Mineração de Processos:
O elo que faltava na gestão de processos de negócio

Diogo R. Ferreira

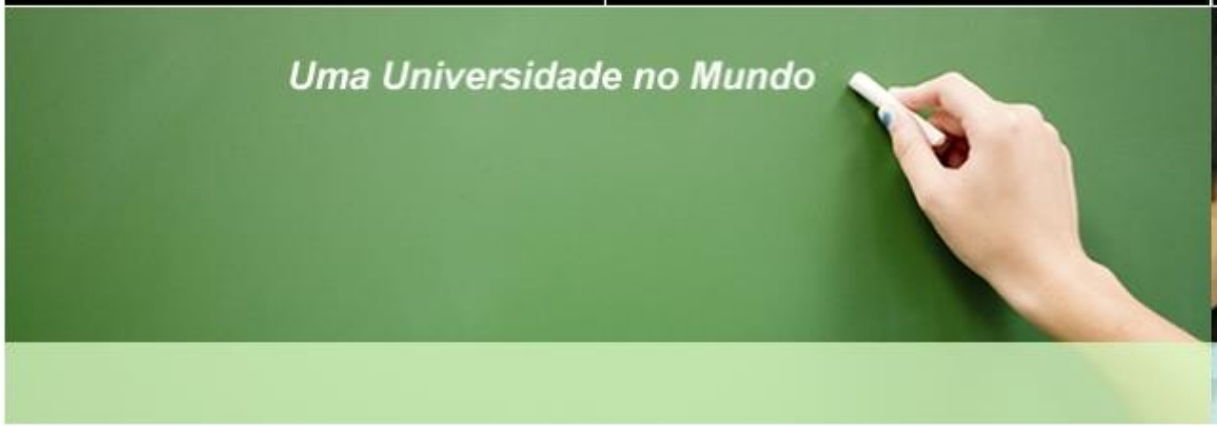
Instituto Superior Técnico, Universidade Técnica de Lisboa

VI Simpósio Brasileiro de Sistemas de Informação (SBSI 2010)
Marabá, 16 a 18 de Junho de 2010

UTL

UNIVERSIDADE TÉCNICA DE LISBOA

► Notícias ► Eventos ► Quicklinks



Eventos

26 de Set. 2009 — **Festival Voo Livre'09 - Covilhã**

25 de Set. 2009 — **Seminário de Tecnologias de Informação na Promoção das Áreas Protegidas (TIPAP-2009)**

7 a 8 de Set. 2009 — **CEI2009 - Encontro Nacional do Colégio de Engenharia Informática da Ordem dos Engenheiros**

7 a 7 de Set. 2009 — **Sessão de Homenagem ao Professor João Martins**

Notícias

IST Press edita o Livro "Ecologia Industrial: princípios e ferramentas"
18 de Set. 2009

IST Press edita o Livro "Análise Complexa e Equações Diferenciais"
18 de Set. 2009

Encontro "O contributo de Bolonha na prossecução da Qualidade no Ensino Superior", dia 21 de Outubro de 2009
17 de Set. 2009

3ª Edição do Livro "Vitrúvio-Tratado de Arquitectura"
16 de Set. 2009

Google™ Custom Search

Localização

Campus Alameda
Campus Taguspark

Iniciativas + Recursos

Observatório de Empregabilidade
Recrutamento
Biblioteca
Transporte "Alameda-Taguspark"
Webmail

Subscrever as feeds de RSS
Siga o Técnico no Twitter

História do IST

- 1911: fundado com os cursos de Engenharia de Minas, Civil, Mecânica, Electrotécnica e Química
- 1930s: construção do campus da Alameda

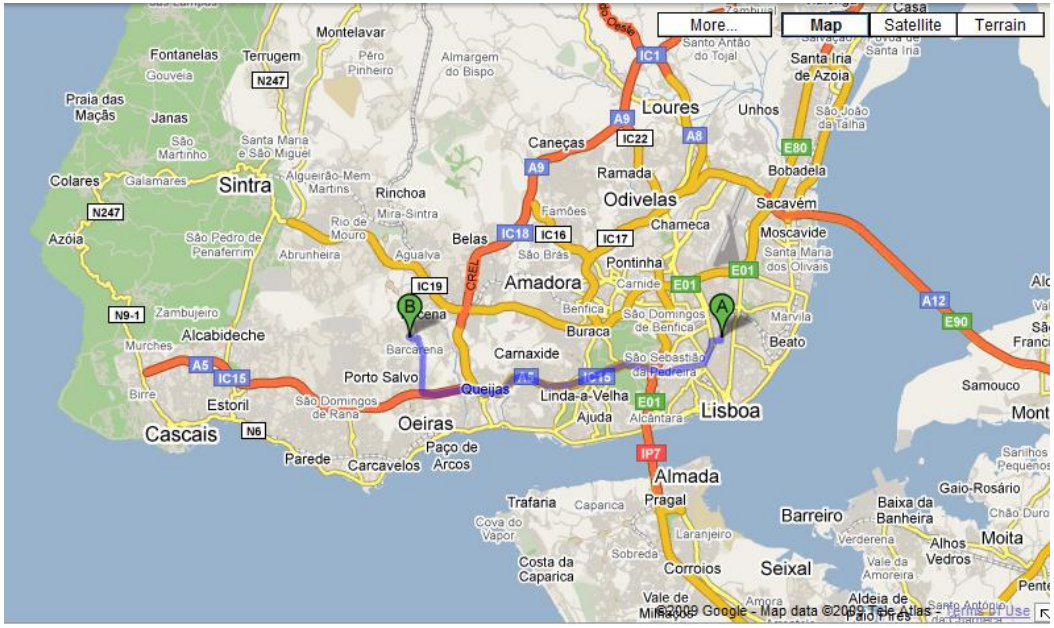
"Engenho e Obra", Dom Quixote

Bing Maps

História do IST

- 1930s: integra a Universidade Técnica de Lisboa (UTL)
- 1950-1970: desenvolvimento dos centros de pesquisa
- 1970: cursos passam de 6 para 5 years (no presente: 3+2)
- 1990s: construção do campus do Taguspark (Oeiras)

Bing Maps



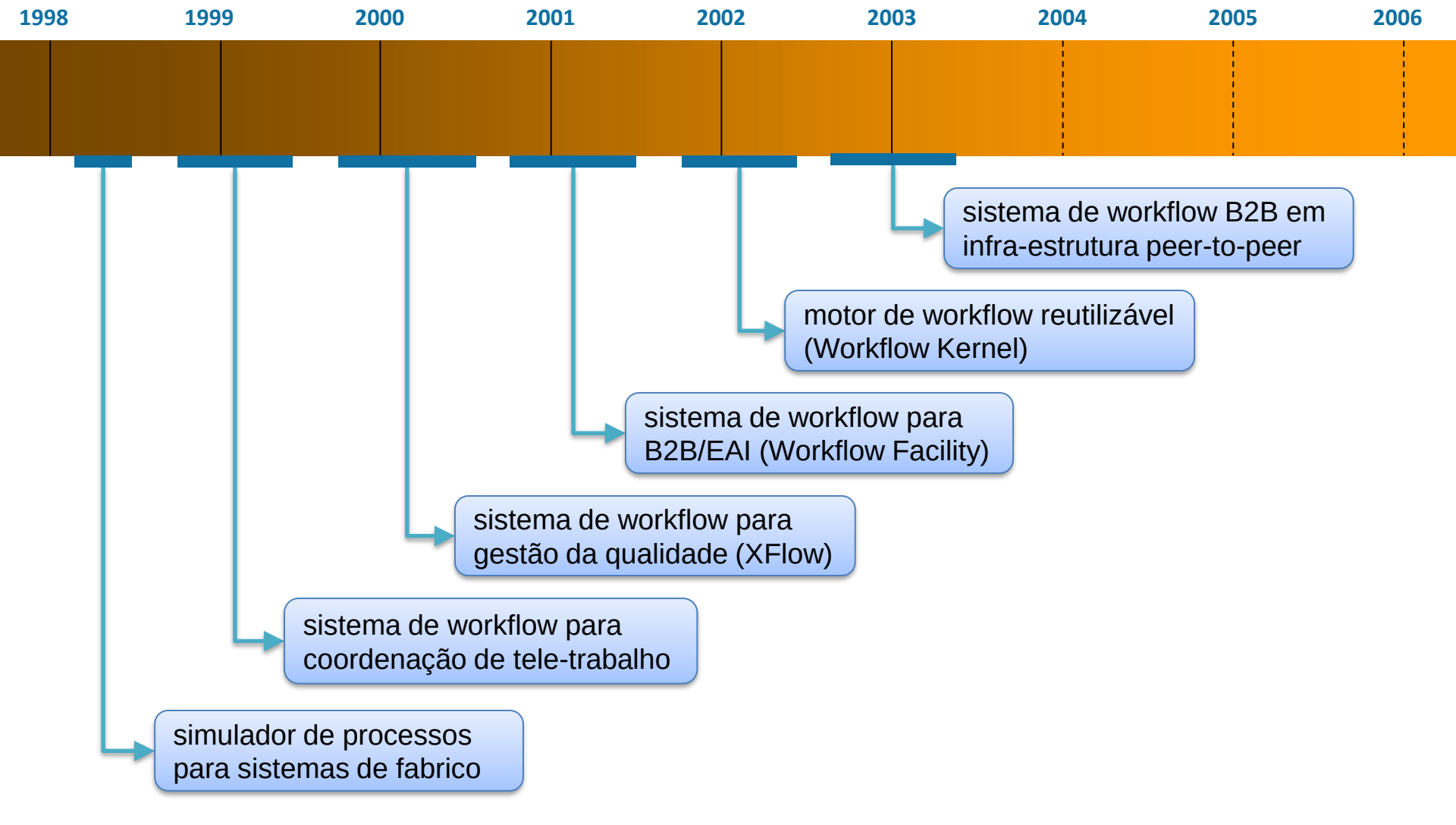
Google Maps

Pesquisa na UTL

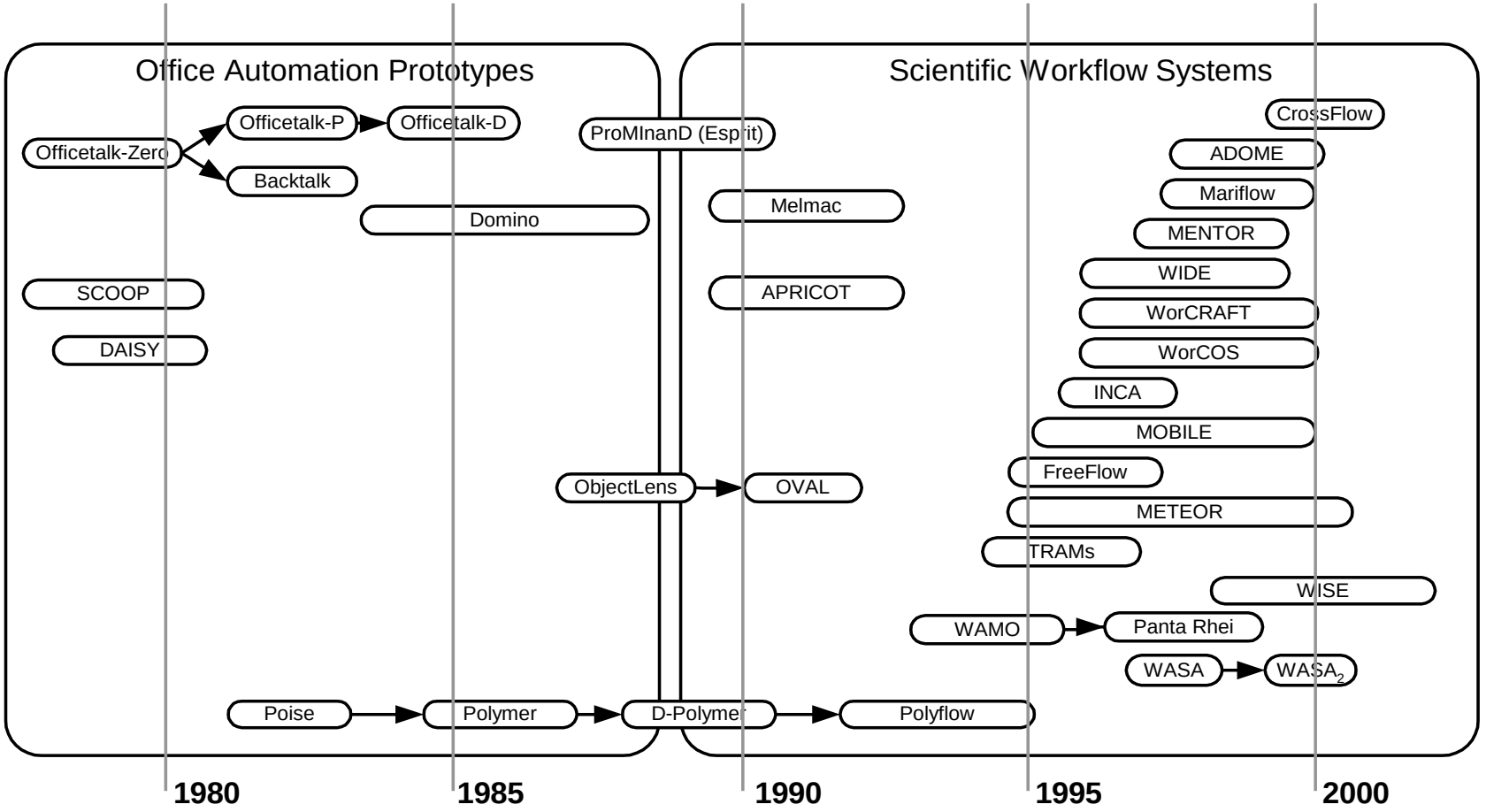
- Várias escolas, várias disciplinas
  - física, química, matemática
  - economia, gestão, ciências humanas
  - biologia, biotecnologia, bioinformática
  - ambiente e sustentabilidade
  - saúde e ciências do desporto
  - arquitectura, urbanismo, transportes
  - …

Manuel S. Pereira (Ed.)
622 pgs, Springer 2007

A minha pesquisa, no passado

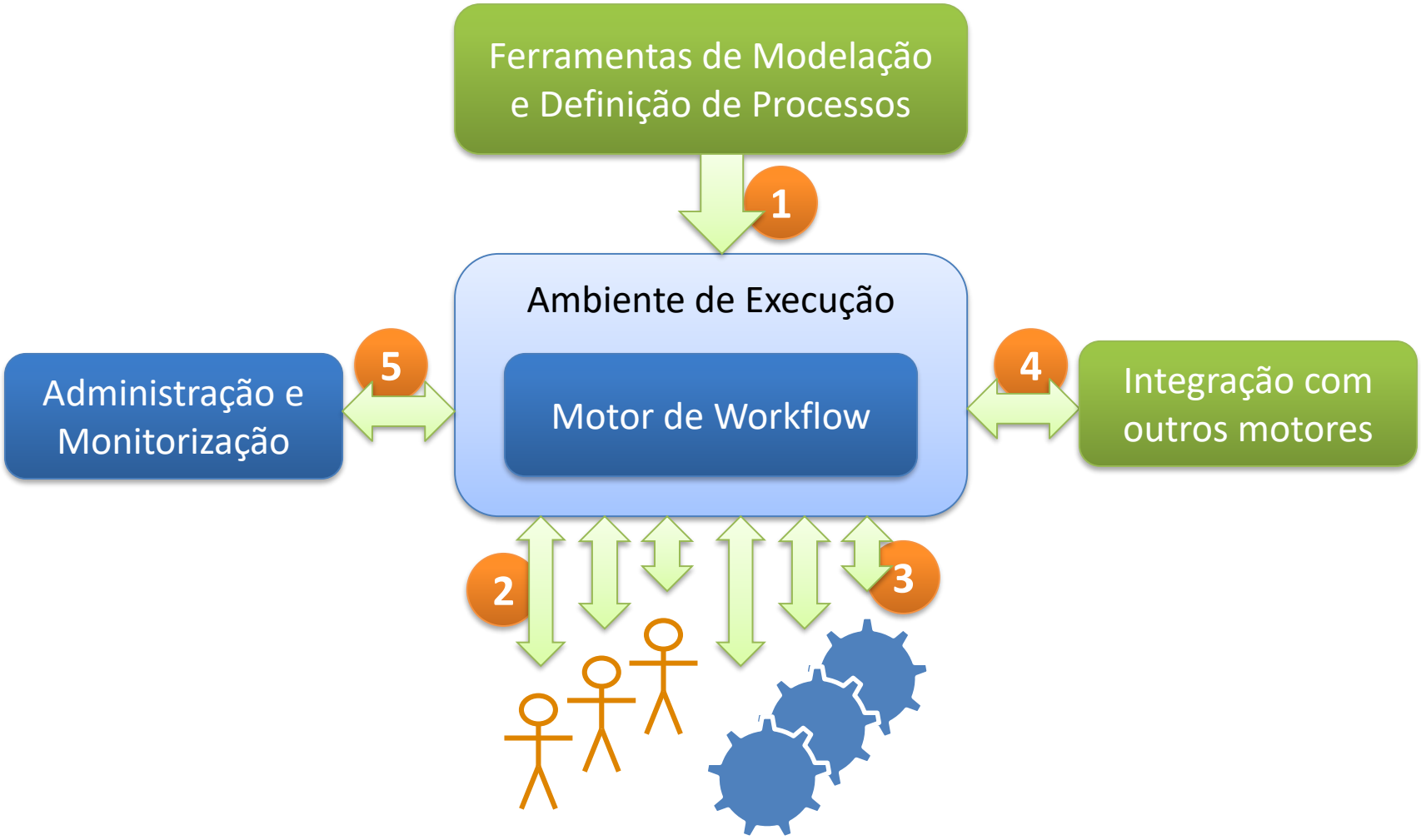


Pesquisa em sistemas de workflow

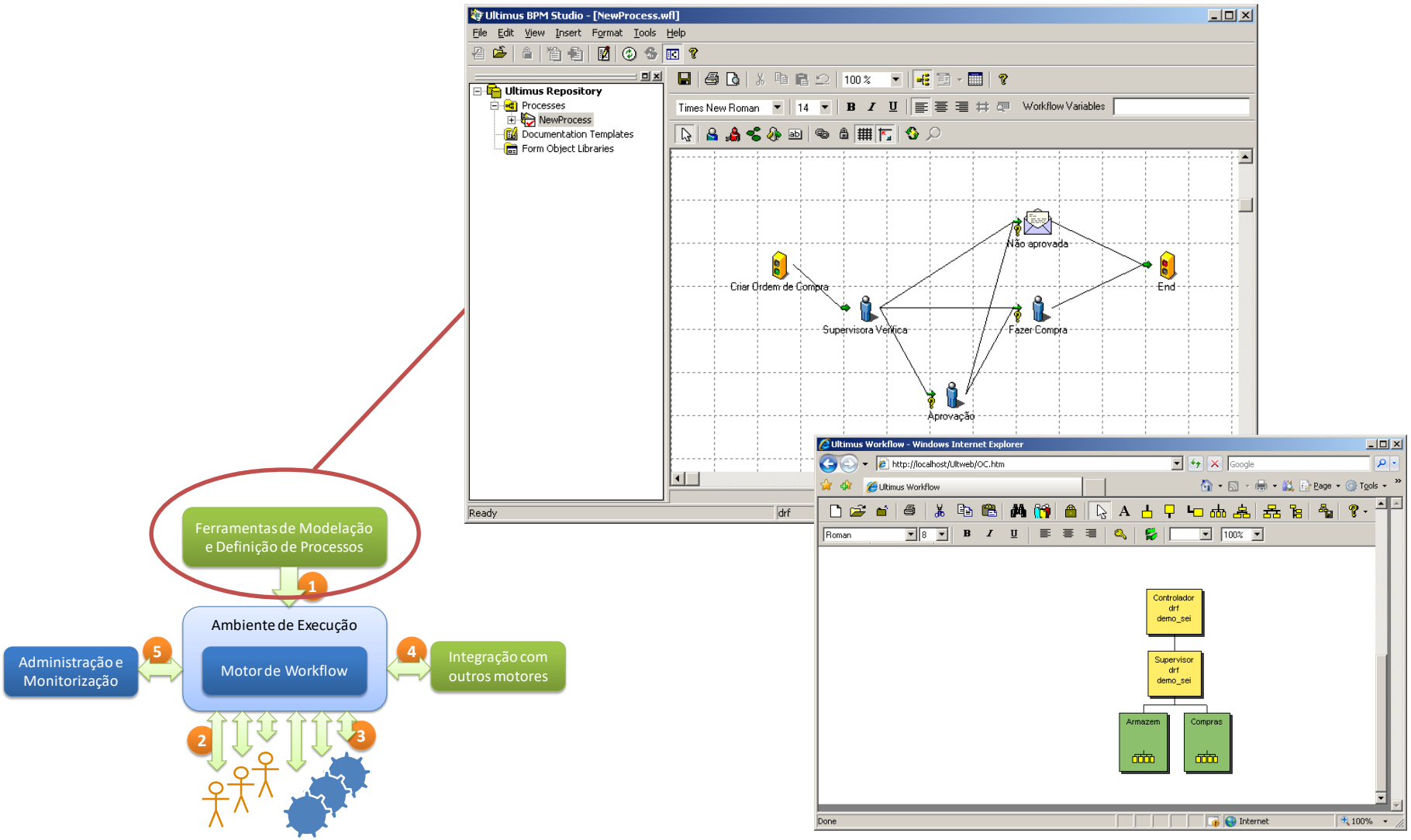


(Fonte: M. zur Muehlen, 2003)

