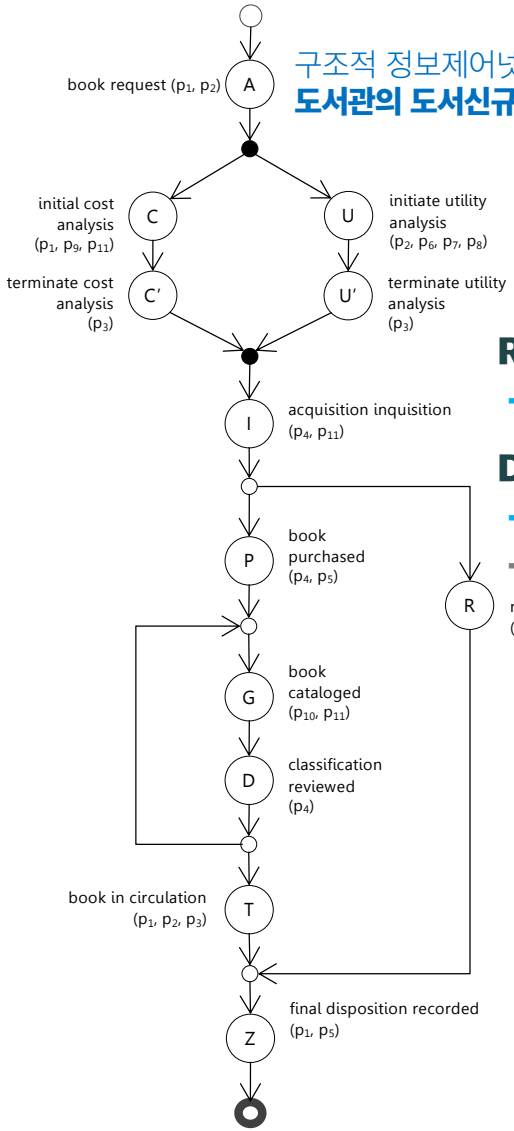
## ❖ A CONCEPT OF WORKFLOW-SUPPORTED SOCIAL NETWORK DISCOVERY

- ○ 워크플로우기반 소셜네트워크 지식 발견과 분석을 위한 이론적 프레임워크와 구현 시스템
- ○ 워크플로우기반 소셜네트워크 지식 발견 알고리즘



# A PROPORTINAL INFORMATION CONTROL NET MODEL

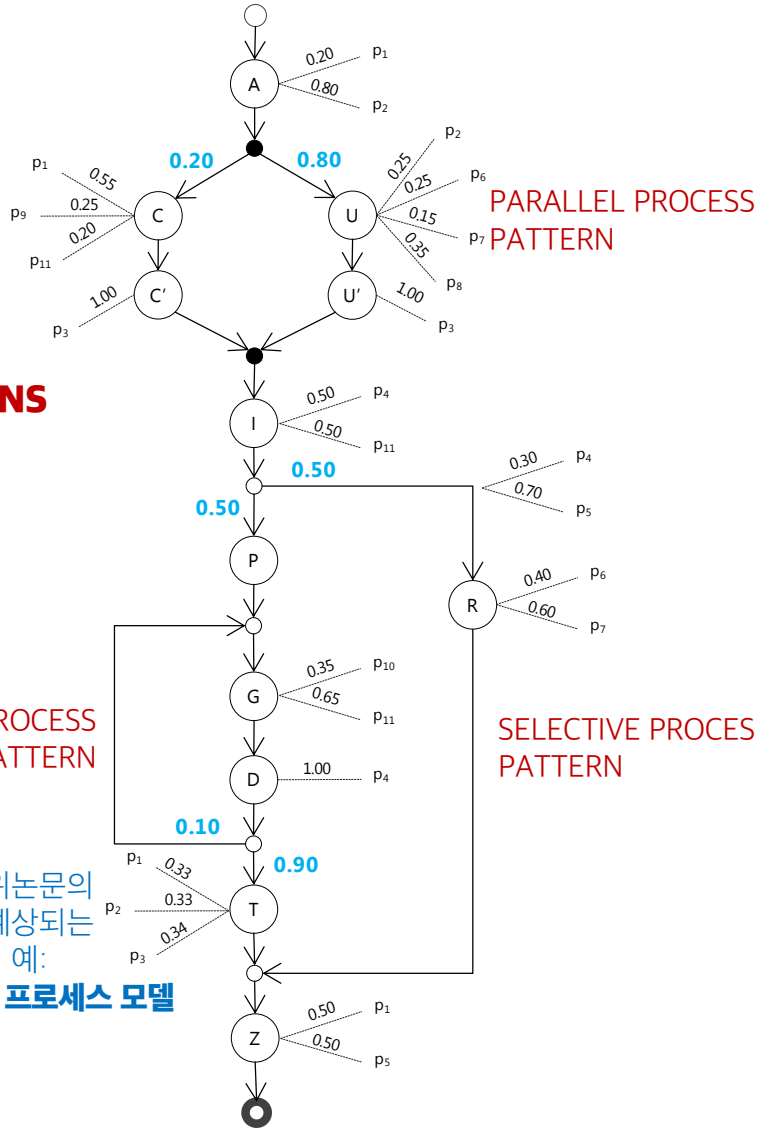
**A. ρ-Algorithm: 구조적 정보제어넷기반 프로세스 패턴 재발견 알고리즘**
**B. Proportions on Process Patterns: 프로세스 실행이력으로부터 프로세스 패턴의 실행비율 지식 발견**




구조적 정보제어넷기반 프로세스 모델의 대표적인 예:
**도서관의 도서신규구입 및 등록 프로세스 모델**

**REDISCOVERING ICN-BASED**
**- STRUCTURED PROCESS PATTERNS**

**DISCOVERING PROPORTIONS ON**
**- PROCESS PATTERNS**
**- PERFORMERS (FUTURE)**




PARALLEL PROCESS PATTERN

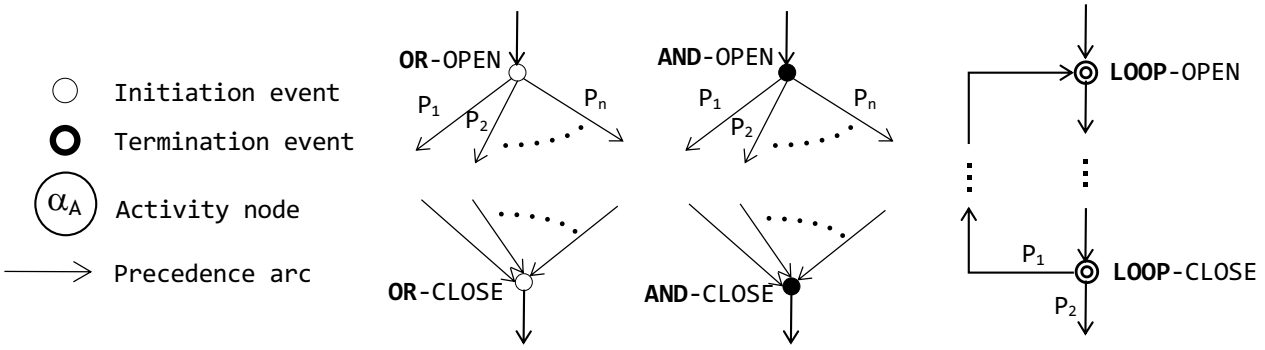
ITERATIVE PROCESS PATTERN

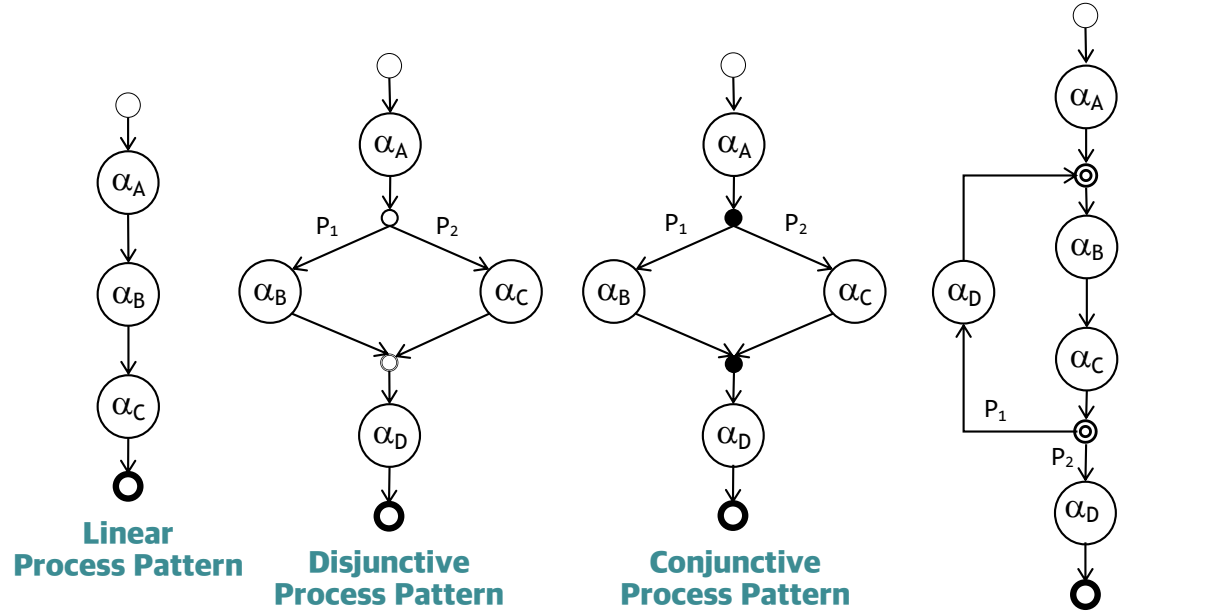
SELECTIVE PROCES PATTERN

해당 모델의 실행이력로그로부터 본 학위논문의 연구결과를 적용하여 재발견될 것으로 예상되는 구조적 정보제어넷기반 프로세스 모델의 예:
**재발견될 도서관의 도서신규구입 및 등록 프로세스 모델**

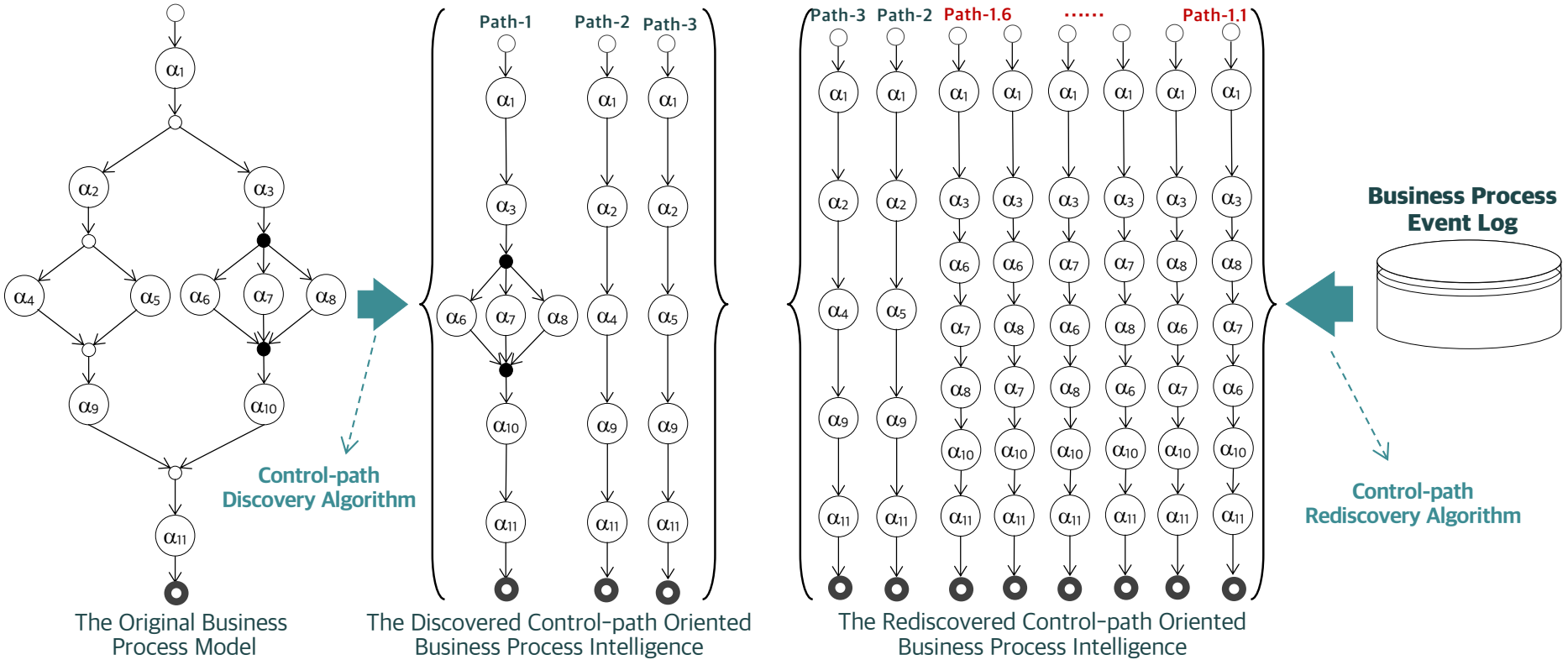
# A PROPORTIONAL PROCESS PATTERN DISCOVERY FRAMEWORK

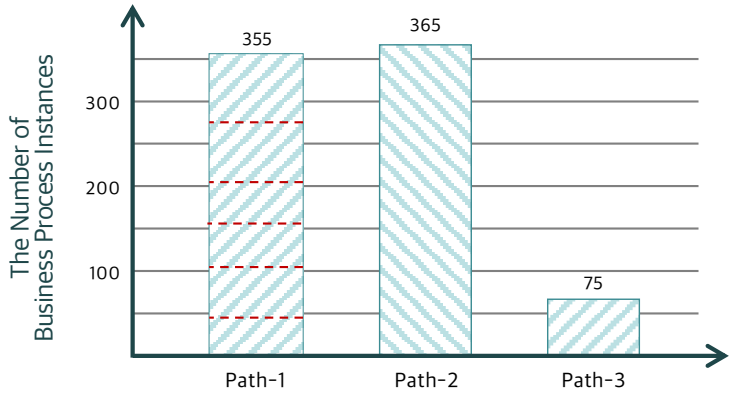
# STRUCTURED ICN-BASED PROPORTIONAL PROCESS PATTERNS