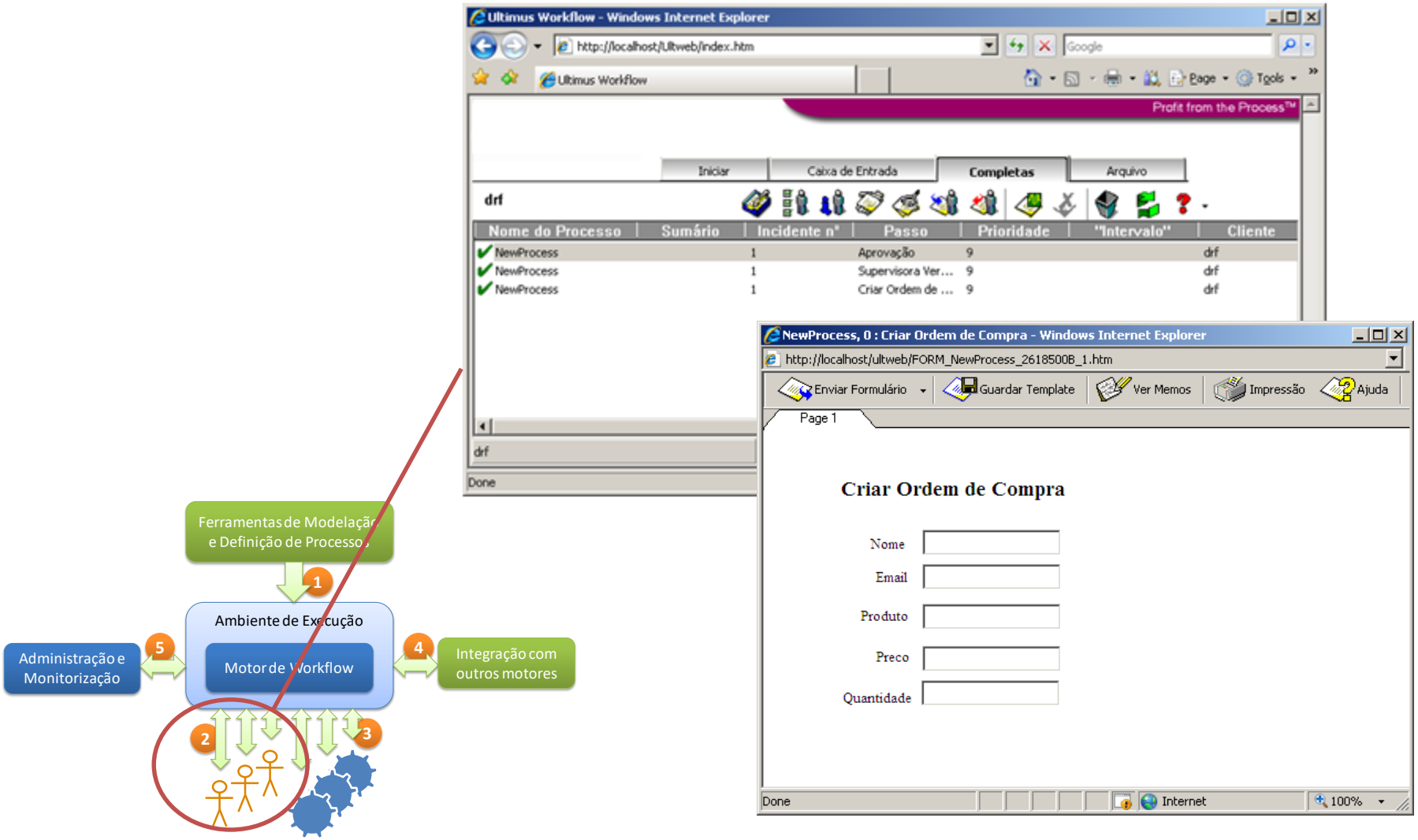
Sistemas de workflow



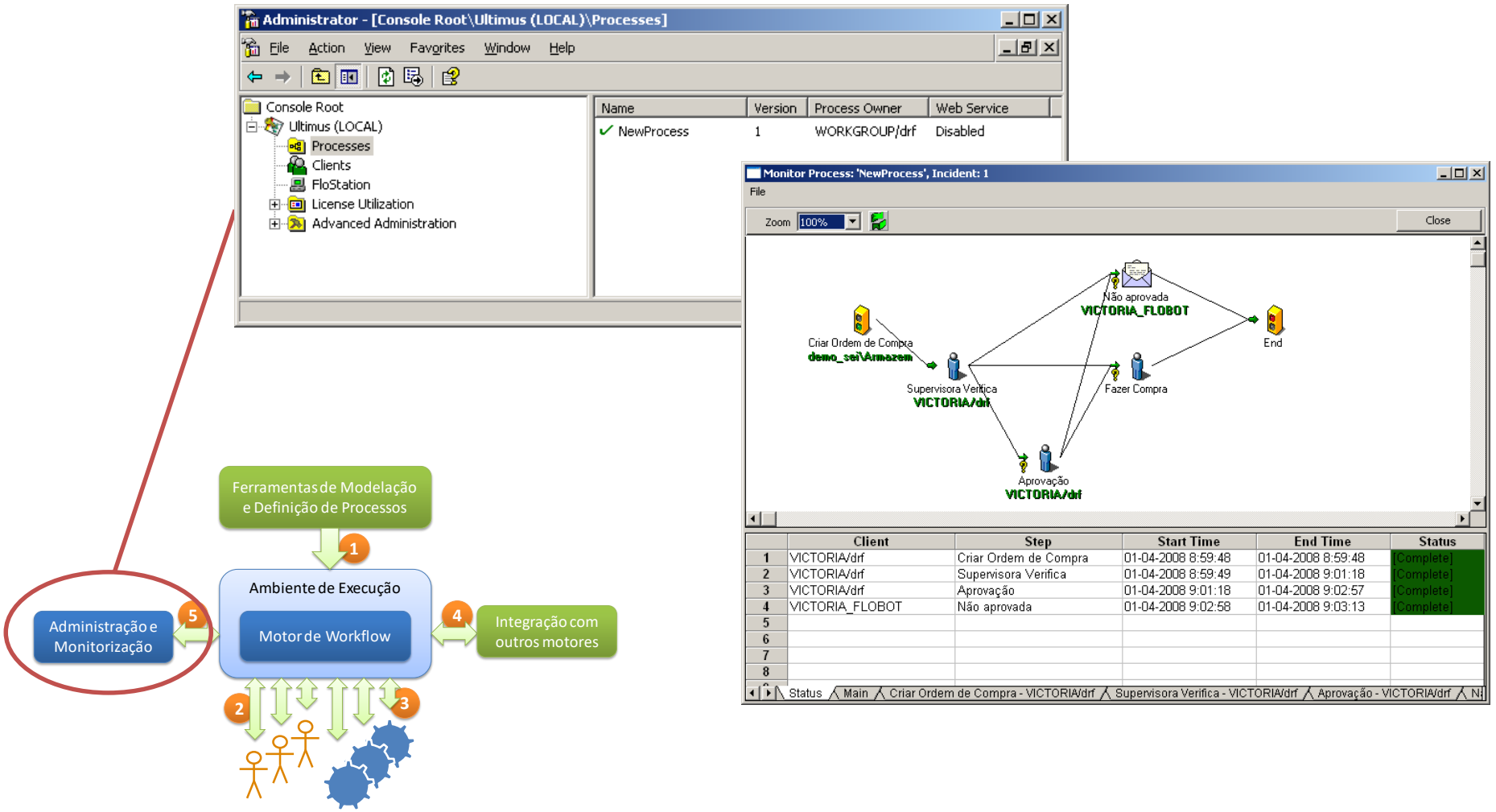
Sistemas de workflow



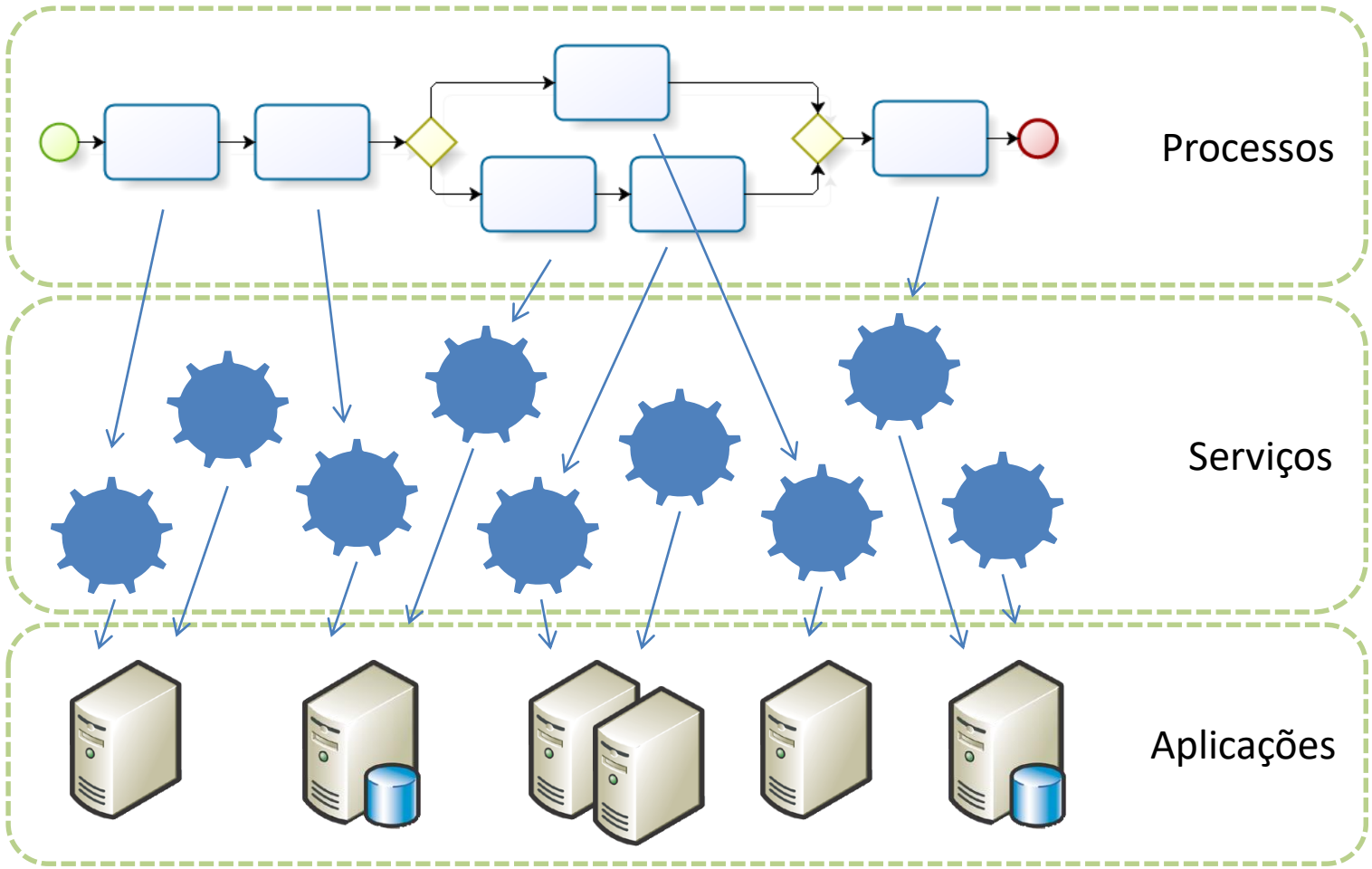
Sistemas de workflow



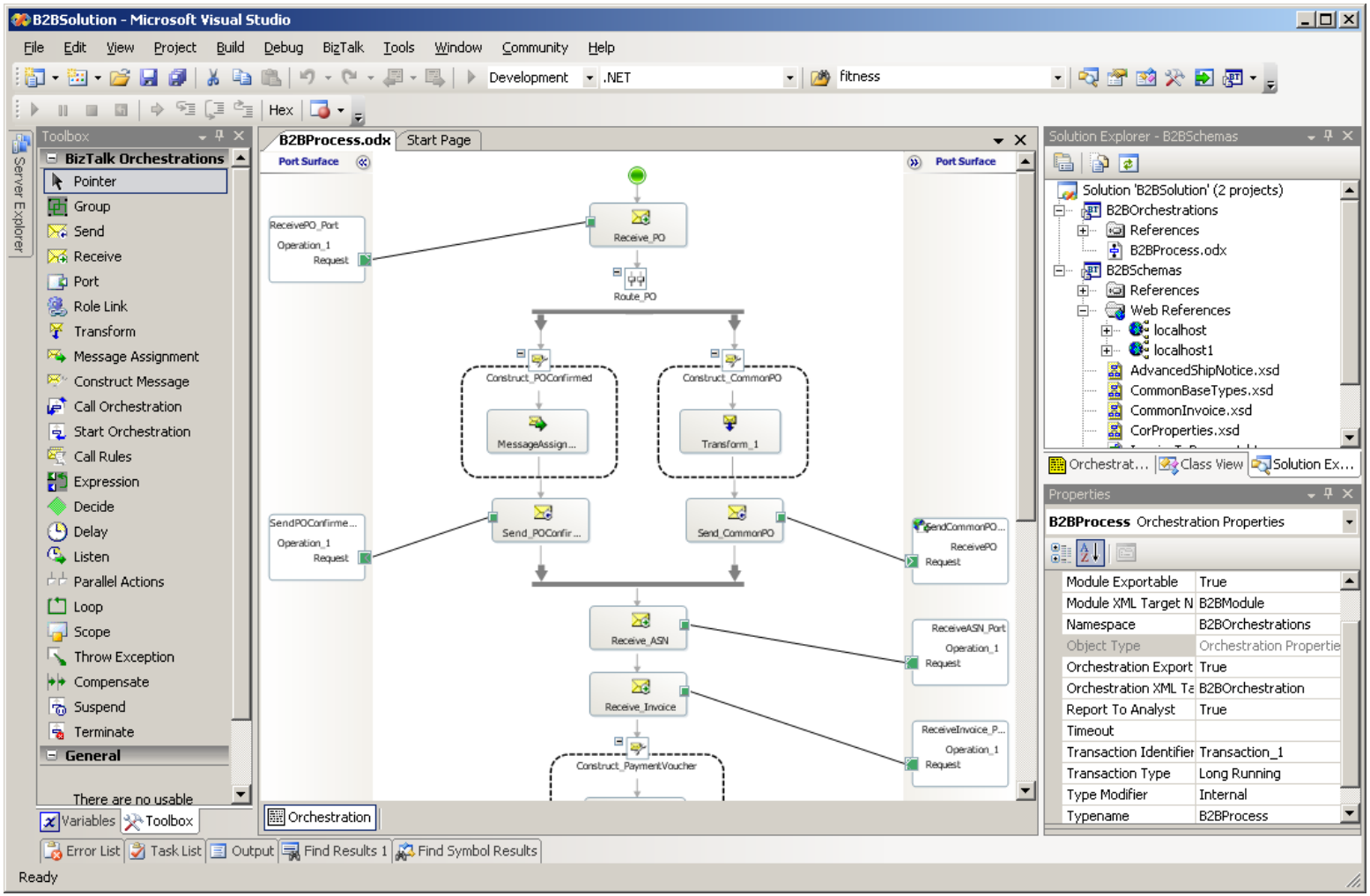
Sistemas de workflow



Arquitecturas SOA

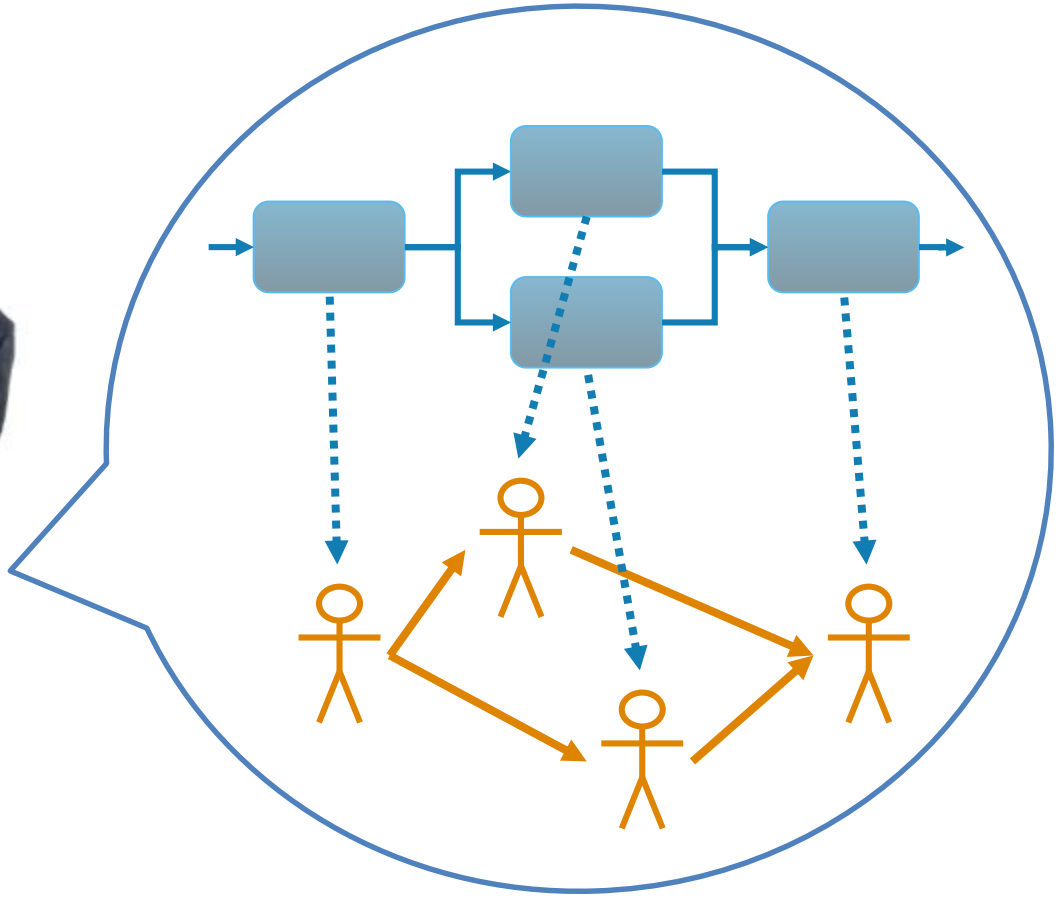


Arquitecturas SOA



Processos

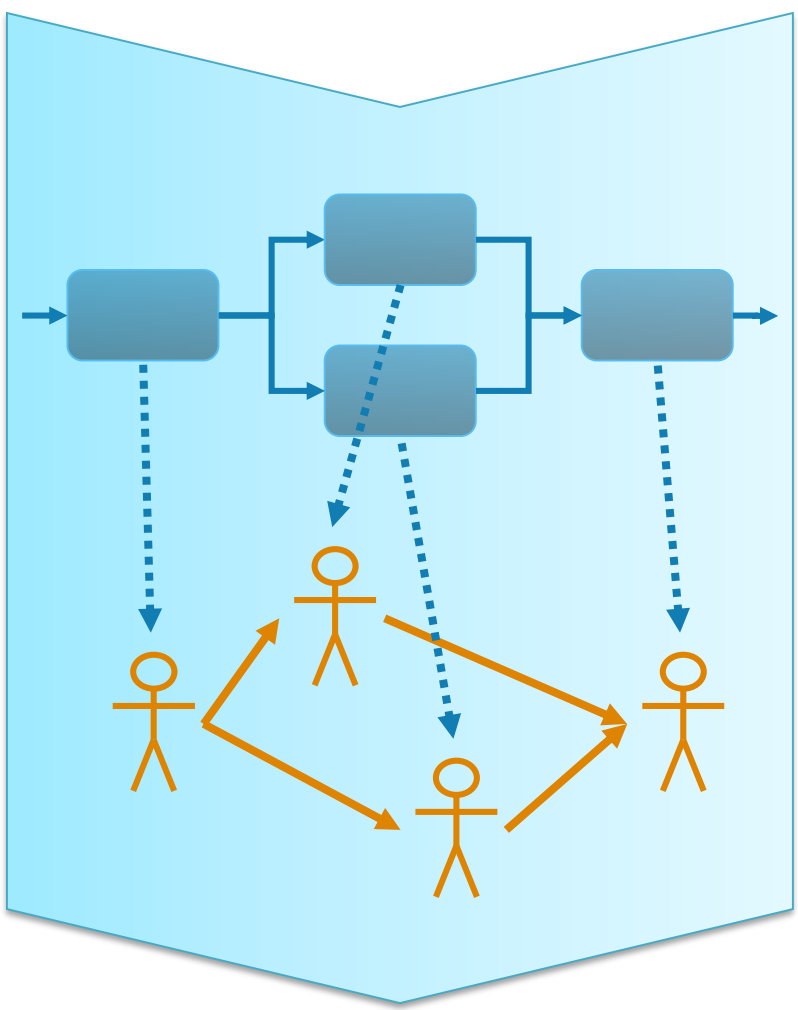
- um plano para a organização

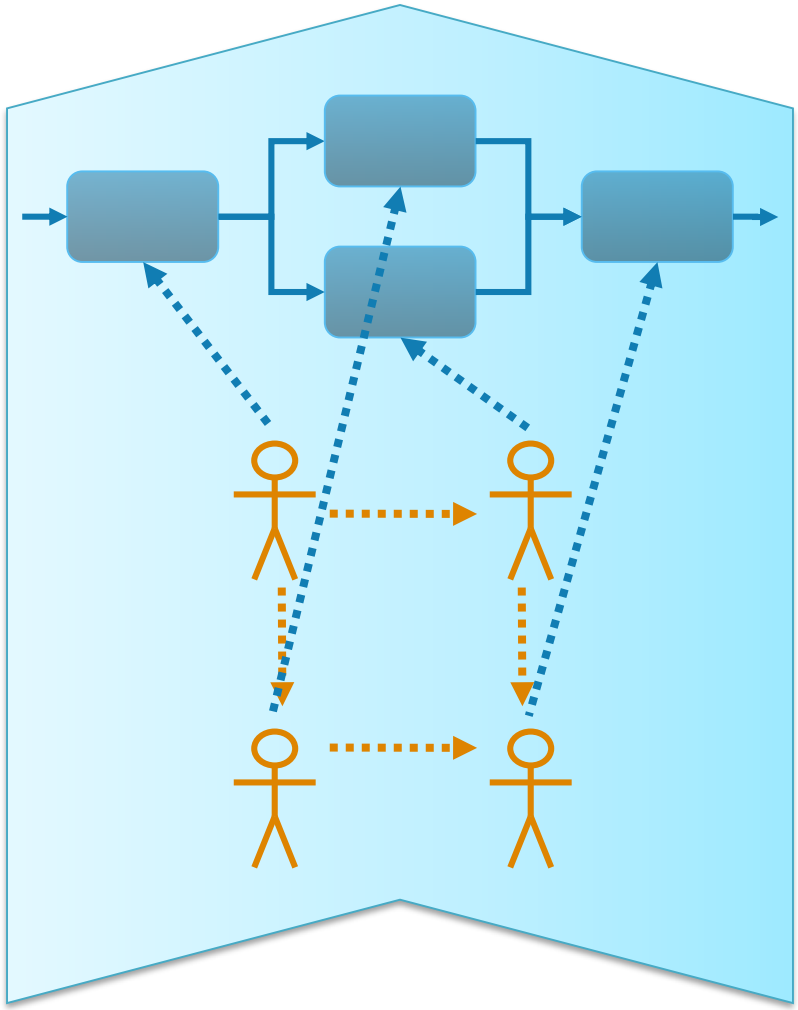


Processos

- um quebra-cabeças para a organização

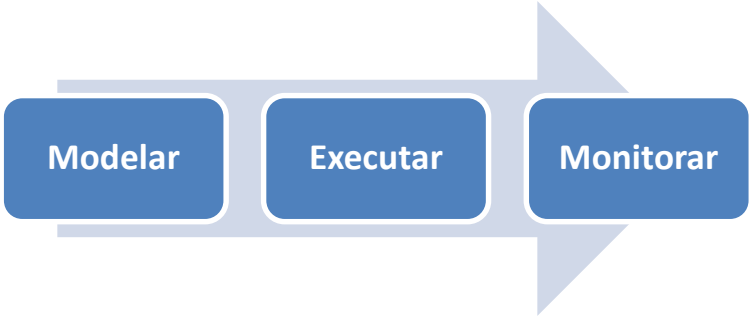
Top-down vs. bottom-up



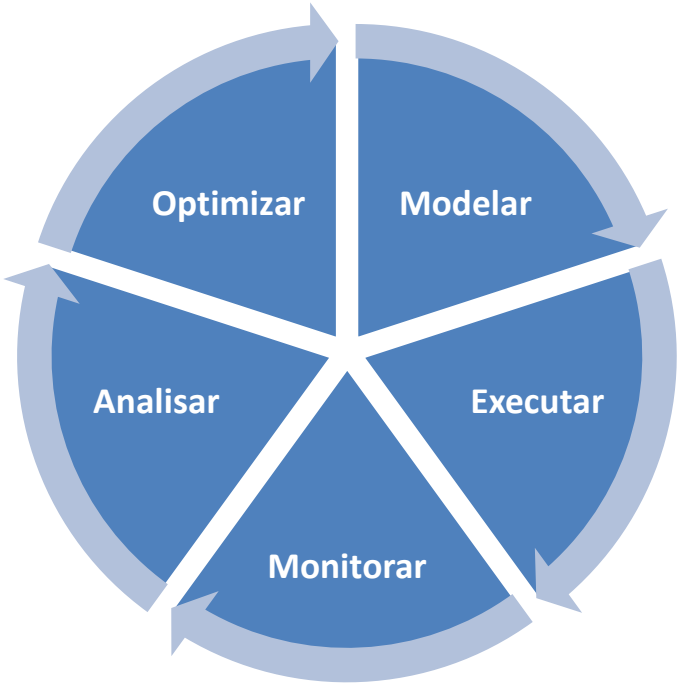


Sistemas de workflow e BPM

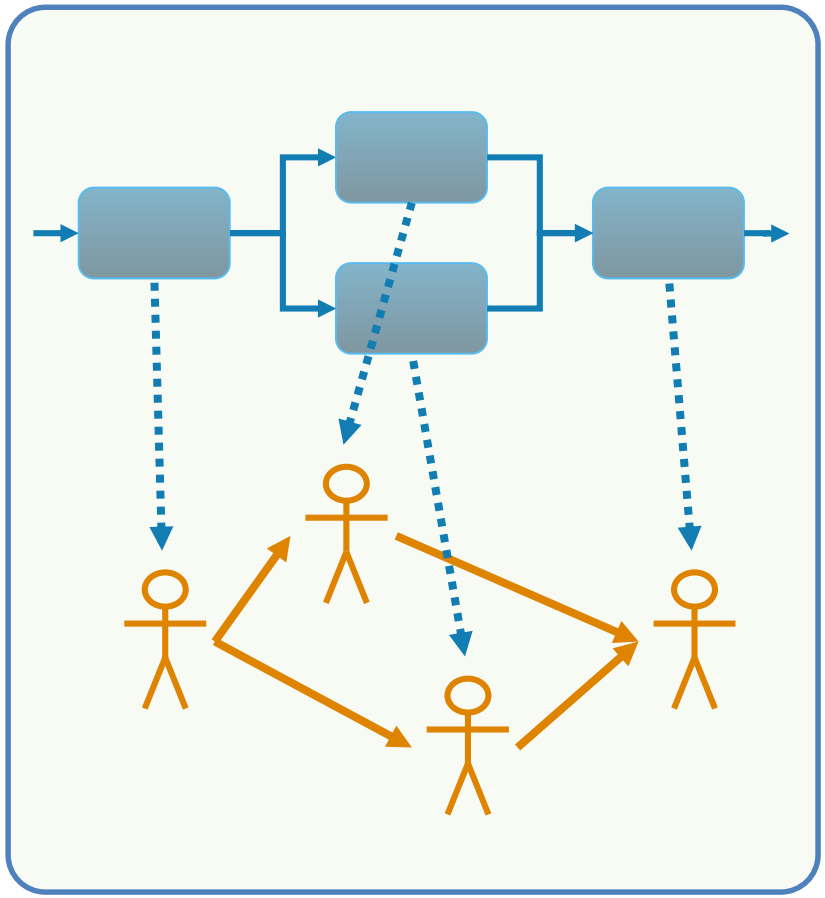