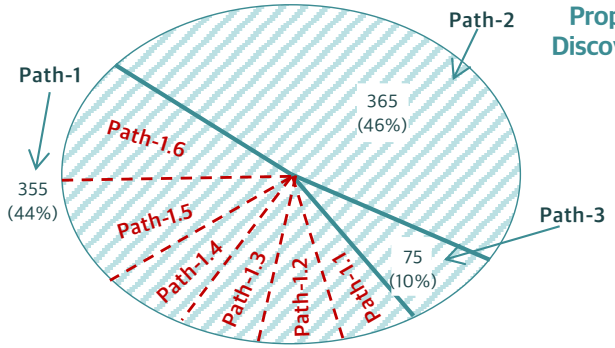


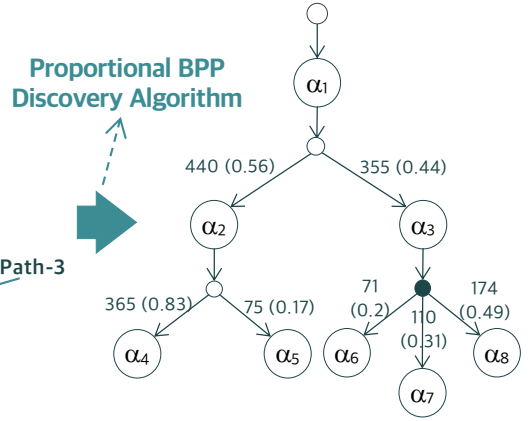



# A CONCEPTUAL INITIATION OF PROPORTIONAL BUSINESS PROCESSES



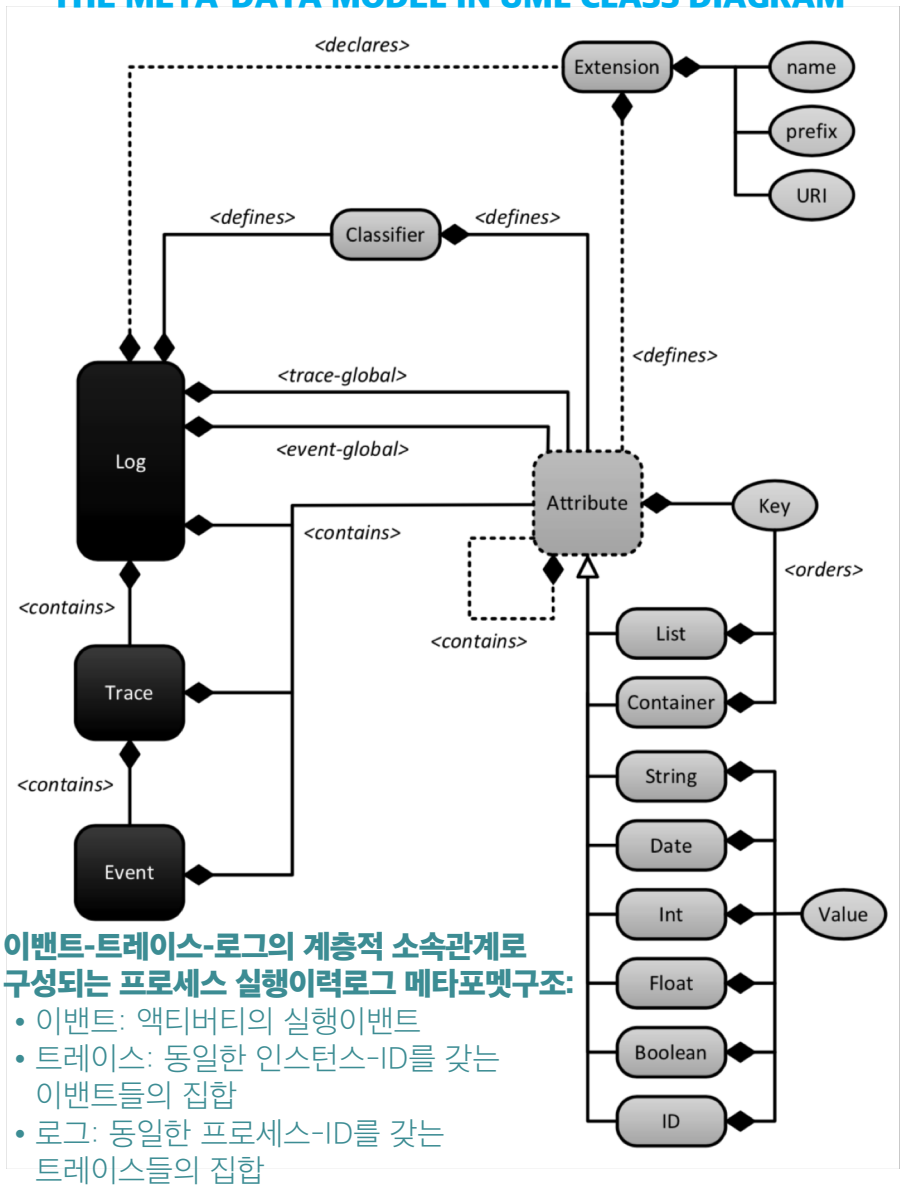




**Discovery of Process Enactment Proportions from Business Process Logs**

**Proportional
Business Process Patterns**

# THE IEEE XES EVENT LOG FORMAT

## THE IEEE XES EVENT STREAM FORMAT: THE META-DATA MODEL IN UML CLASS DIAGRAM




**이밴트-트레이스-로그의 계층적 소속관계로 구성되는 프로세스 실행이력로그 메타포멧구조:**

- 이밴트: 액티버티의 실행이밴트
- 트레이스: 동일한 인스턴스-ID를 갖는 이밴트들의 집합
- 로그: 동일한 프로세스-ID를 갖는 트레이스들의 집합

## A SAMPLE OF XES-FORMATTED EVET TRACES IN THE 2013 BPI CHALLENGE DATASET

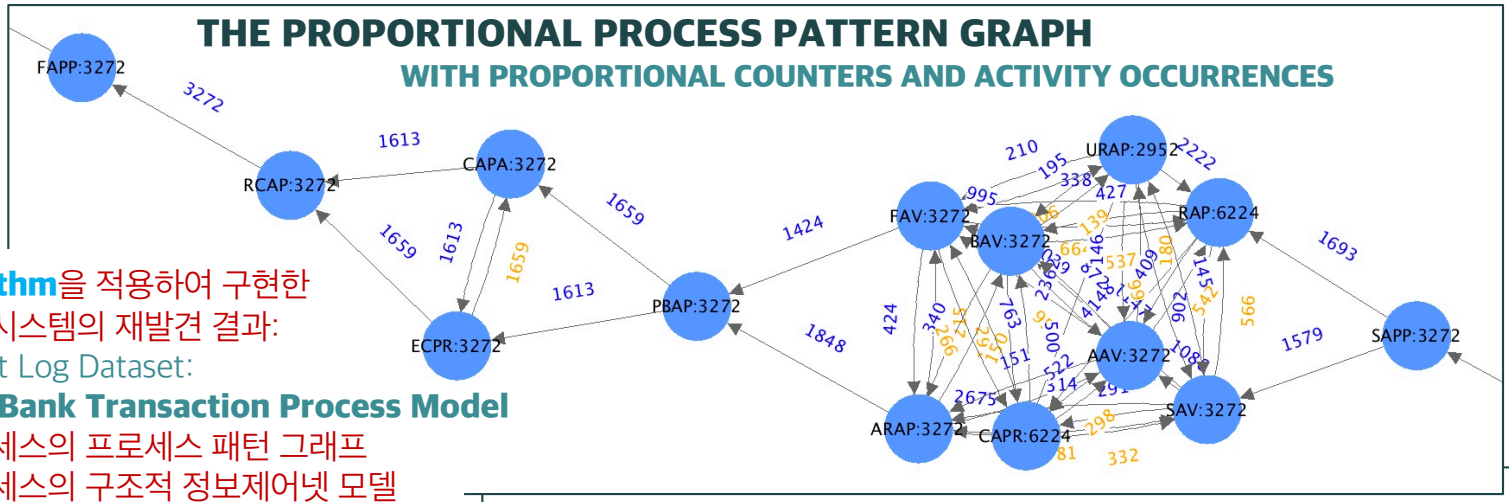
BPI_Challenge_2013_open_problems.xes

```
            <string key="lifecycle:transition" value="Assigned"/>
        </event>
</trace>
<trace>
        <string key="concept:name" value="1-535418135"/>
        <event>
                <string key="org:group" value="Org line A2"/>
                <string key="resource country" value="POLAND"/>
                <string key="org:resource" value="Michal"/>
                <string key="oranization country" value="pl"/>
                <string key="org:role" value="A2_1"/>
                <string key="concept:name" value="Accepted"/>
                <string key="impact" value="Medium"/>
                <string key="product" value="PROD791"/>
                <date key="time:timestamp" value="2011-06-21T12:01:27+02:00"/>
                <string key="lifecycle:transition" value="In Progress"/>
        </event>
        <event>
                <string key="org:group" value="Org line A2"/>
                <string key="resource country" value="POLAND"/>
                <string key="org:resource" value="Michal"/>
                <string key="oranization country" value="pl"/>
```