- Sistemas de workflow
  - permitem executar de forma automatizada um processo especificado



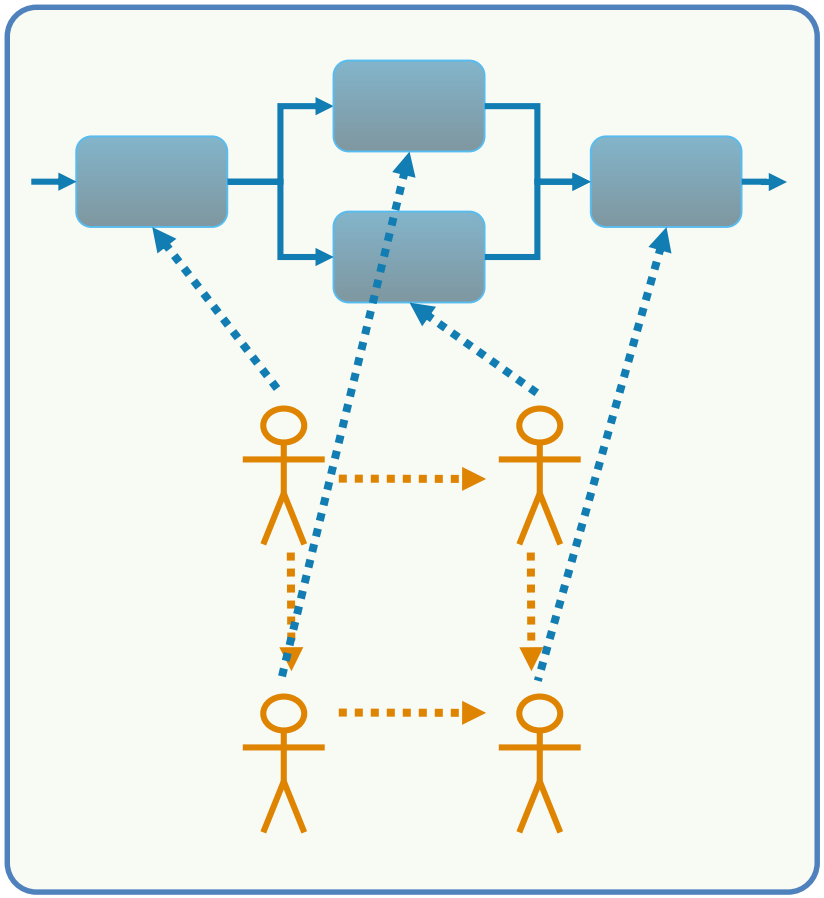
- Sistemas de BPM
  - dão suporte ao ciclo de vida de gestão e melhoria dos processos



Top-down vs. bottom-up

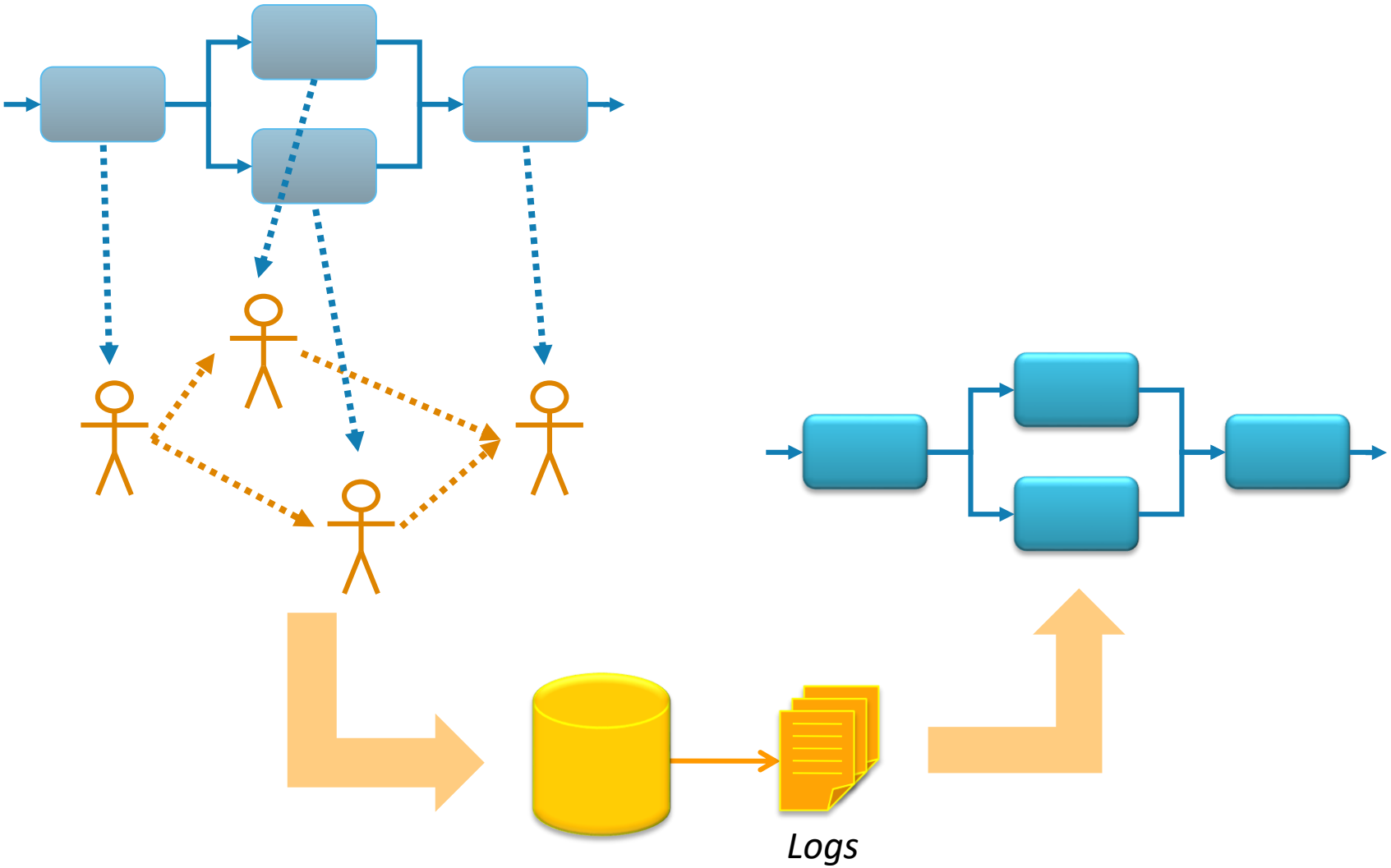


Problema resolvido!



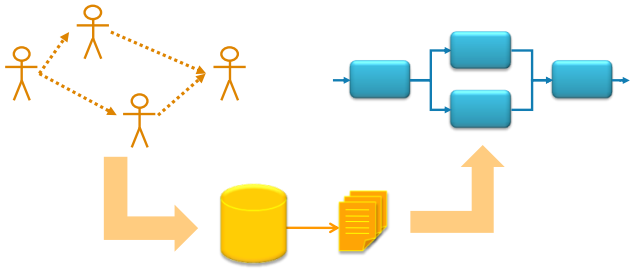
Problema por resolver…

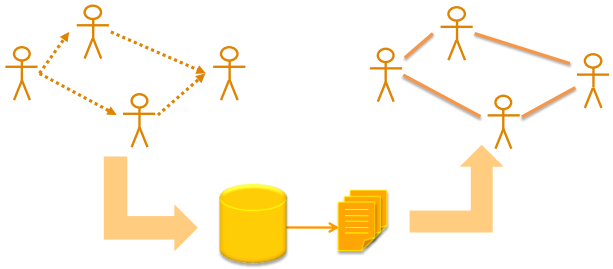
Mineração de processos

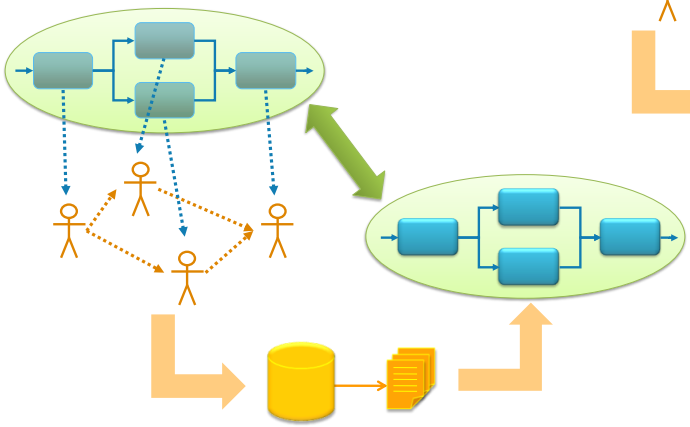


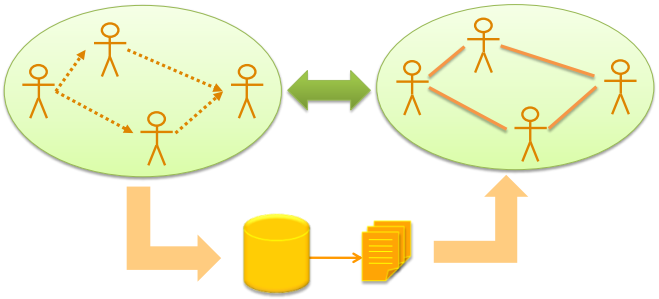
Mineração de processos

- Torna-se possível estudar:
  - descoberta de processos
  - descoberta de redes sociais
  - conformidade de processos
  - conformidade de redes sociais