**국제표준 IEEE XES 포맷 기반의 프로세스 실행이력로그 샘플:**

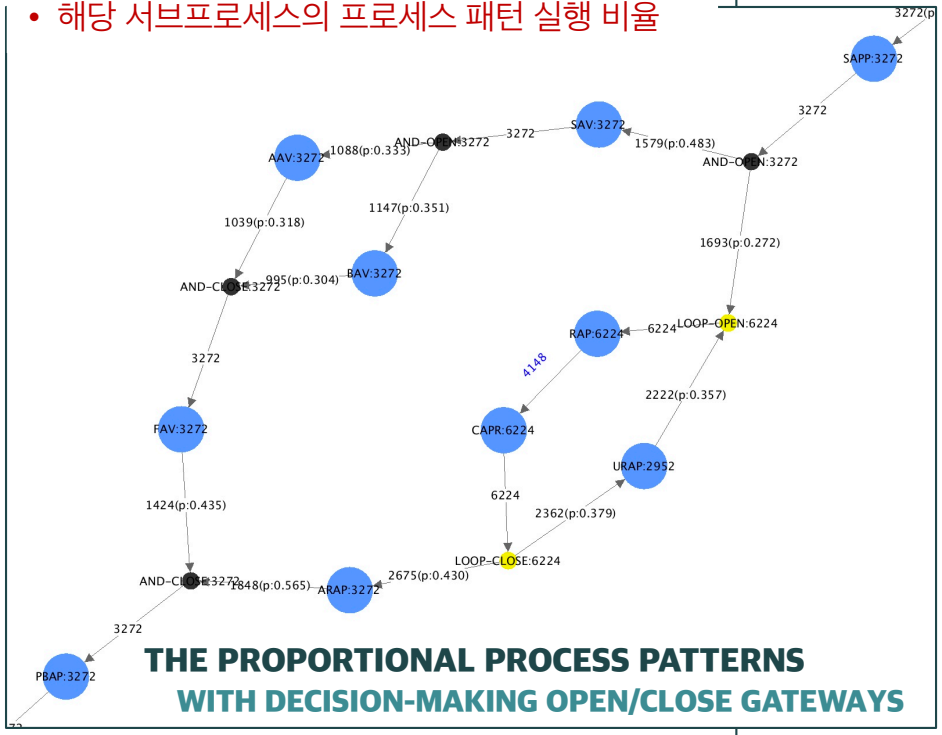
- 트레이스: `<trace> .... </trance>`
- 이밴트: `<event> .... < /event>`

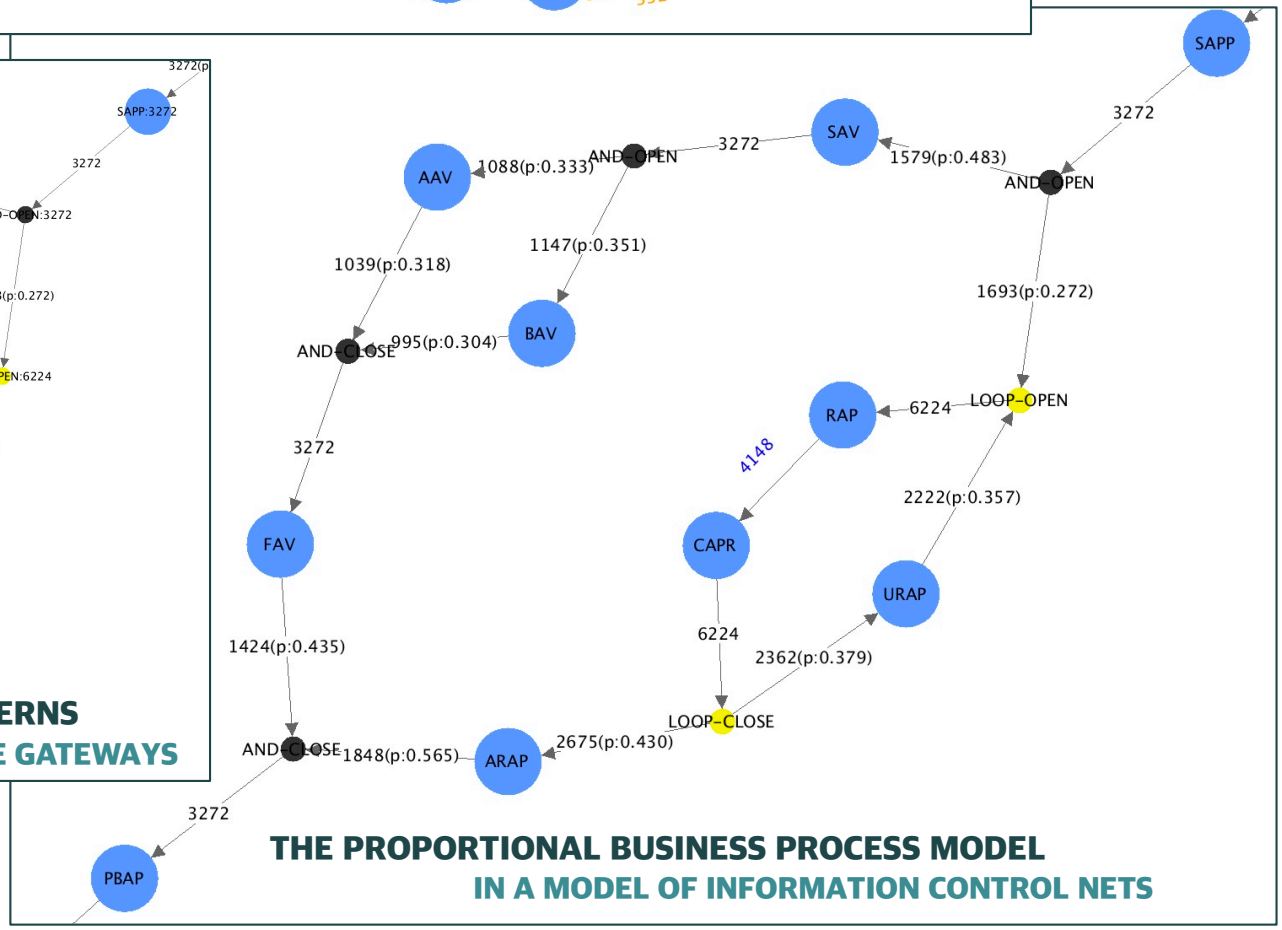
```
                <string key="org:group" value="Org line A2"/>
                <string key="resource country" value="POLAND"/>
                <string key="org:resource" value="Michal"/>
                <string key="oranization country" value="pl"/>
                <string key="org:role" value="A2_1"/>
                <string key="concept:name" value="Accepted"/>
                <string key="impact" value="Medium"/>
                <string key="product" value="PROD791"/>
                <date key="time:timestamp" value="2011-07-28T14:24:26+02:00"/>
                <string key="lifecycle:transition" value="Wait"/>
        </event>
        <event>
                <string key="org:group" value="Org line A2"/>
                <string key="resource country" value="POLAND"/>
                <string key="org:resource" value="Michal"/>
                <string key="oranization country" value="pl"/>
                <string key="org:role" value="A2_1"/>
                <string key="concept:name" value="Accepted"/>
                <string key="impact" value="Medium"/>
                <string key="product" value="PROD791"/>
                <date key="time:timestamp" value="2012-01-19T15:23:31+01:00"/>
                <string key="lifecycle:transition" value="In Progress"/>
        </event>
        <event>
                <string key="org:group" value="Org line A2"/>
                <string key="resource country" value="POLAND"/>
                <string key="org:resource" value="Michal"/>
                <string key="oranization country" value="pl"/>
                <string key="org:role" value="A2_1"/>
                <string key="concept:name" value="Completed"/>
                <string key="impact" value="Medium"/>
                <string key="product" value="PROD791"/>
                <date key="time:timestamp" value="2012-01-19T15:25:35+01:00"/>
                <string key="lifecycle:transition" value="Closed"/>
        </event>
</trace>
<trace>
        <string key="concept:name" value="1-535439201"/>
        <event>
                <string key="org:group" value="Org line C"/>
```