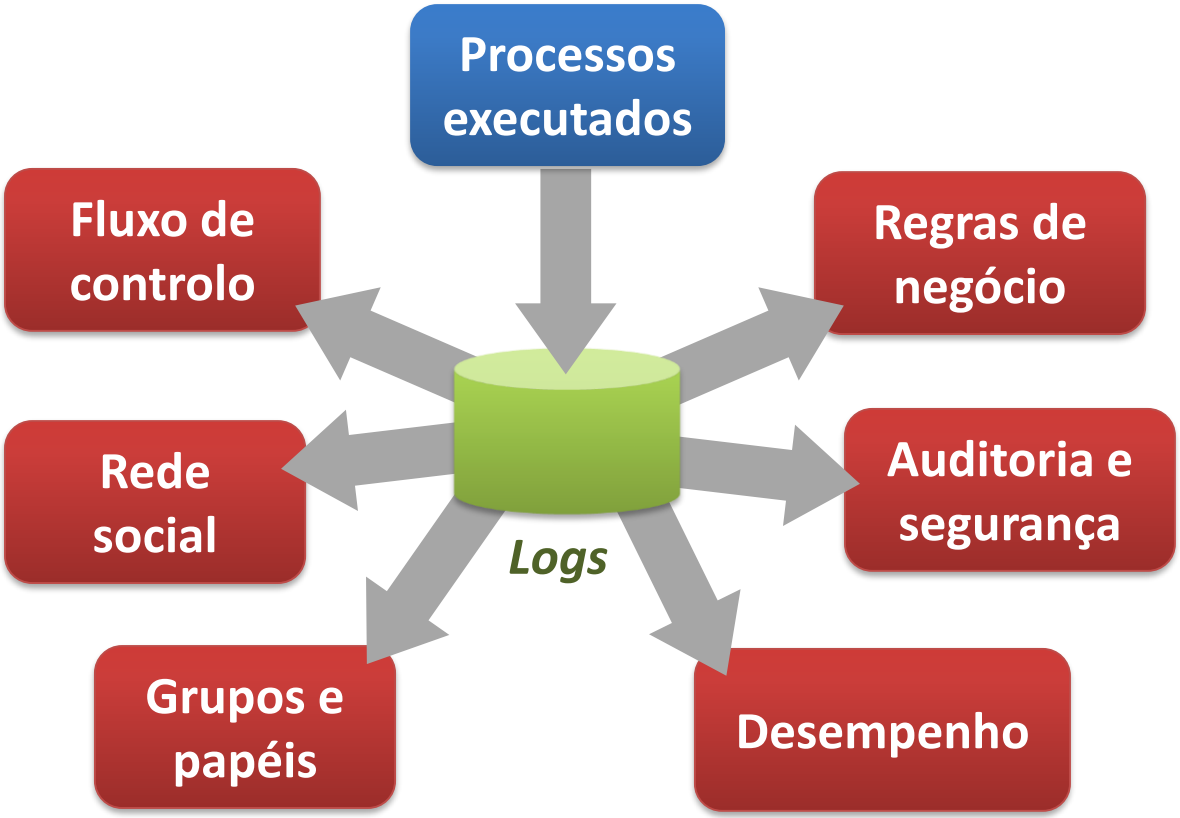






Mineração de processos

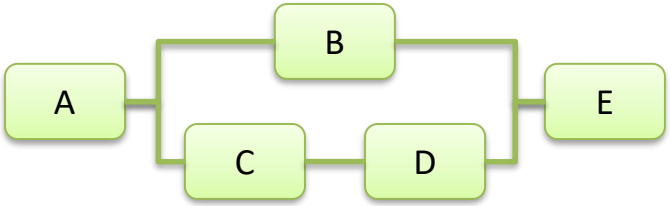
- Potencial da mineração de processos



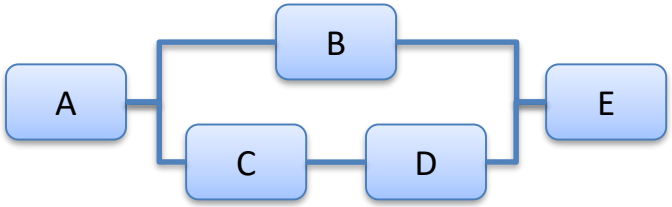
Mineração de processos

- Processos e instâncias (casos)

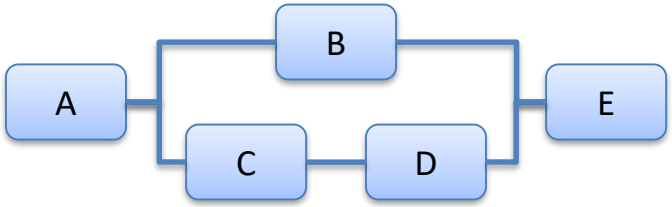
Modelo do processo



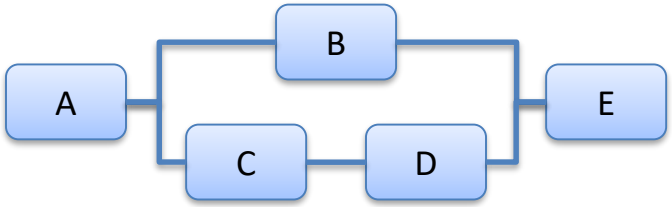
Instâncias do processo



Caso 1



Caso 2



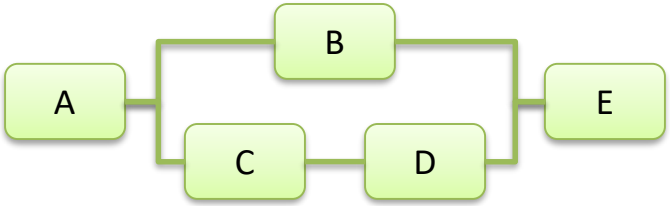
Caso 3

...

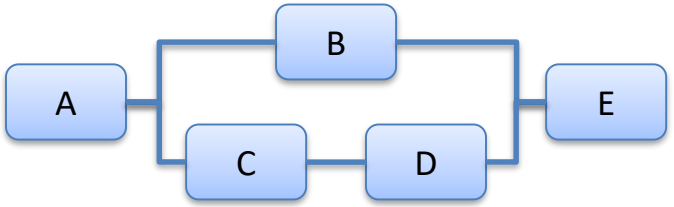
Mineração de processos

- Processos e instâncias (casos)

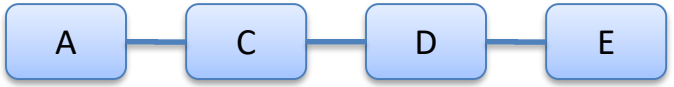
Modelo do processo



Instâncias do processo

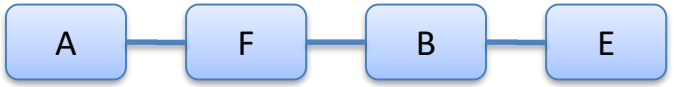


Caso 1



Caso 4

Caso 5



Caso 6

...

Registo de eventos (*Log*)

Caso	Actividade	Utilizador	Data/hora
1	Preencher requisição	João	2010-03-29 10:15
1	Aprovar requisição	Ana	2010-03-30 09:05
1	Encomendar produto	João	2010-03-30 14:20
2	Preencher requisição	Mariana	2010-04-02 11:40
1	Receber mercadoria	Pedro	2010-04-05 08:00
1	Actualizar inventário	Pedro	2010-04-05 08:10
2	Aprovar requisição	Ana	2010-04-05 09:30
2	Arquivar requisição	Pedro	2010-04-06 12:20
1	Processar pagamento	Ana	2010-04-07 08:10
3	Preencher requisição	Mariana	2010-04-09 15:40
…	…	…	…

Registo de eventos (*Log*)

	Caso	Actividade	Utilizador	Data/hora	
Caso 1	1	Preencher requisição	João	2010-03-29 10:15	
	1	Aprovar requisição	Ana	2010-03-30 09:05	
	1	Encomendar produto	João	2010-03-30 14:20	
	2	Preencher requisição	Mariana	2010-04-02 11:40	Caso 2
Caso 1	1	Receber mercadoria	Pedro	2010-04-05 08:00	
	1	Actualizar inventário	Pedro	2010-04-05 08:10	
	2	Aprovar requisição	Ana	2010-04-05 09:30	
	2	Arquivar requisição	Pedro	2010-04-06 12:20	Caso 2
Caso 1	1	Processar pagamento	Ana	2010-04-07 08:10	
	3	Preencher requisição	Mariana	2010-04-09 15:40	Caso 3
	…	…	…	…	

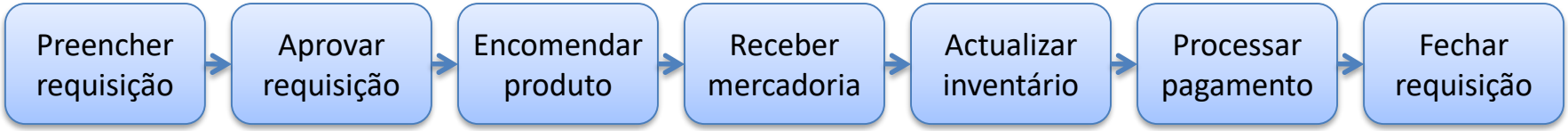
Mineração de processos

- Perspectivas de mineração
  - mineração do fluxo
  - mineração da rede social

Mineração do fluxo

- Caso 1

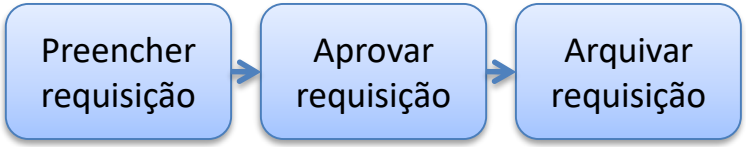
Caso	Actividade
1	Preencher requisição
1	Aprovar requisição
1	Encomendar produto
1	Receber mercadoria
1	Actualizar inventário
1	Processar pagamento
1	Fechar requisição



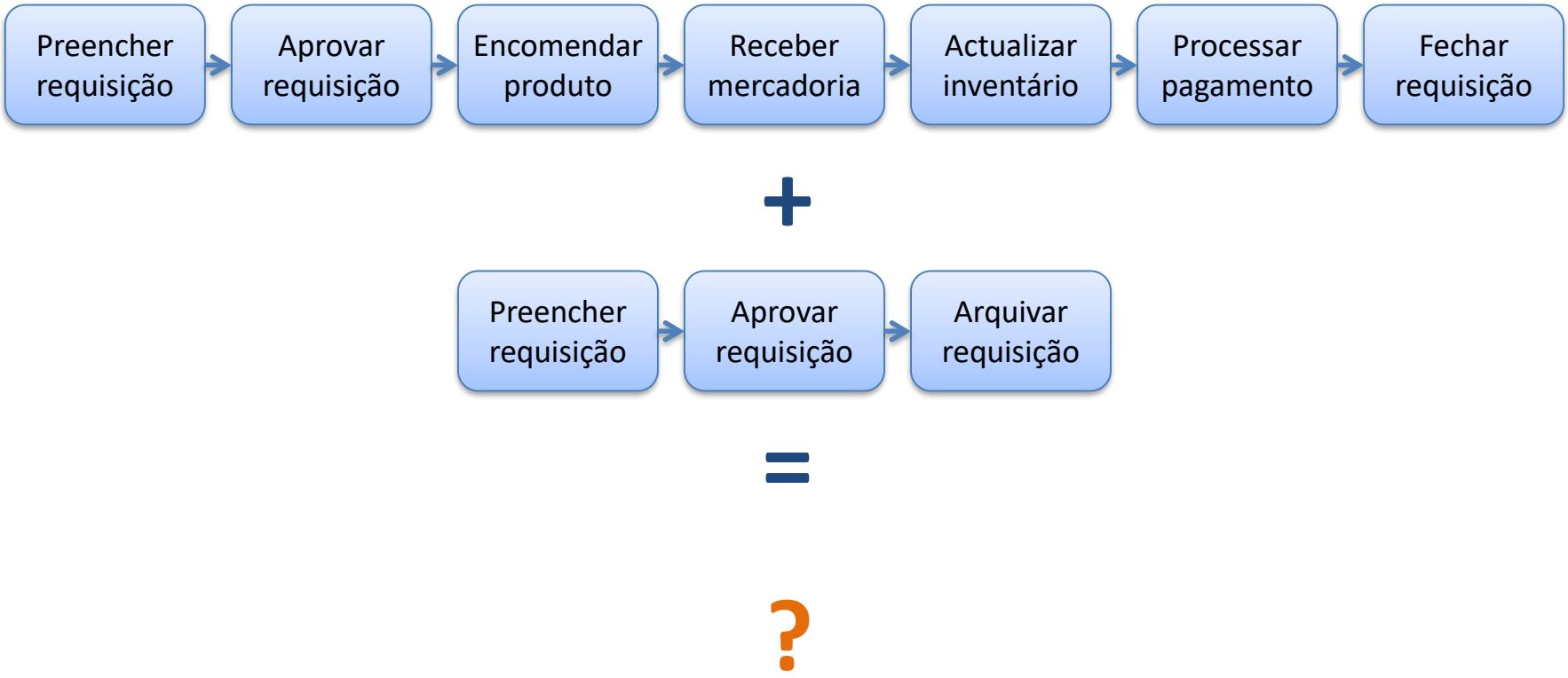
Mineração do fluxo

- Caso 2

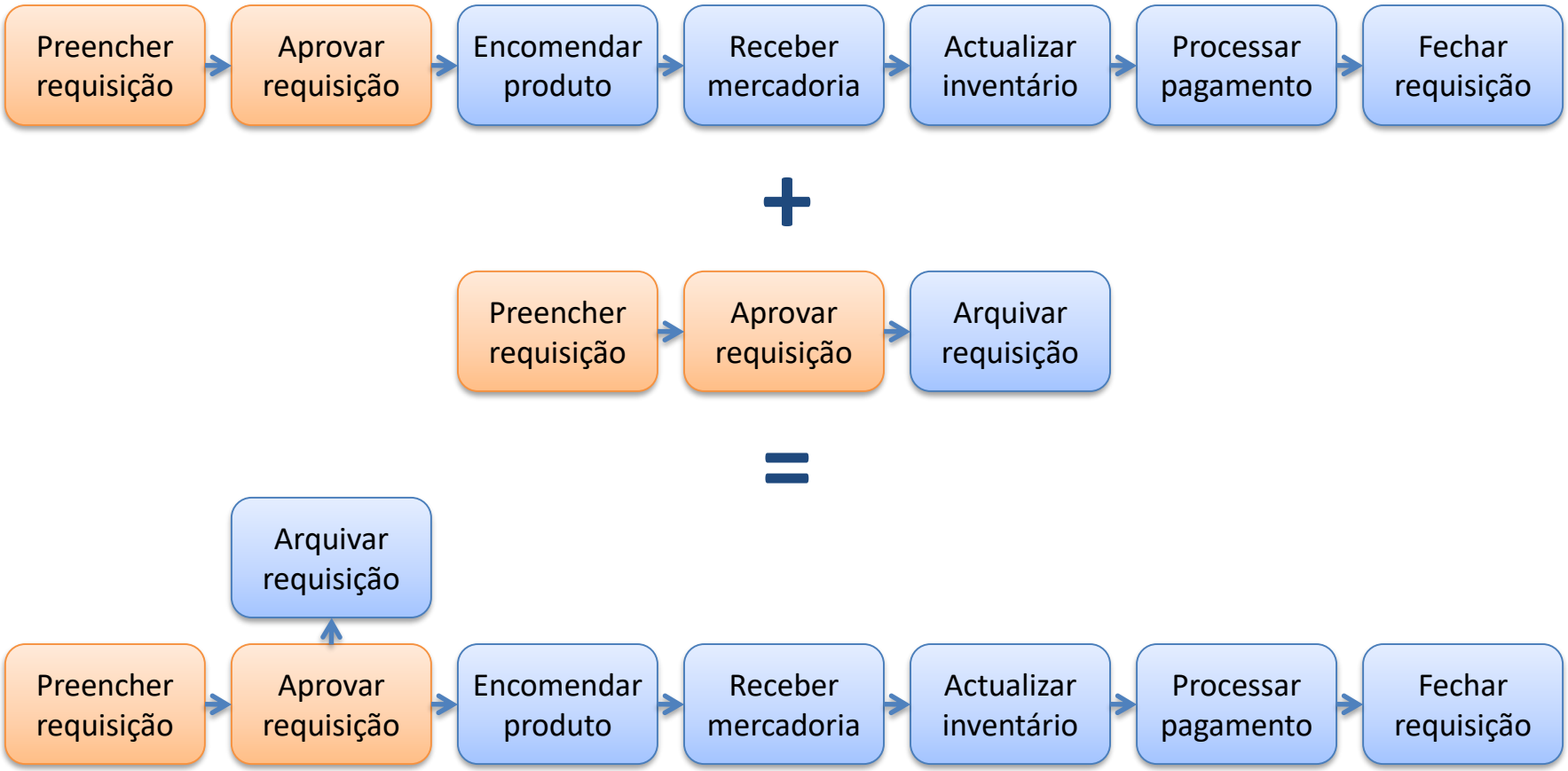
Caso	Actividade
2	Preencher requisição
2	Aprovar requisição
2	Arquivar requisição



Mineração do fluxo



Mineração do fluxo



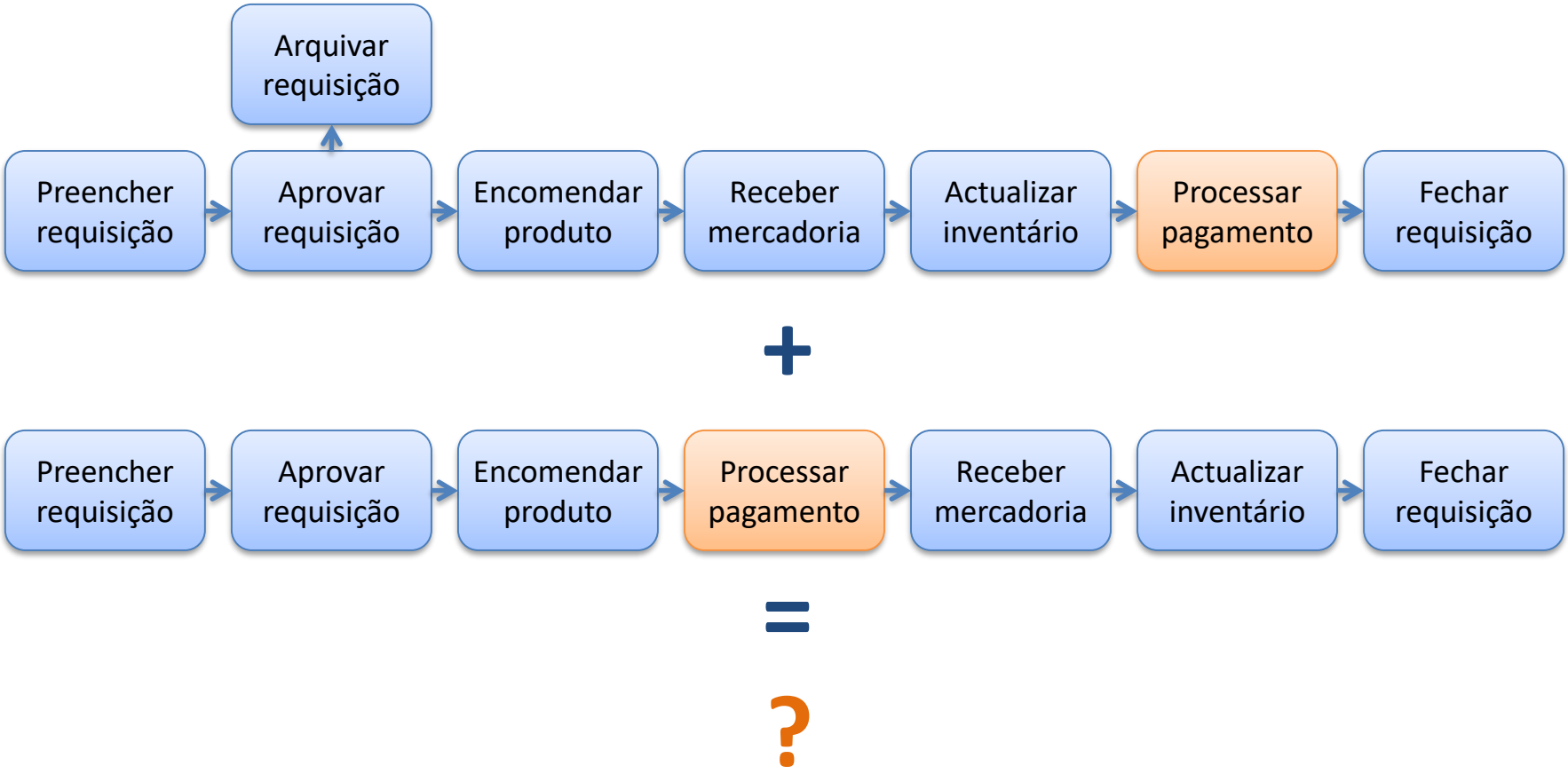
Mineração do fluxo

- Caso 3

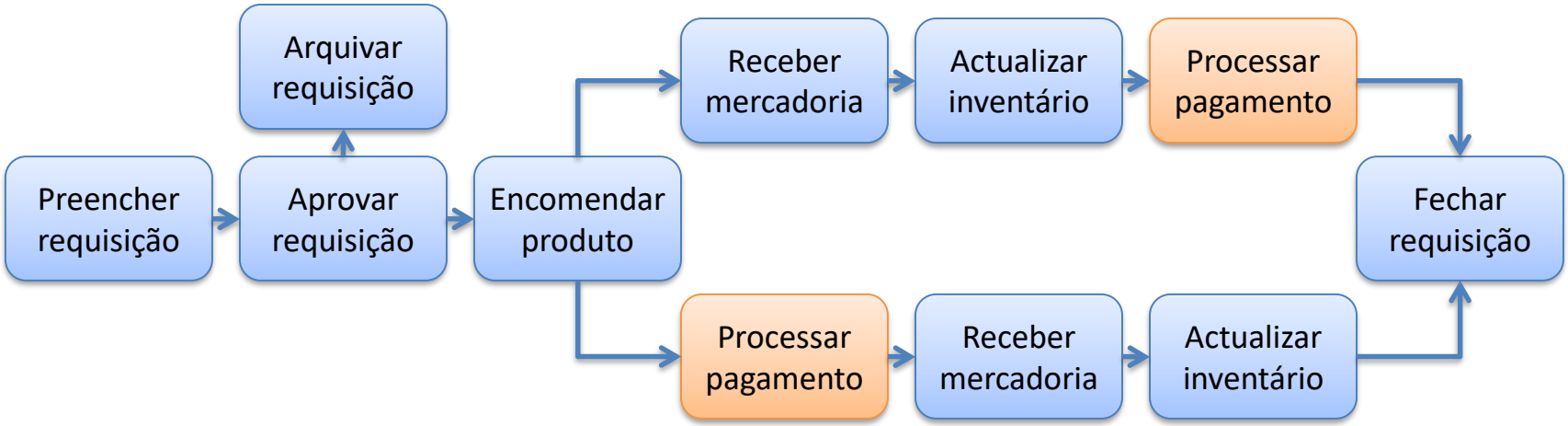
Caso	Actividade
1	Preencher requisição
1	Aprovar requisição
1	Encomendar produto
1	Receber mercadoria
1	Actualizar inventário
1	Processar pagamento
1	Fechar requisição