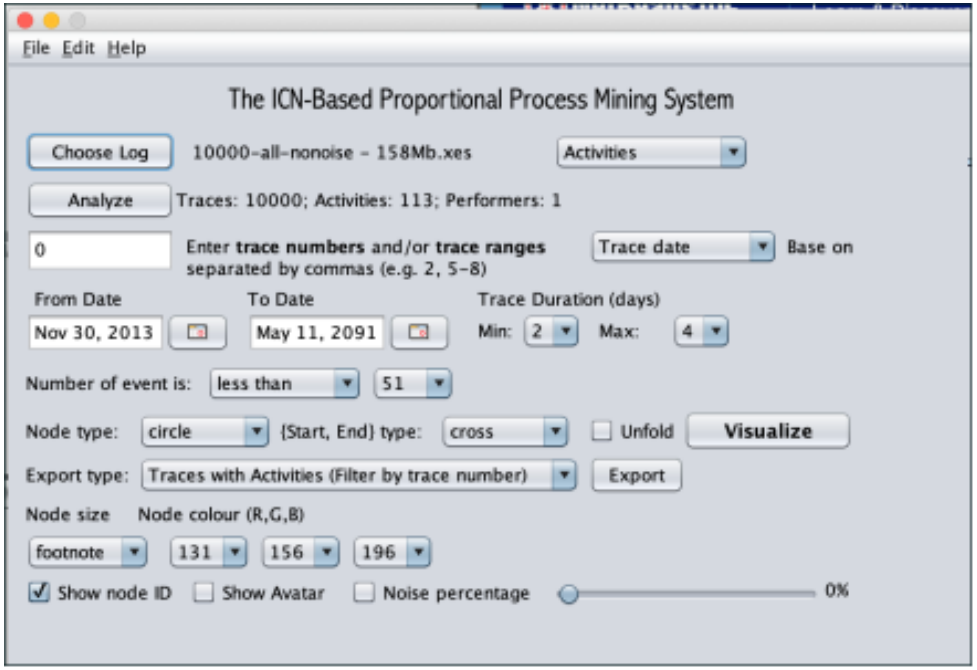



**The rho-Algorithm**을 적용하여 구현한 프로세스 마이닝 시스템의 재발견 결과:

- Process Event Log Dataset:
  **The Large Bank Transaction Process Model**
- 해당 서브프로세스의 프로세스 패턴 그래프
- 해당 서브프로세스의 구조적 정보제어넷 모델
- 해당 서브프로세스의 프로세스 패턴 실행 비율