Caso	Actividade
3	Preencher requisição
3	Aprovar requisição
3	Encomendar produto
3	Processar pagamento
3	Receber mercadoria
3	Actualizar inventário
3	Fechar requisição

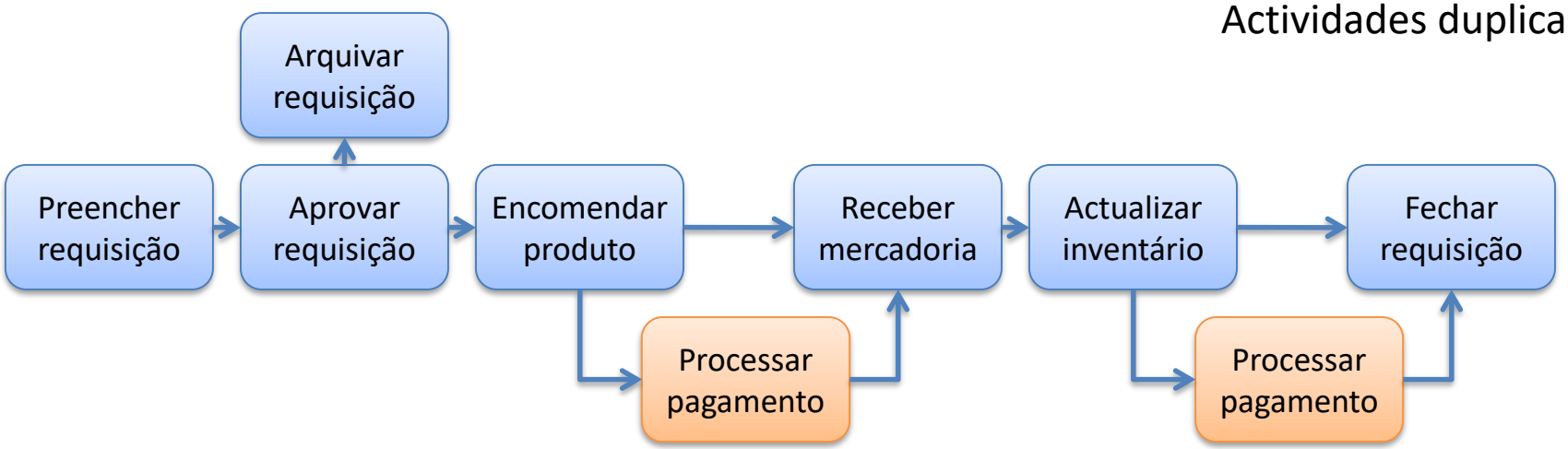
Mineração do fluxo



Mineração do fluxo

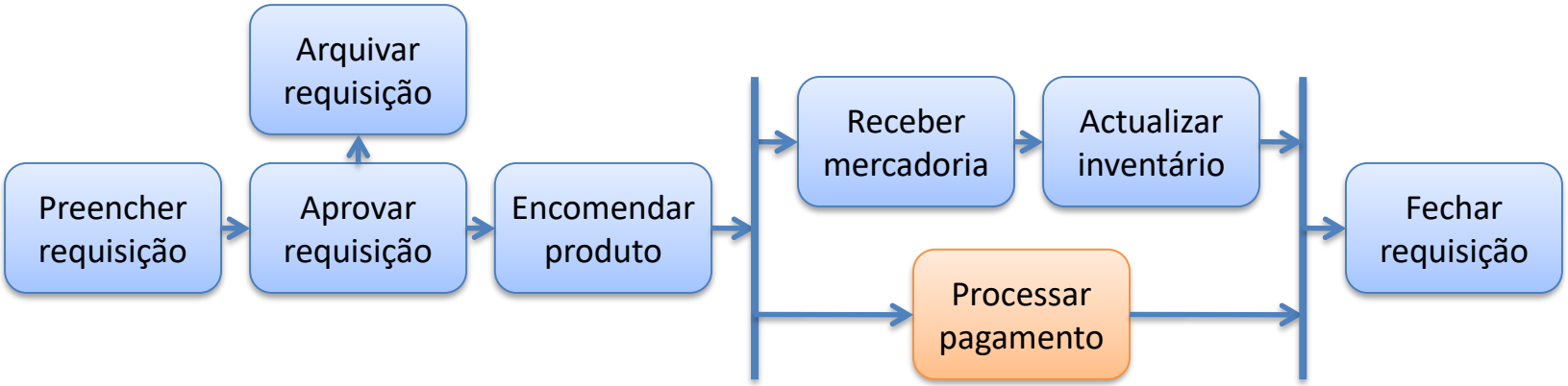


Actividades duplicadas

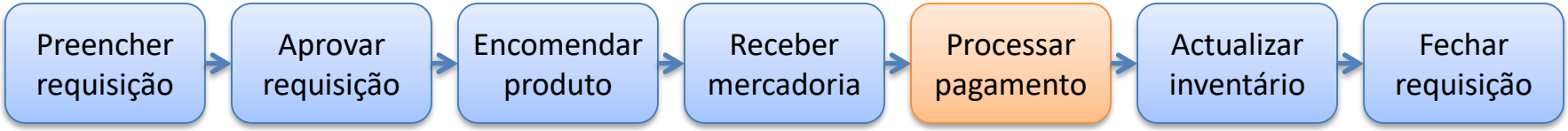


Dependências não-locais

Mineração do fluxo



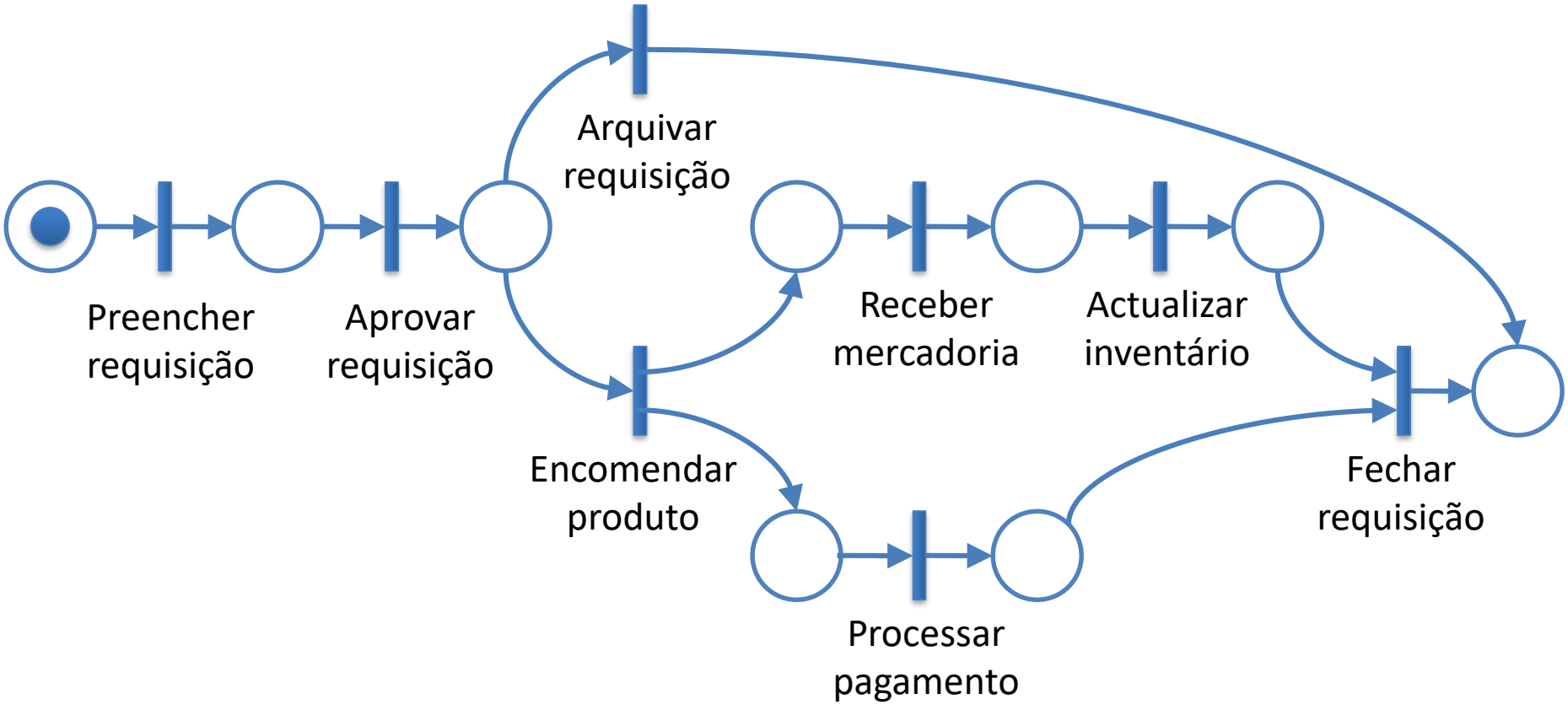
Paralelismo



(possibilidade não observada no *log*)