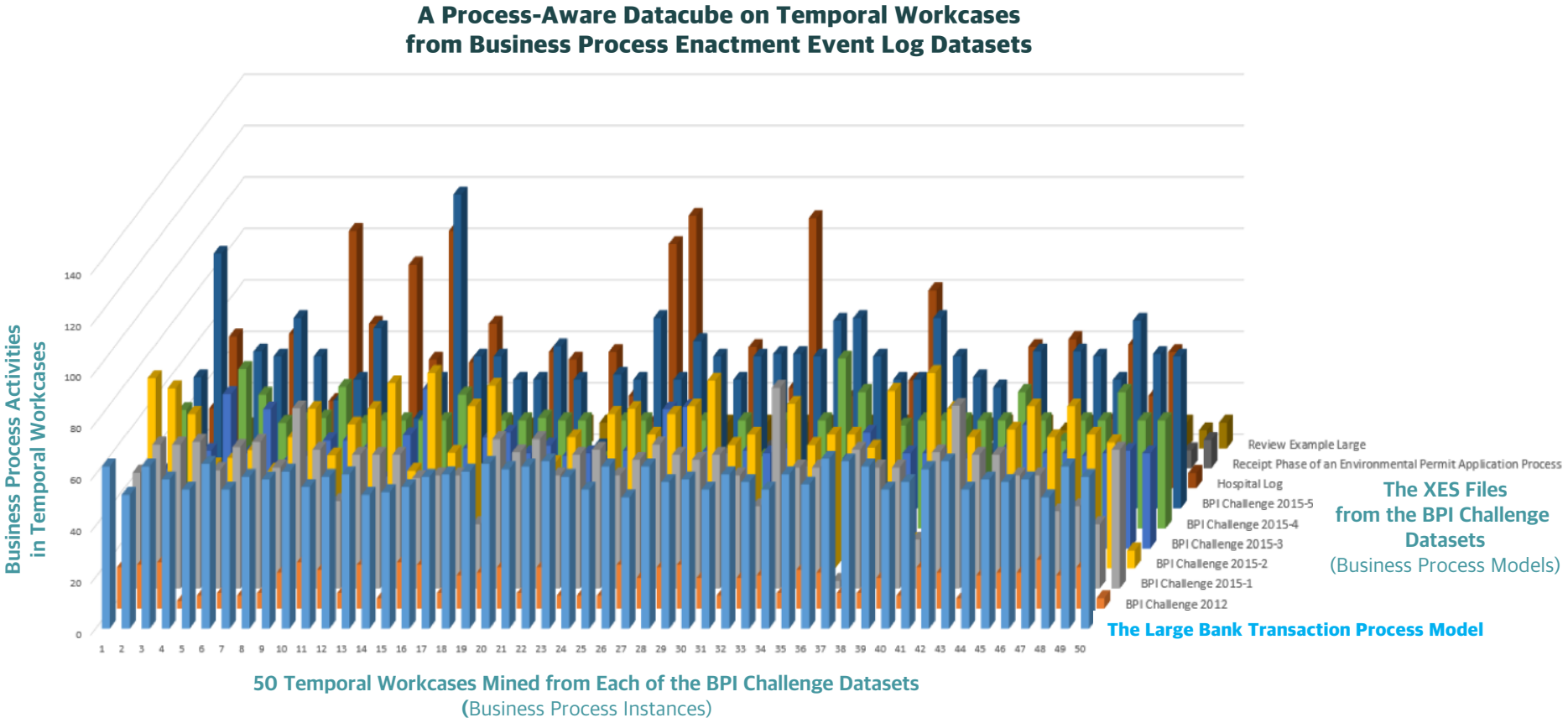






# AN EXPERIMENT FOR PROVING THE CORRECTNESS OF THE ρ-ALGORITHM

# A DATACUBE OF BUSINESS PROCESS LOGS:

## The BPI challenges of 4TU.Centre for Research Data




**THE BUSINESS PROCESS INTELLIGENCE CHALLENGES SINCE 2011 (BPM International Conferences)**

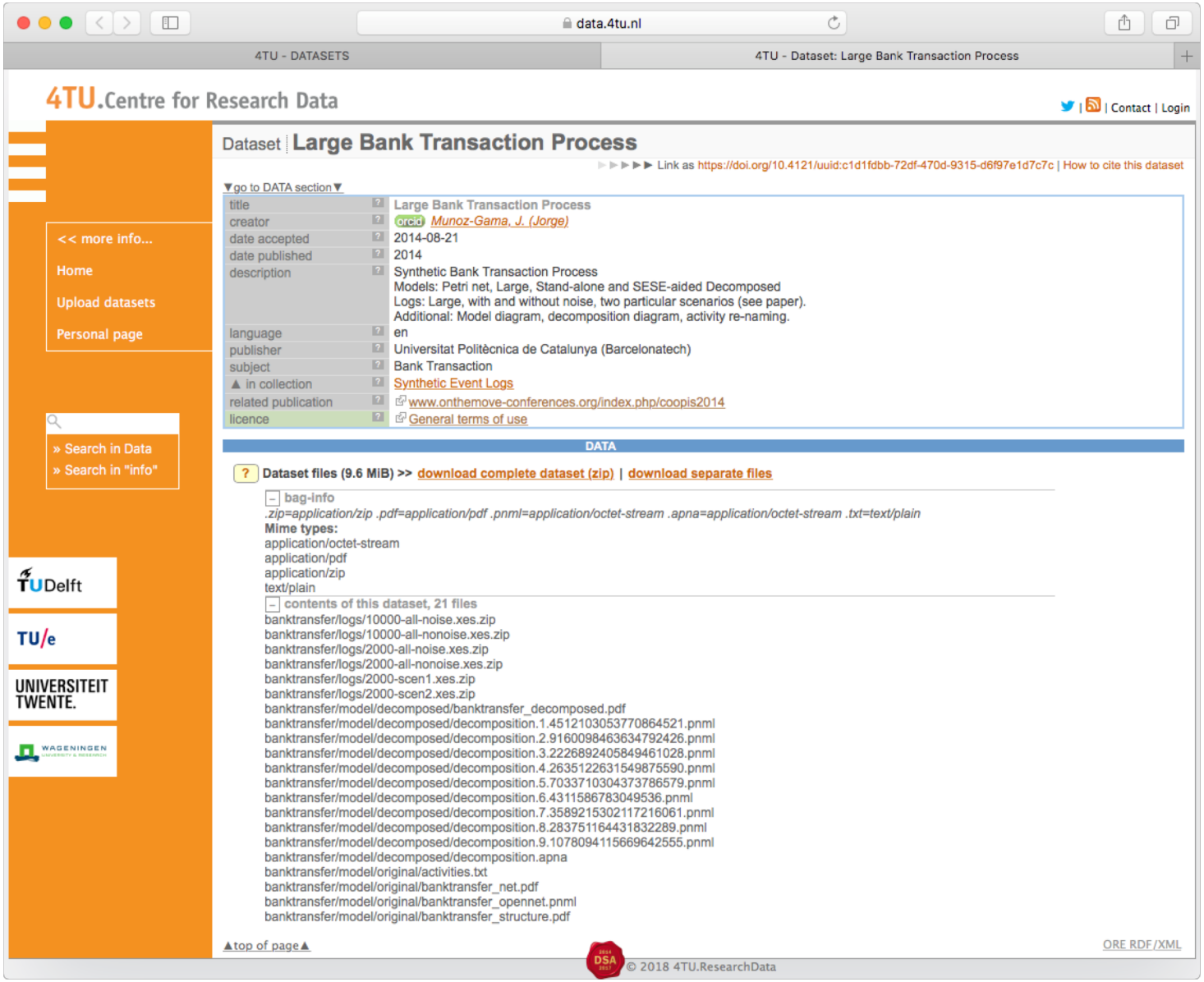
- The challenges are sponsored by Minit and Celonis, that provide participants with a real-life event log.
- The challenges strongly encourage people to use any tools, techniques, methods at their disposal.
- The sponsors provide access to their tools for use with the BPI Challenge dataset.
- The challenges have three categories, namely **students**, **academics** and **professionals**, for awards decisions.

## The BPI Challenge Dataset for the Experiment

- The Large Bank Transaction Process Model: 10,000 PROCESS INSTANCES ON THE DATASET
- 4TU.Centre for Research Data: https://data.4tu.nl/repository/

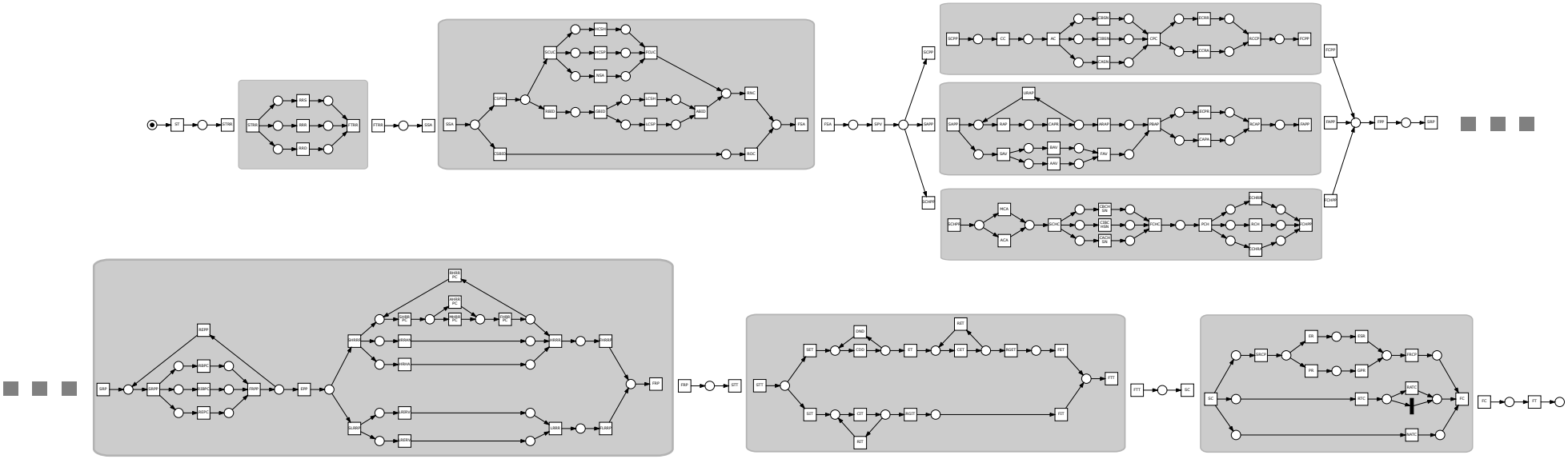
# THE DATASET FROM the 4TU.Centre for Research Data

## THE LARGE BANK TRANSACTION PROCESS MODEL



# AN EXPERIMENTAL DATASET:

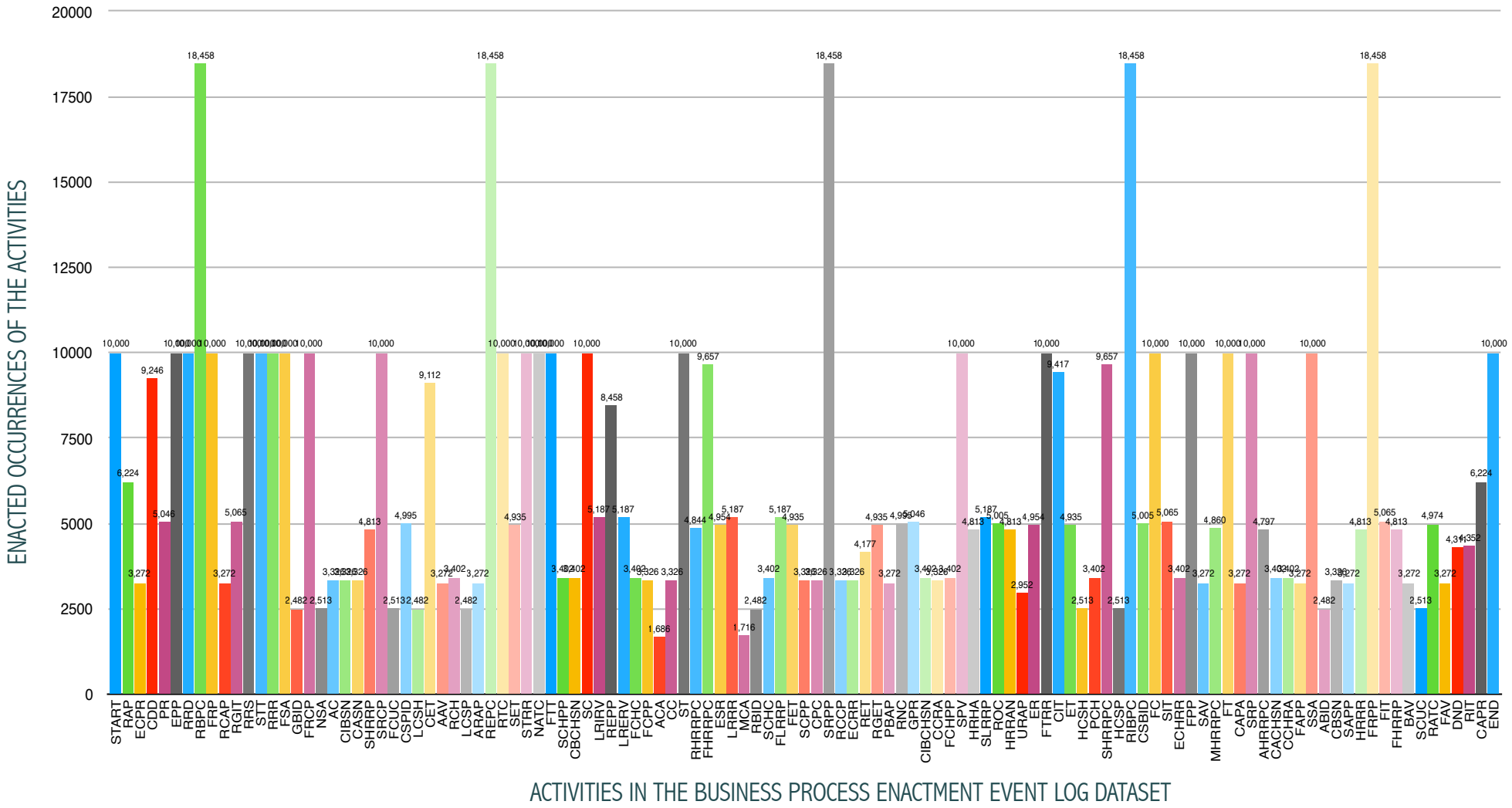
## THE LARGE BANK TRANSACTION PROCESS MODEL