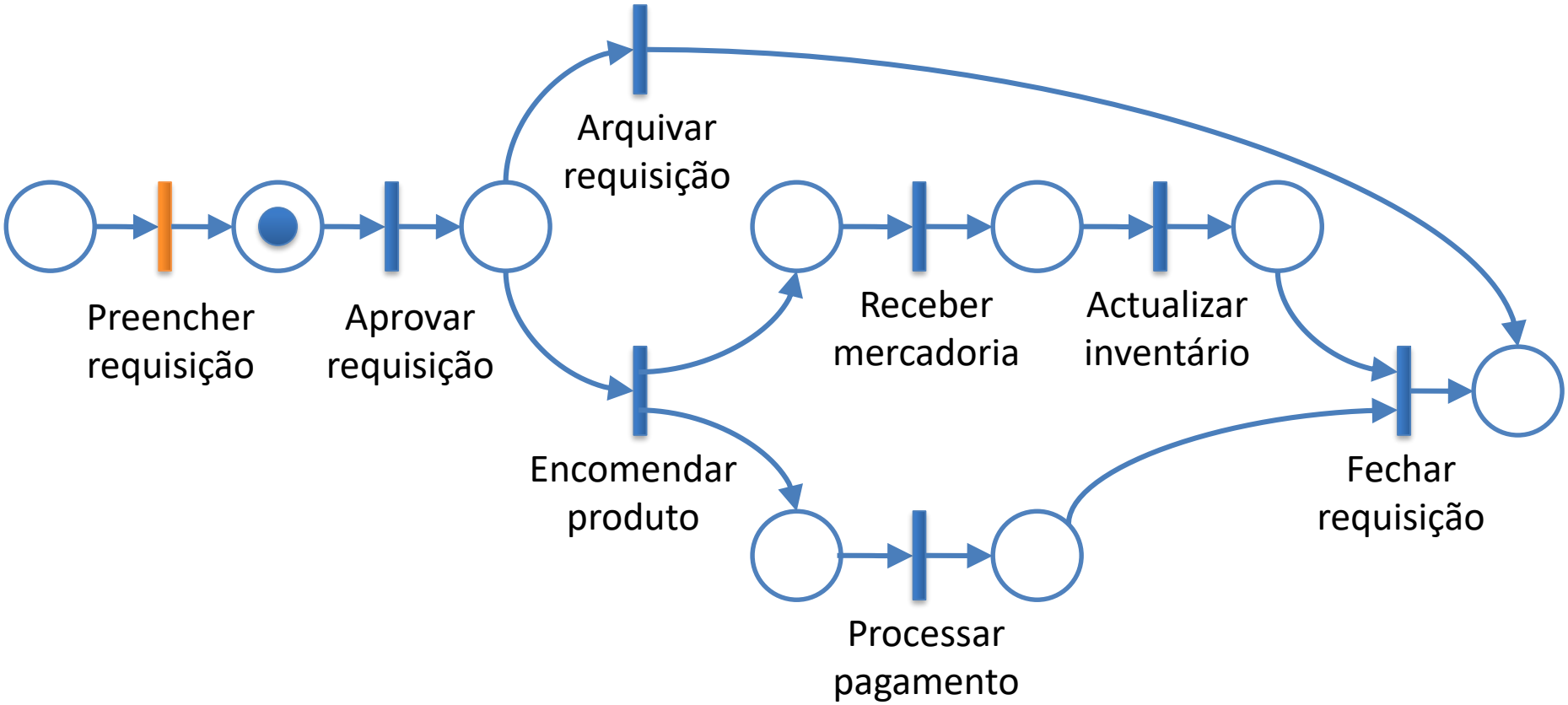
Mineração do fluxo

- Modelo em rede de Petri



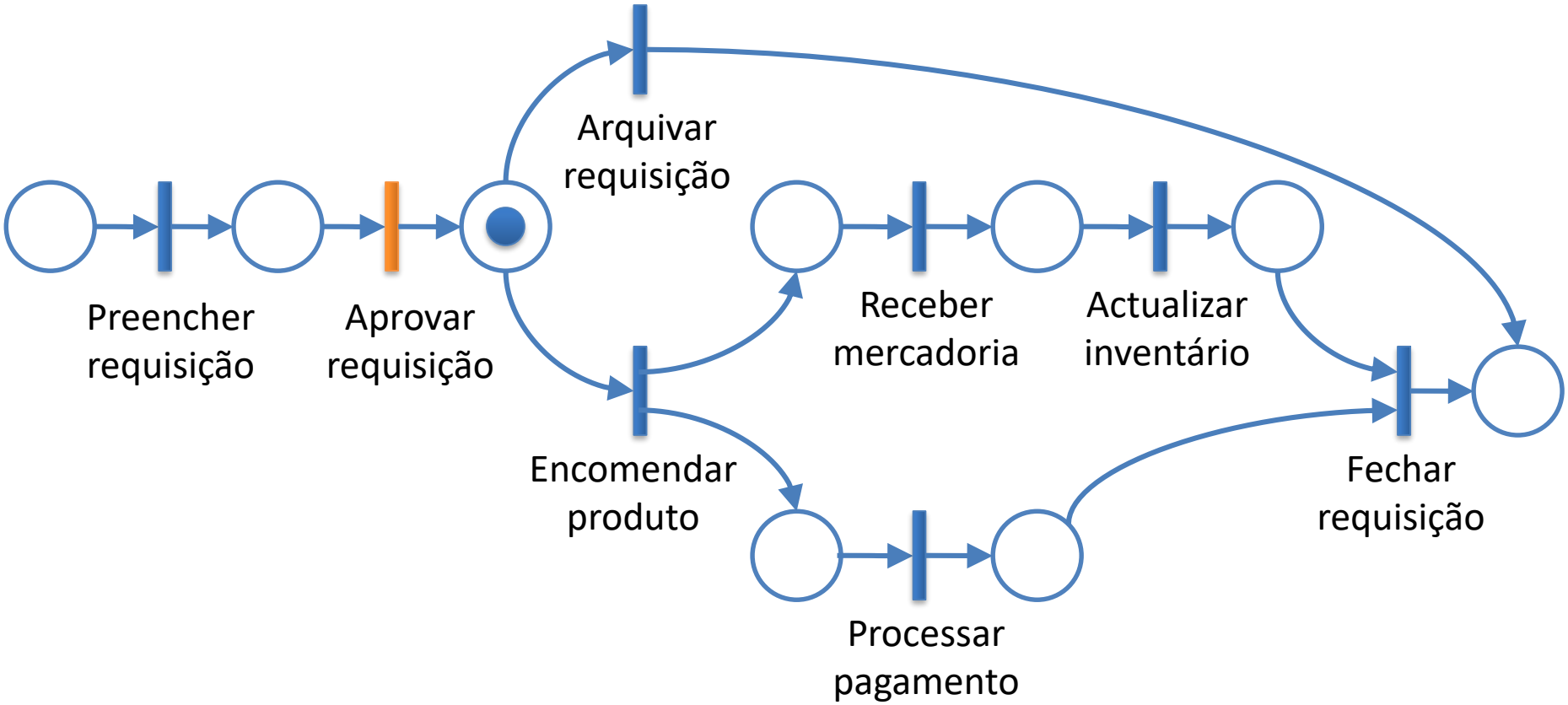
Mineração do fluxo

- Modelo em rede de Petri



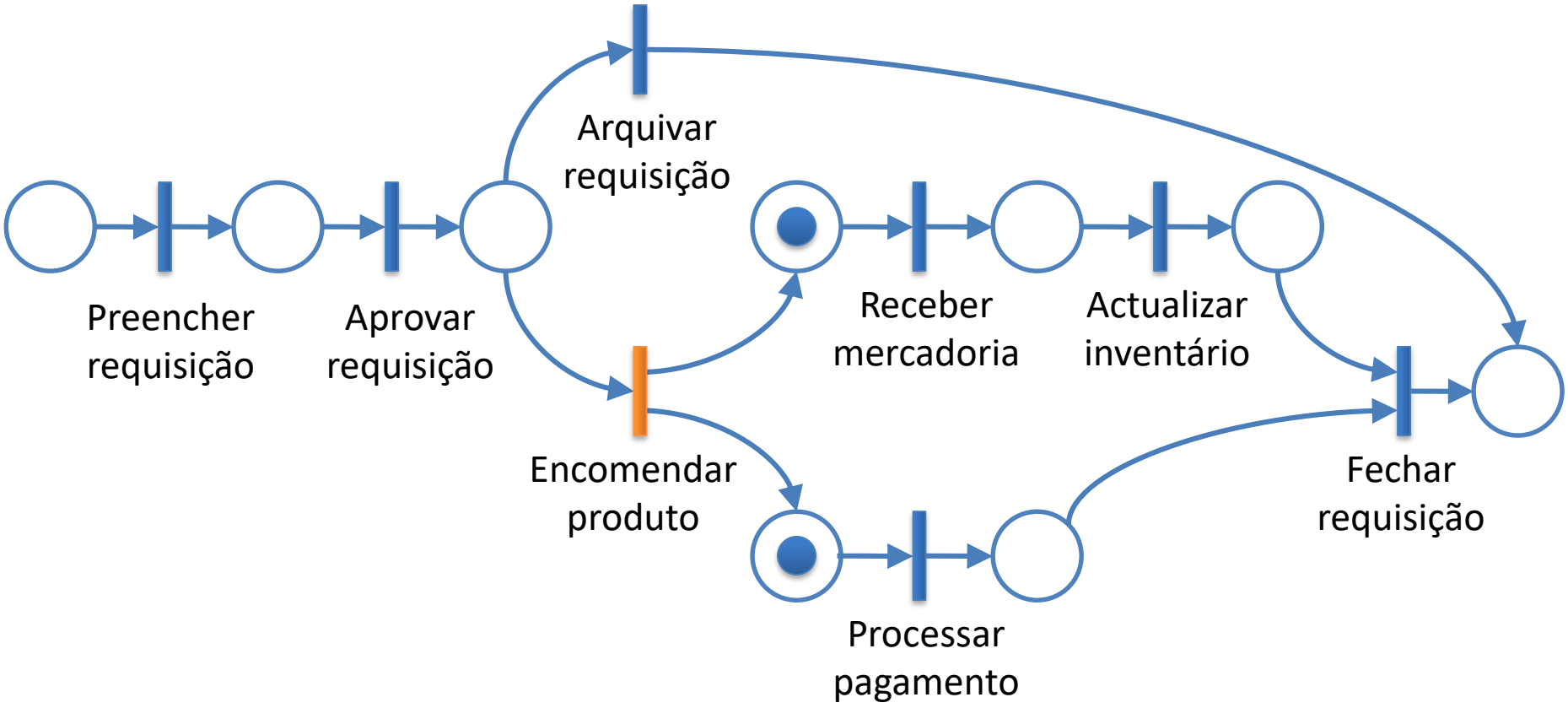
Mineração do fluxo

- Modelo em rede de Petri



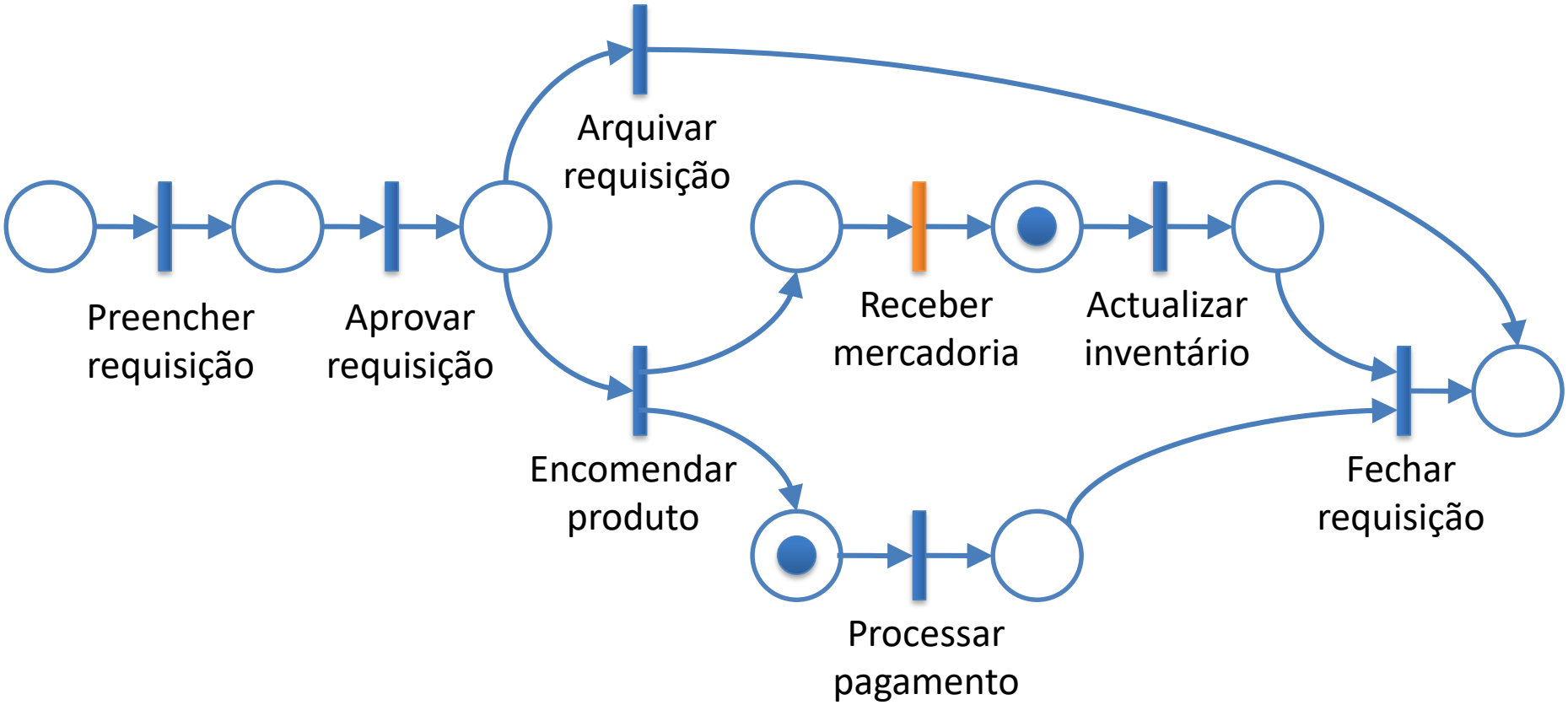
Mineração do fluxo

- Modelo em rede de Petri



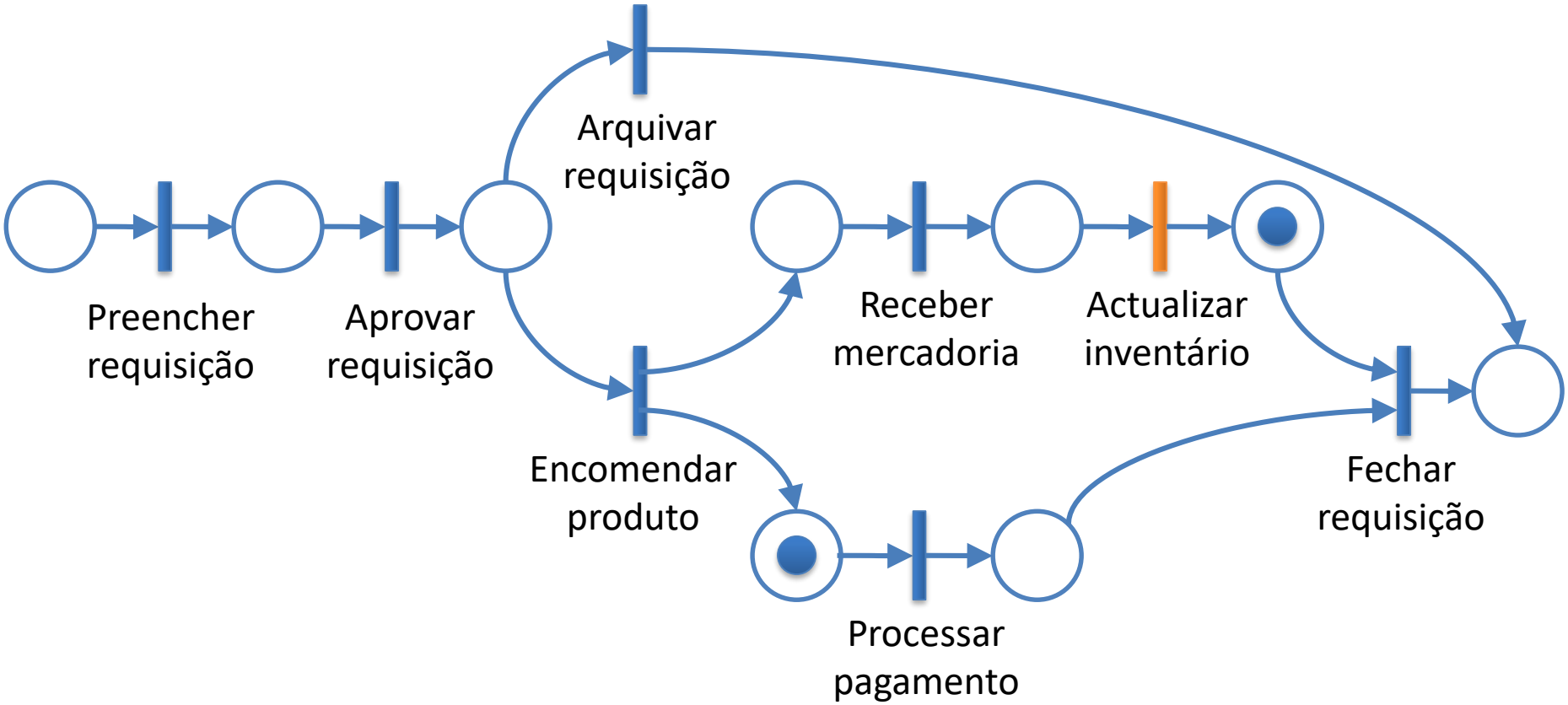
Mineração do fluxo

- Modelo em rede de Petri



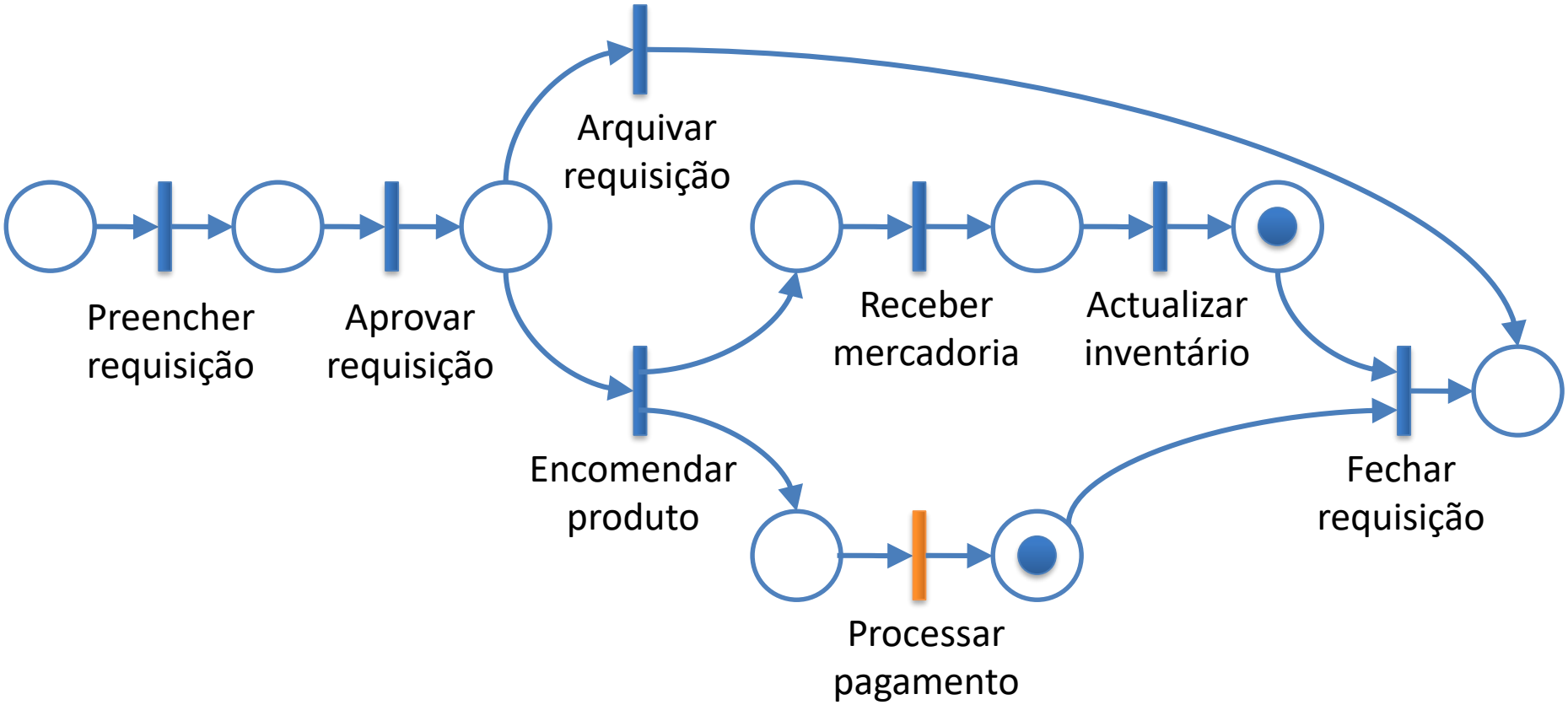
Mineração do fluxo

- Modelo em rede de Petri



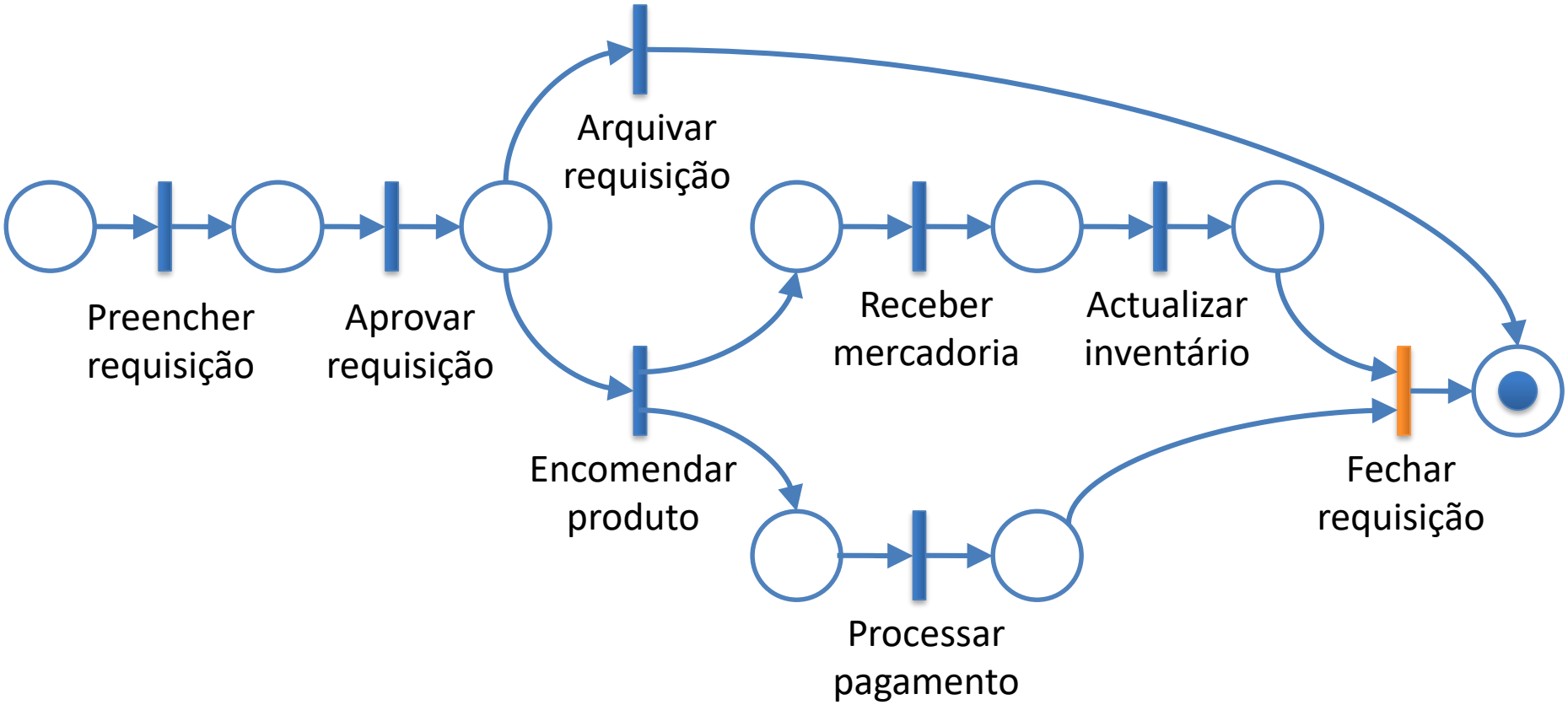
Mineração do fluxo

- Modelo em rede de Petri



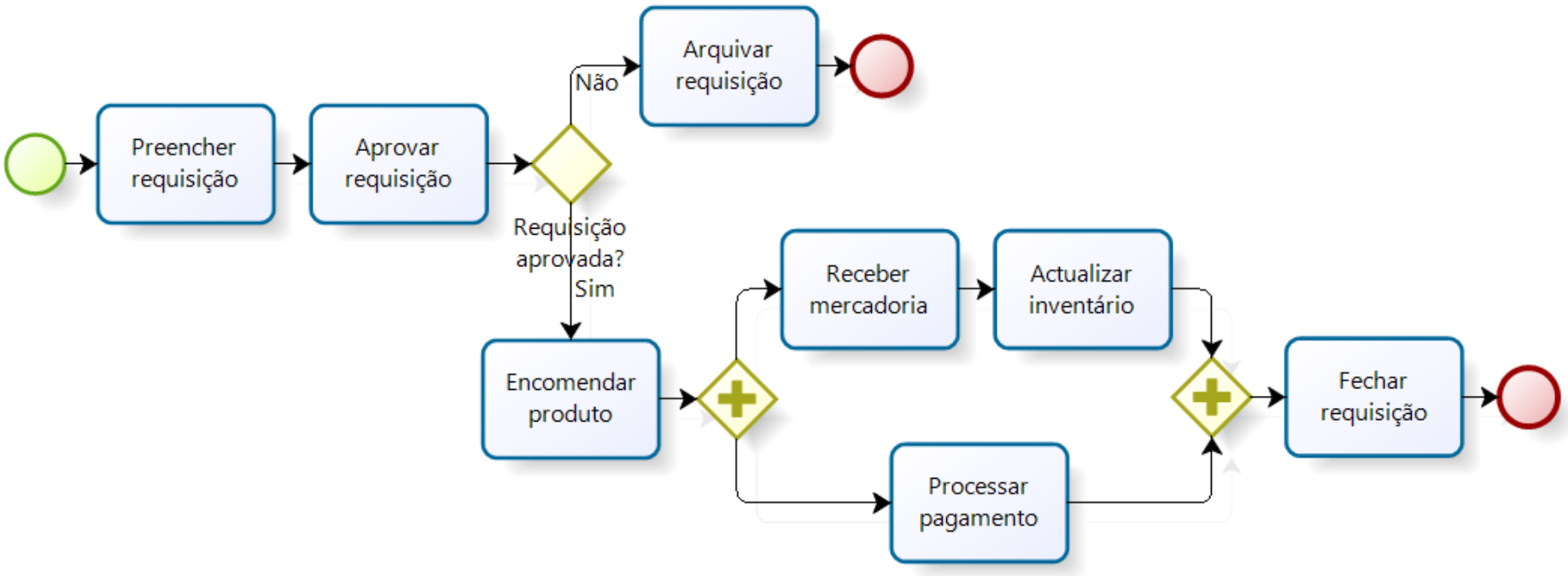
Mineração do fluxo

- Modelo em rede de Petri



Mineração do fluxo

- Modelo em BPMN

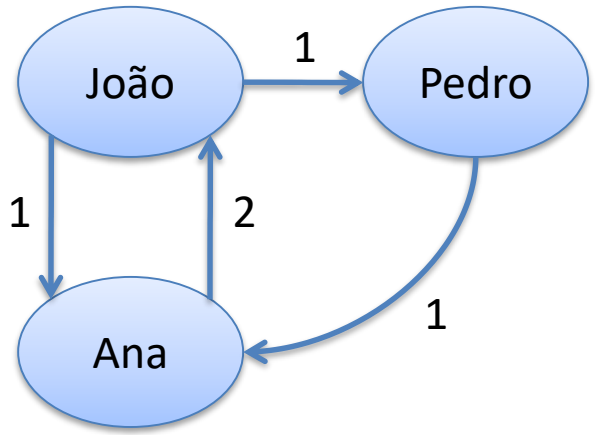


Mineração da rede social

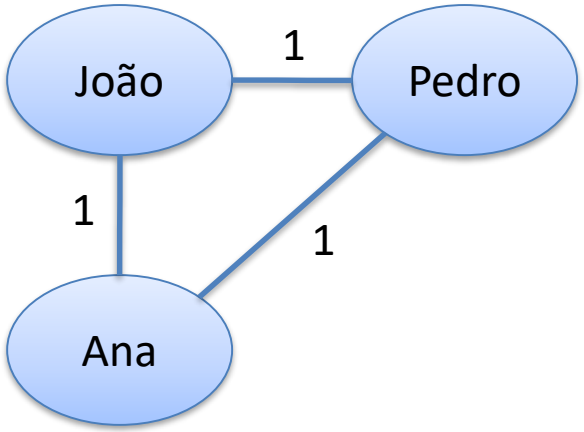
- Caso 1

Caso	Actividade	Utilizador
1	Preencher requisição	João
1	Aprovar requisição	Ana
1	Encomendar produto	João
1	Receber mercadoria	Pedro
1	Actualizar inventário	Pedro
1	Processar pagamento	Ana
1	Fechar requisição	João