- There are 113 activities in the chosen business process model, and these activities are subdivided into 8 subprocesses.
- After applying our rediscovery framework to the dataset, which is implemented as a name of the event trace mining algorithm, we found out that the chosen event history file contains all the enactment event logs stored from enacting the exact number of 10,000 business process instances.

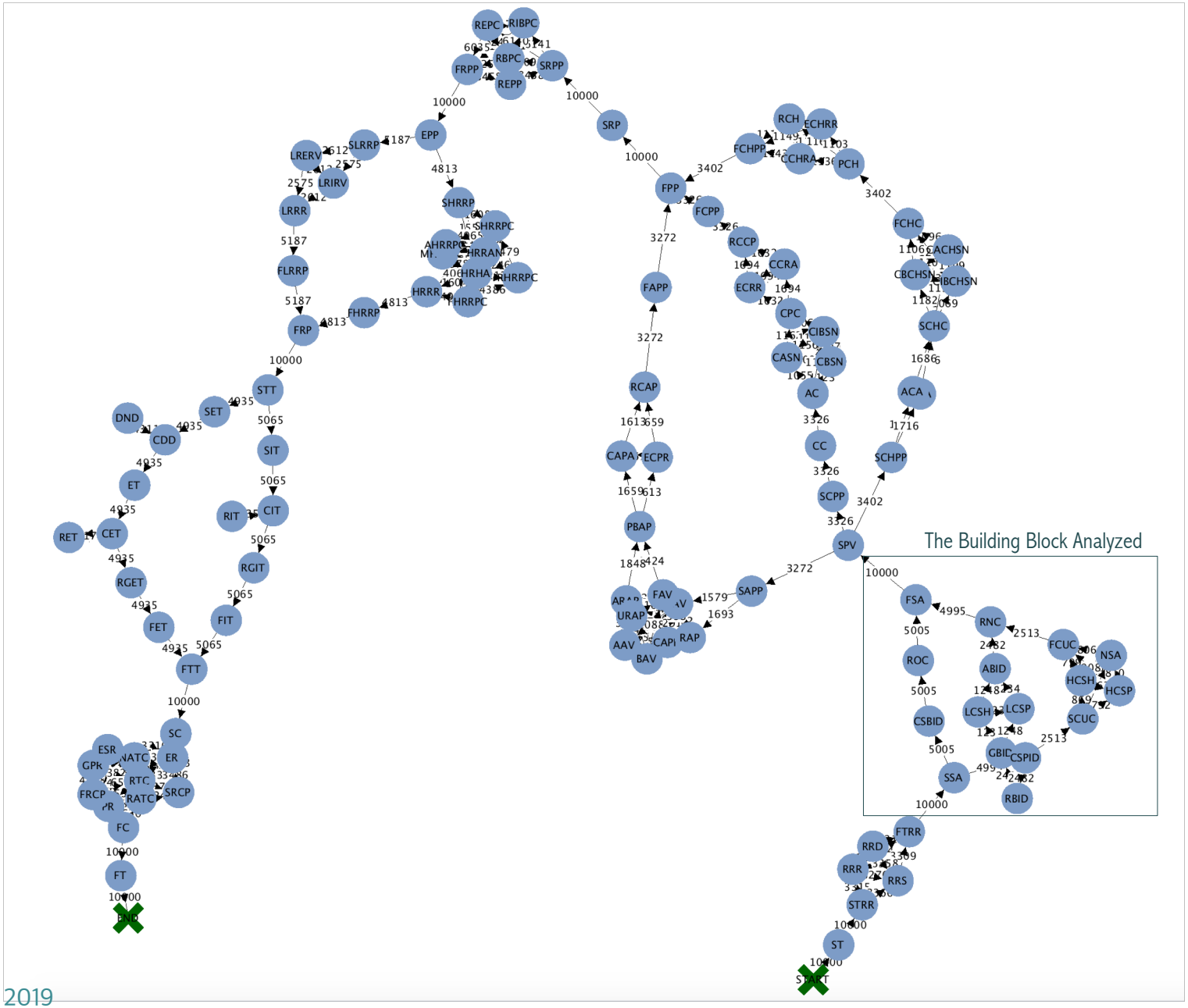
# EXPERIMENTAL MINING RESULT: OCCURRENCES

## THE ACTIVITIES AND THEIR OCCURRENCES FROM THE LBTPM DATASET

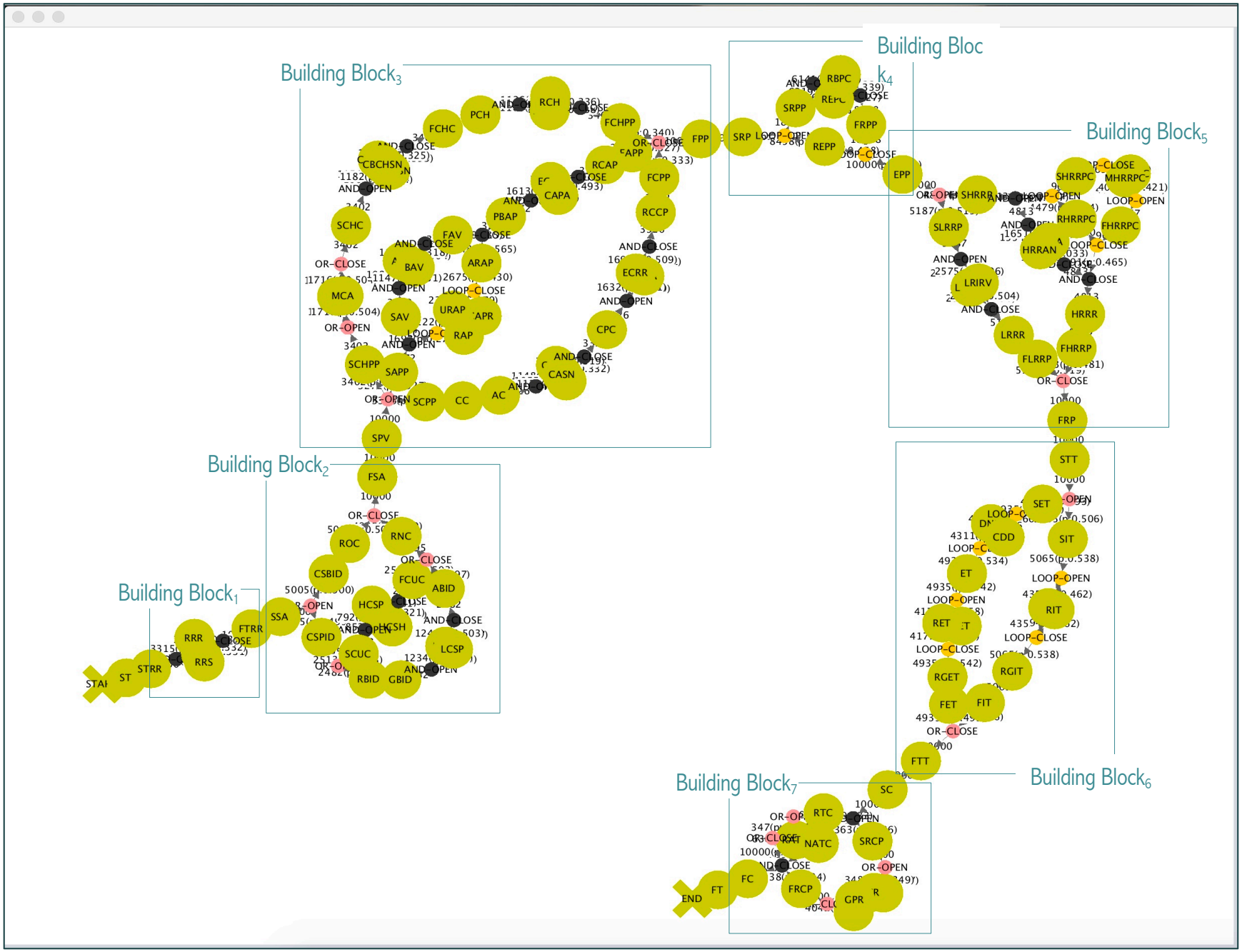


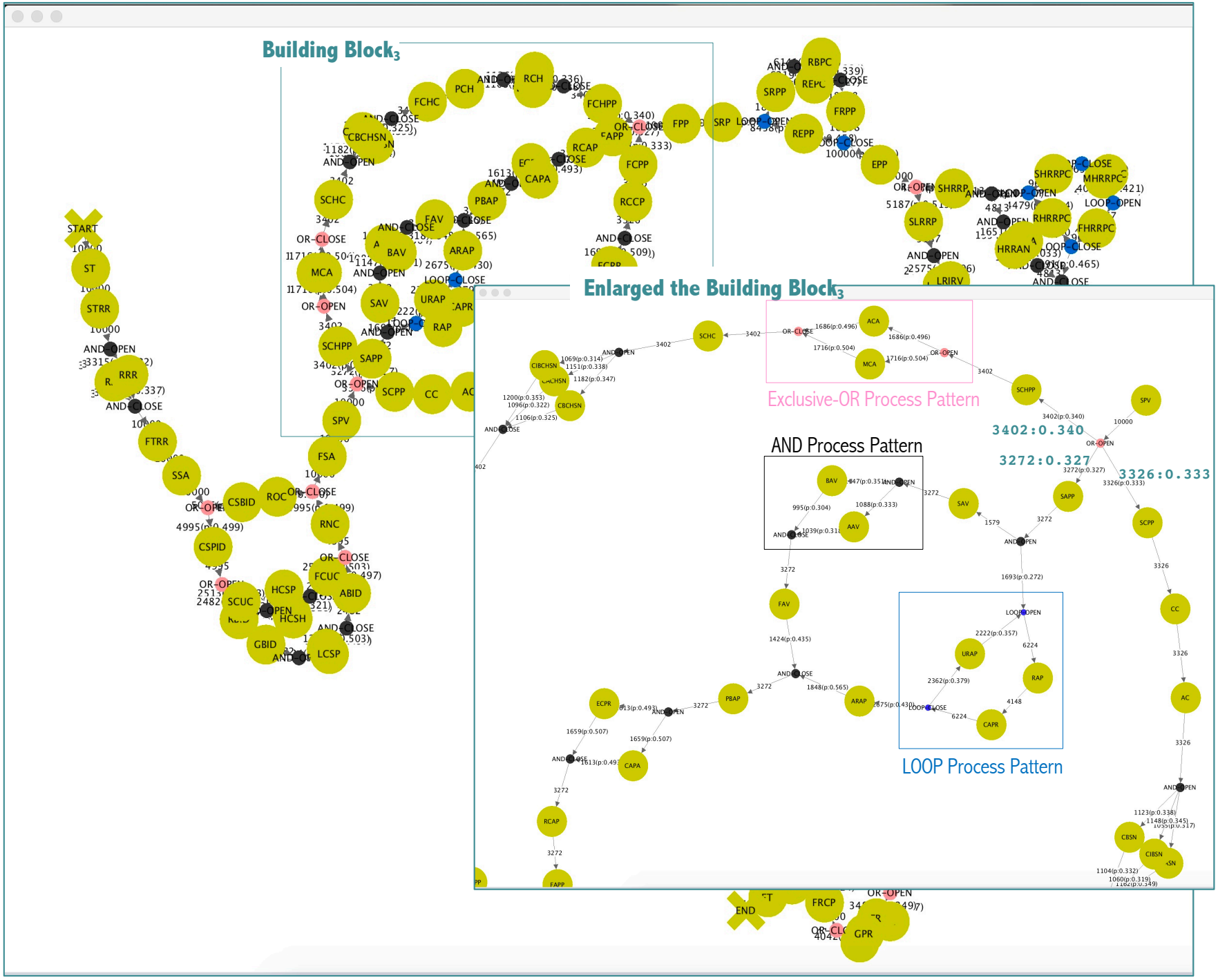
# EXPERIMENTAL MINING RESULT: PROPORTIONAL COUNTS

## A PROPORTIONAL PROCESS PATTERN GRAPH FROM THE LBTPM DATASET



# A PROPORITONAL INFORMATION CONTROL NET MODEL FROM THE LBTPM DATASET




Building Block3
Enlarged the Building Block3
Exclusive-OR Process Pattern
AND Process Pattern
LOOP Process Pattern
3402:0.340
3272:0.327
3326:0.333

# WORKFLOW AND BUSINESS PROCESS MINING SYSTEM DEMO:

# THE ICN-BASED PROCESS MINING SYSTEM

**PHAM DINH LAM, Ph.D. STUDENT**

Collaboration Technology Research Lab.
Department of Computer Science and Engineering
KYONGGI UNIVERSITY, REPUBLIC OF KOREA

# CONCLUSIONS

## THE MAIN CONTRIBUTION OF THE TUTORIAL:

- INITIATED THE CONCEPT OF PROPORTIONAL BUSINESS PROCESS MODELS
  - DEFINED A FORMAL MODEL OF PROPORTIONAL INFORMATION CONTROL NETS
- PROPOSED THE FRAMEWORKS FOR REDISCOVERING p-ICNs
  - A CONCEPTUAL FRAMEWORK for discovering proportional process patterns
  - AN ALGORITHMIC FRAMEWORK for discovering proportional process patterns
- DEVISED THE ALGORITHMS (ρ-Algorithm) IN THE FRAMEWORK
  - Mining all the four types of proportional process patterns, such as linear, disjunctive, conjunctive and iterative process patterns, from business process enactment event logs formatted in the IEEE XES.
- CARRIED OUT THE EXPERIMENTAL STUDY WITH AN ANALYSIS ON A REAL DATASET
  - Carrying successfully out an experiment on a business process enactment event history dataset (Non-noise) recorded from the execution of the Large Bank Transaction Process Model.
- 제약사항
  - 완전한 프로세스 모델의 재발견을 위해서는 모든 조합의 실행로그가 필요함.

Conclusively,

- IMPACTS UPON EVERY PHASE OF THE BUSINESS PROCESS LIFE-CYCLE
- CONTIRUBUTION TO THE BUSINESS PROCESS REDISCOVERY RESEARCH ARENA
- FUTURE RESEARCH ON
  - REDISCOVERING p-ICNs with Noise Datasets
  - REDISCOVERING p-ICNs with Performers' Work-Done Proportions

QUESTIONS AND ANSWERS

# THANK YOU VERY MUCH!

KWANGHOON PIO KIM, Ph.D.

COLLABORATION TECHNOLOGY RESEARCH LAB.
DIVISION of Computer Science and Engineering, **KYONGGI UNIVERSITY**
**kwang@kgu.ac.kr**
**http://ctrl.kyonggi.ac.kr**