Handover of work



Working together

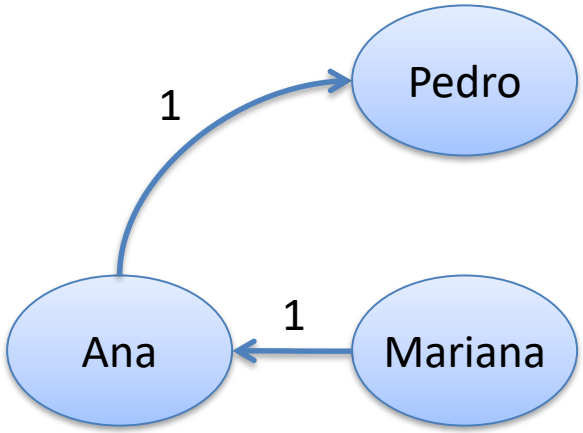


Mineração da rede social

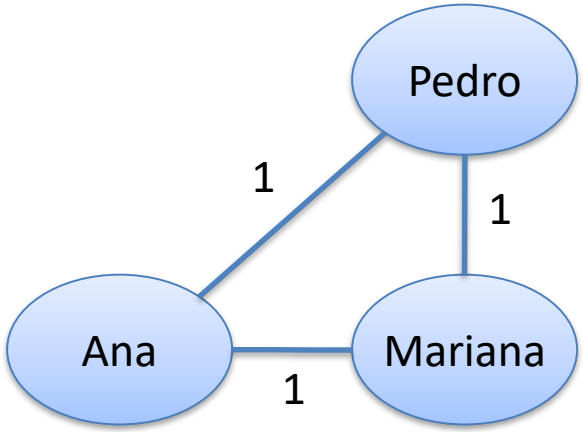
- Caso 2

Caso	Actividade	Utilizador
2	Preencher requisição	Mariana
2	Aprovar requisição	Ana
2	Arquivar requisição	Pedro

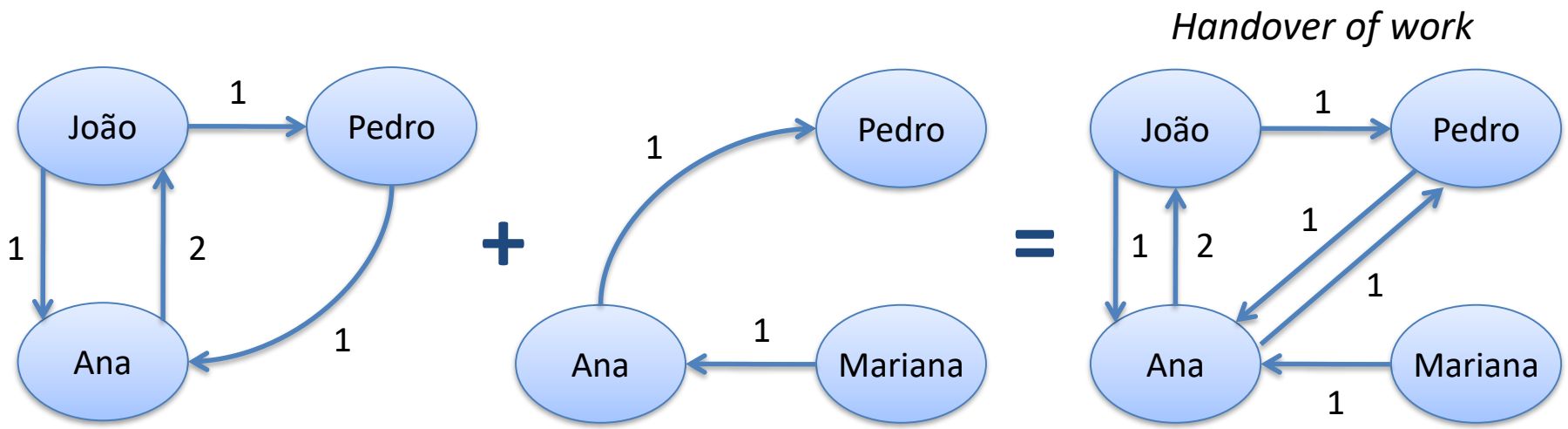
Handover of work

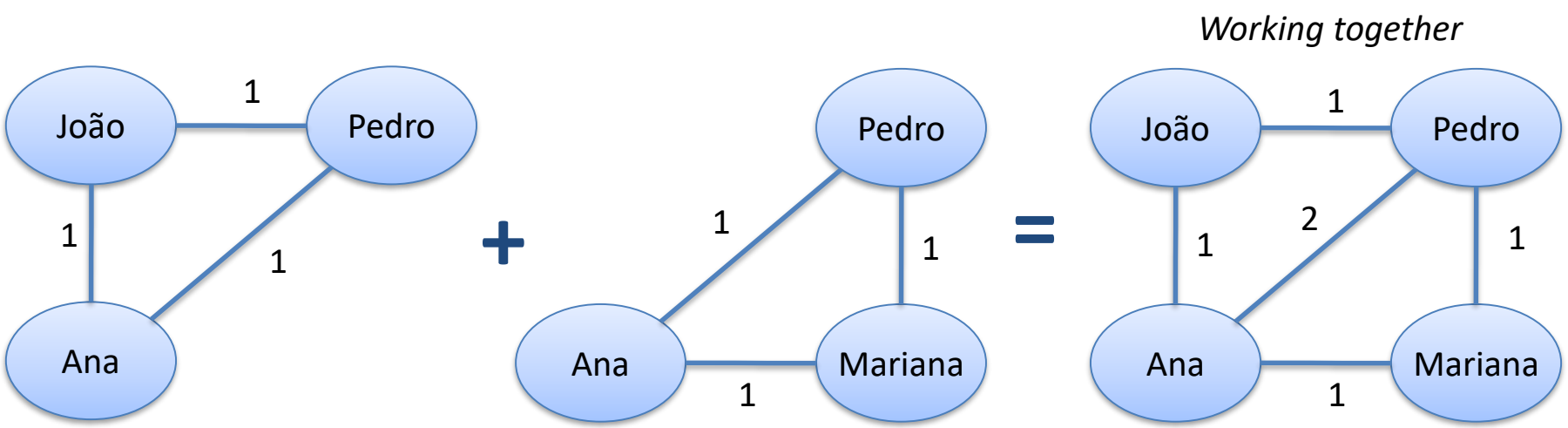


Working together



Mineração da rede social



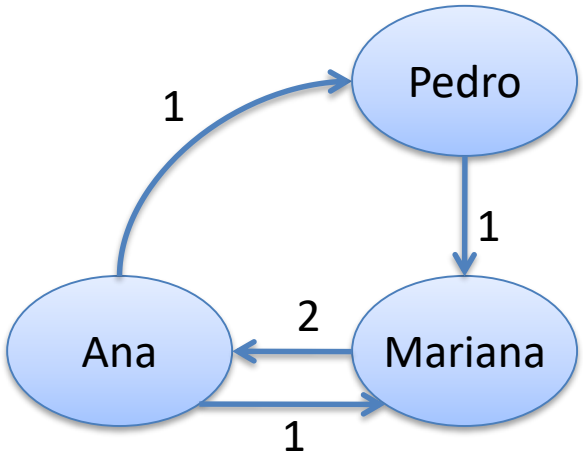


Mineração da rede social

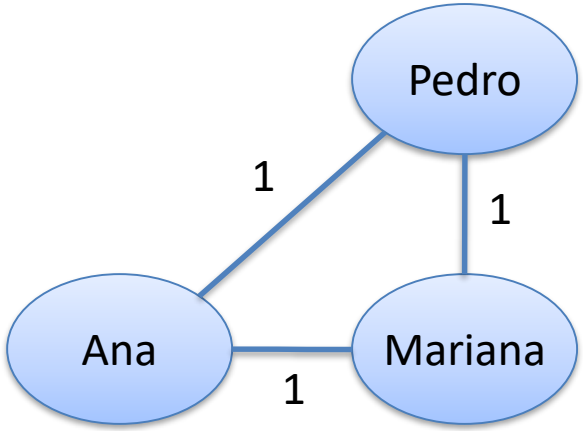
- Caso 3

Caso	Actividade	Utilizador
3	Preencher requisição	Mariana
3	Aprovar requisição	Ana
3	Encomendar produto	Mariana
3	Processar pagamento	Ana
3	Receber mercadoria	Pedro
3	Actualizar inventário	Pedro
3	Fechar requisição	Mariana

Handover of work

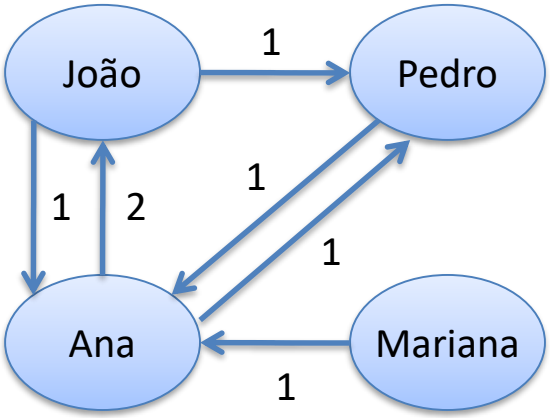


Working together

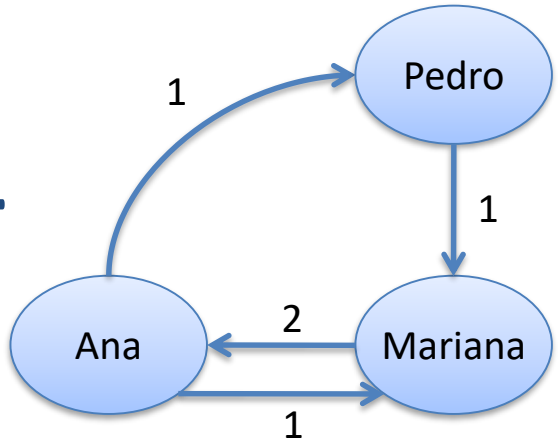


Mineração da rede social

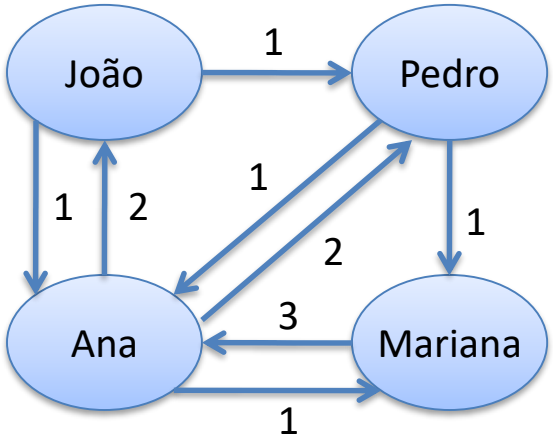
Casos 1 e 2

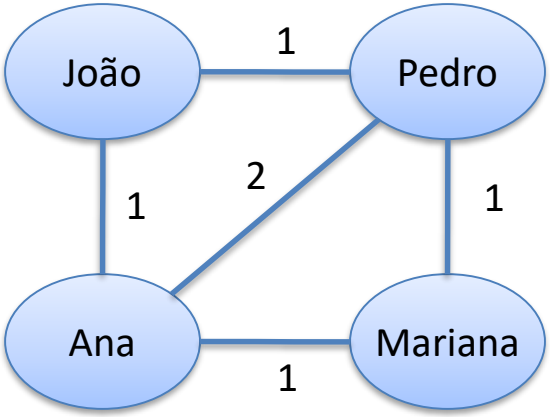


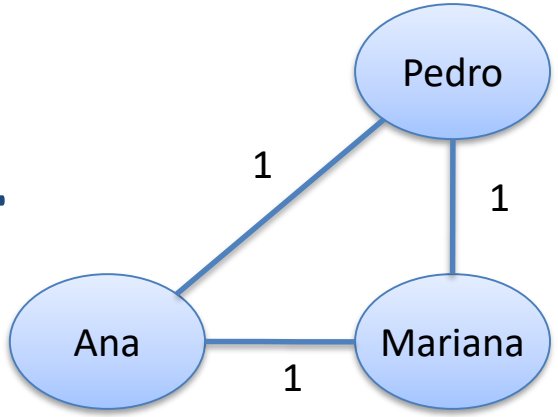
Caso 3



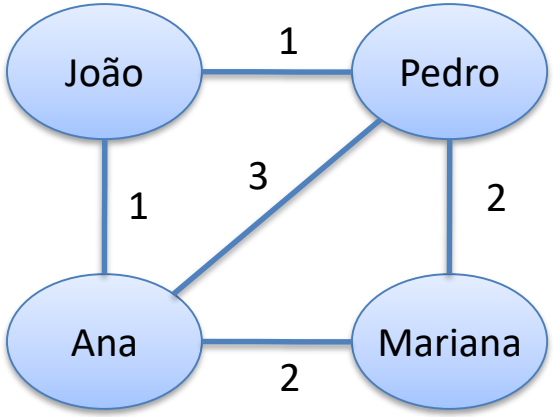
Handover of work





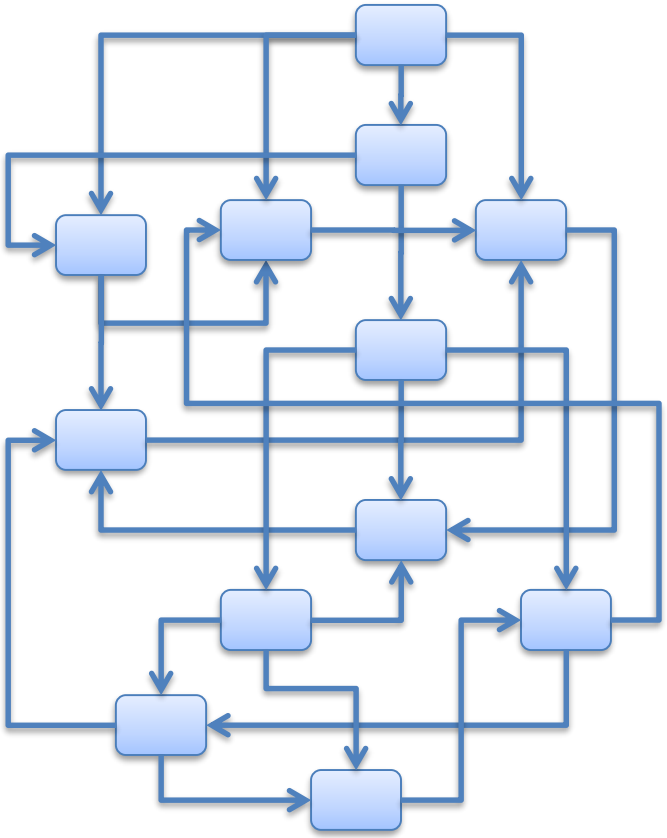


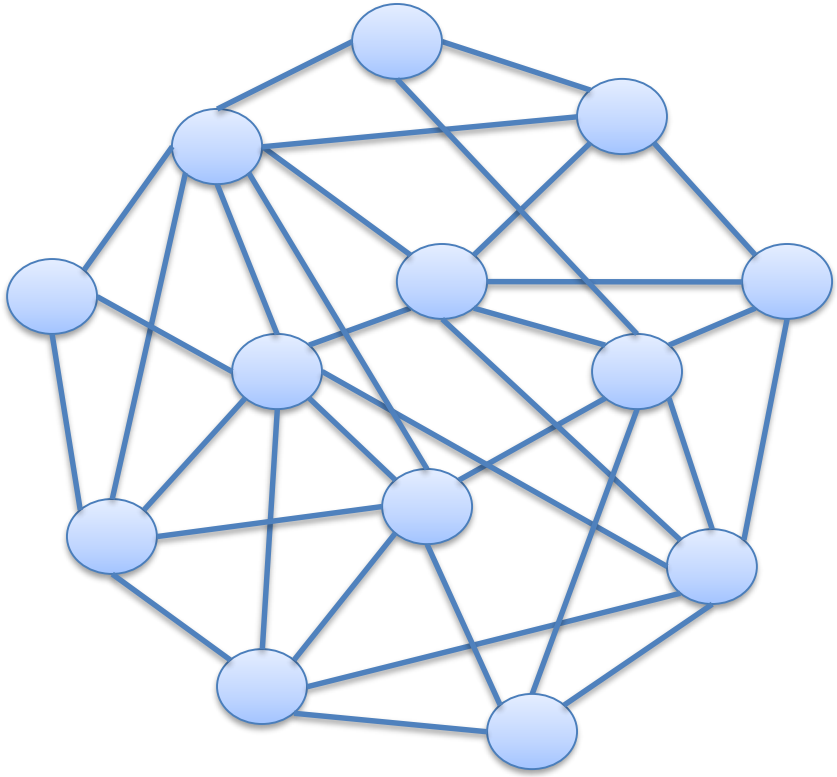
Working together



Mineração de processos

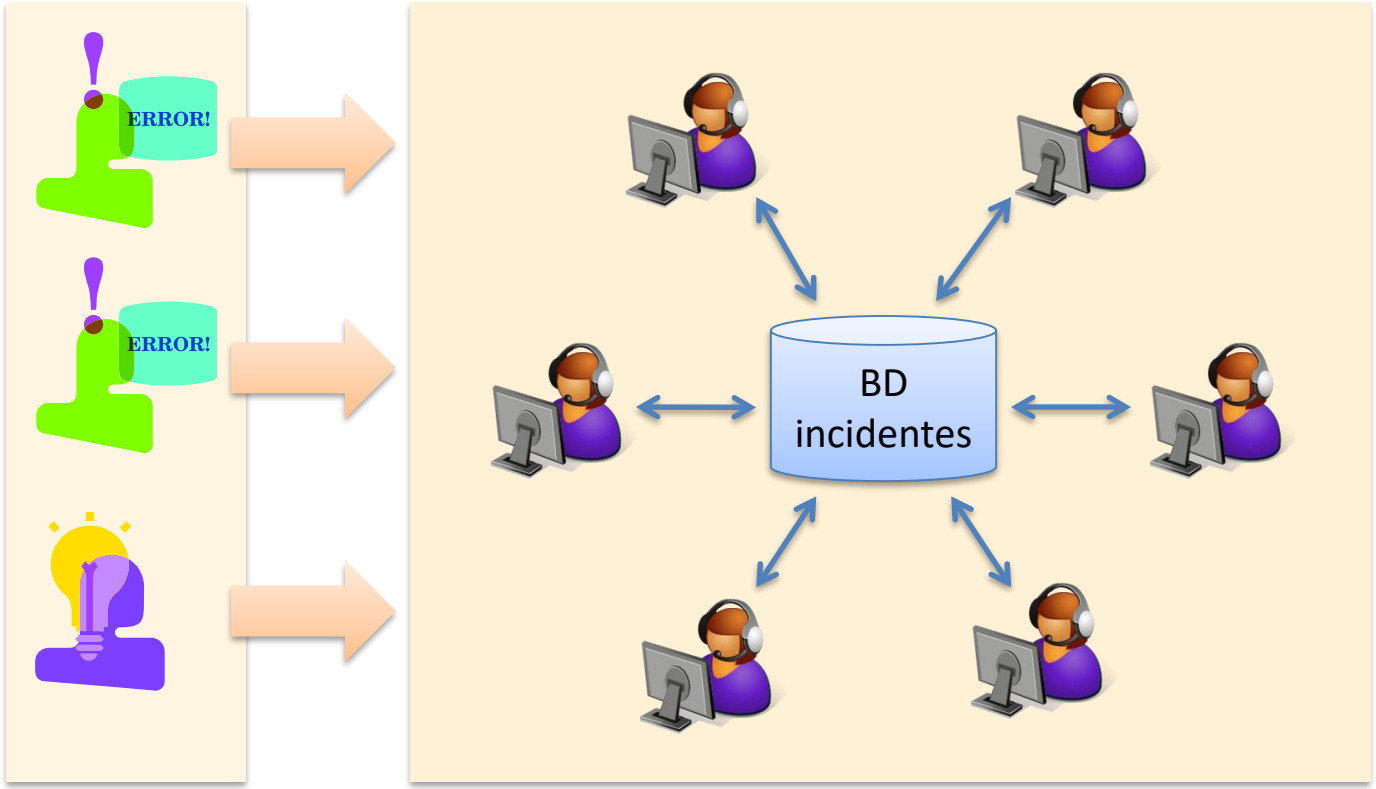
- Em aplicações reais, podemos ter modelos complexos e muito confusos





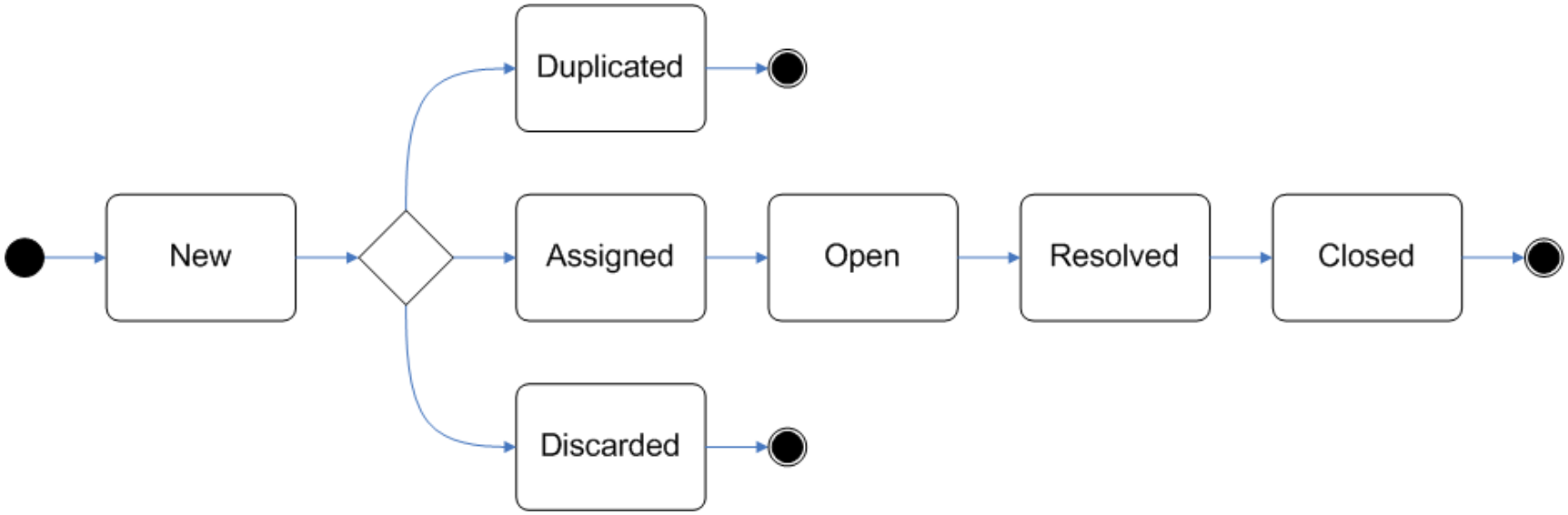
Caso prático I

- Suporte técnico de um produto de software



Caso prático I

- Estado dos incidentes

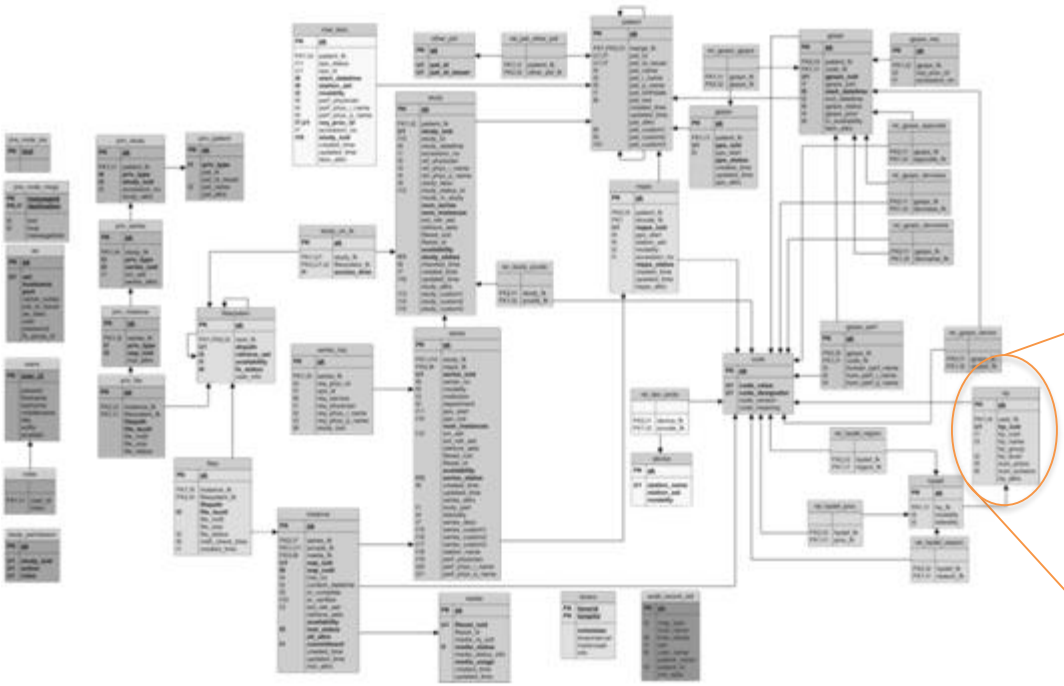


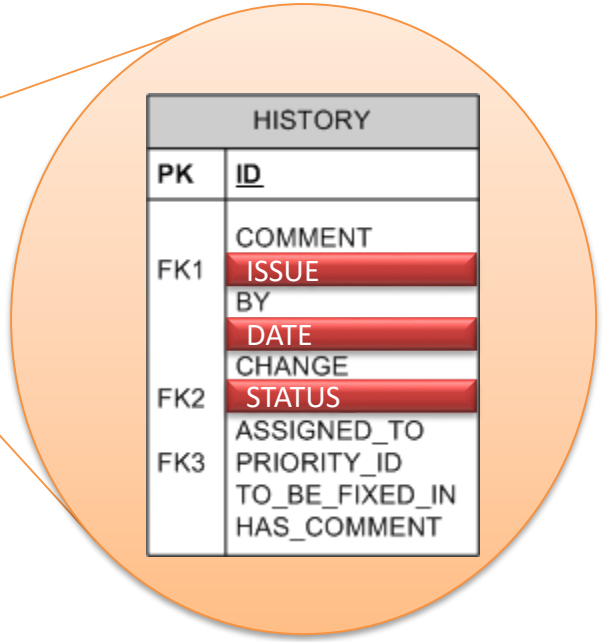
Recording. Matching. Classification. Diagnosis. Resolution. Closure.

(*ITIL Incident Management*)

Caso prático I

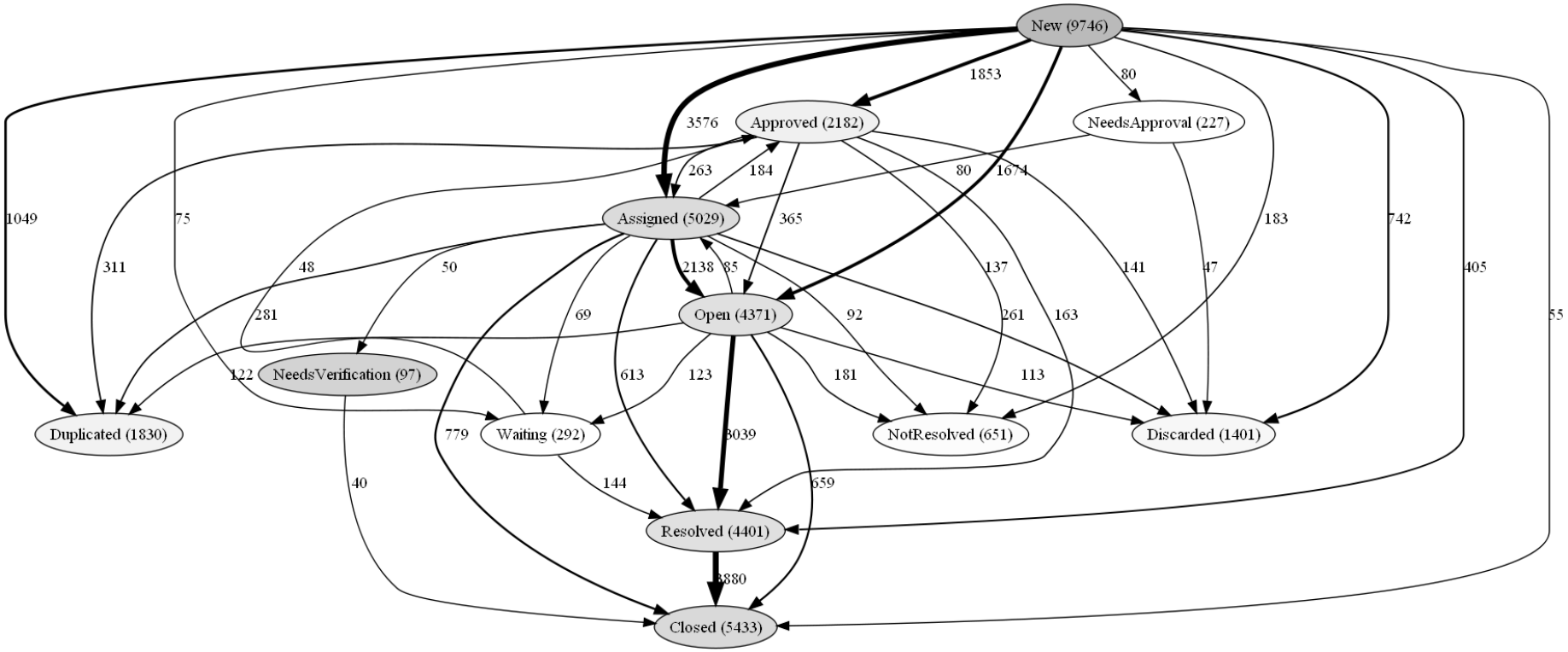
- Base de dados do sistema

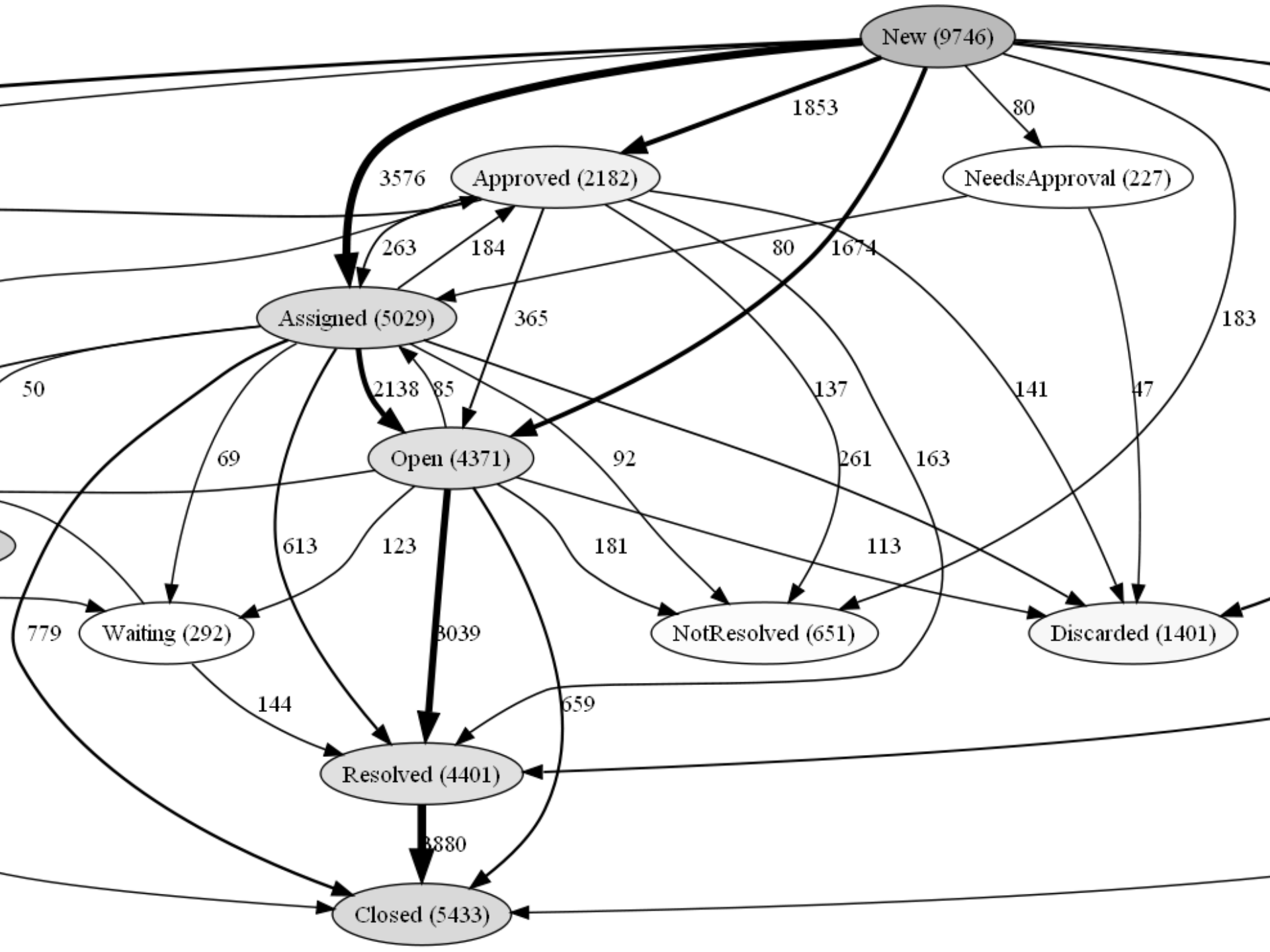




Caso prático I

- Mineração do fluxo

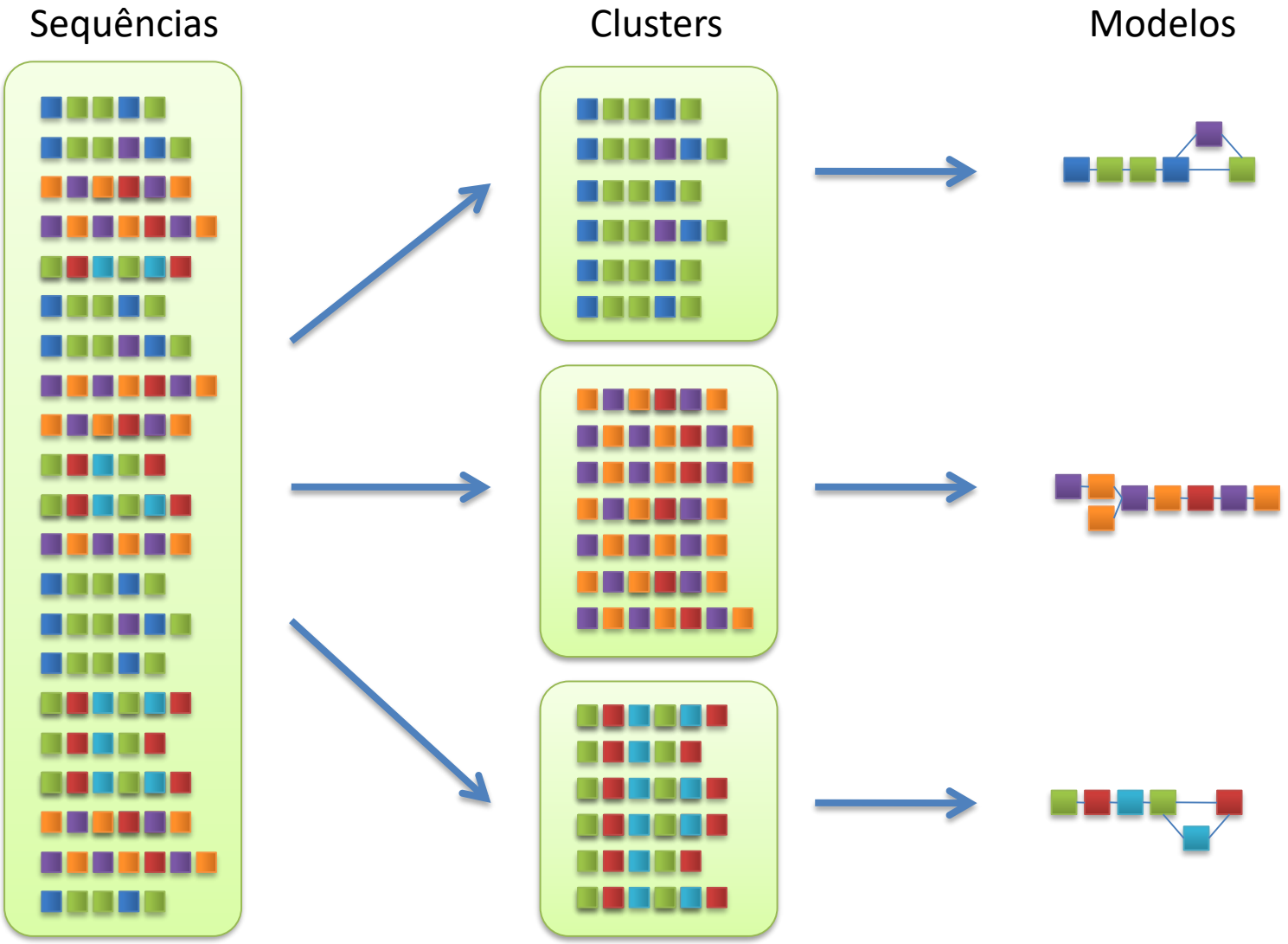



New (9746)
1853
80
3576
Approved (2182)
NeedsApproval (227)
263
184
80
1674
Assigned (5029)
365
183
50
2138
85
137
141
47
69
Open (4371)
92
261
163
613
123
181
113
779
Waiting (292)
3039
NotResolved (651)
Discarded (1401)
144
659
Resolved (4401)
880
Closed (5433)

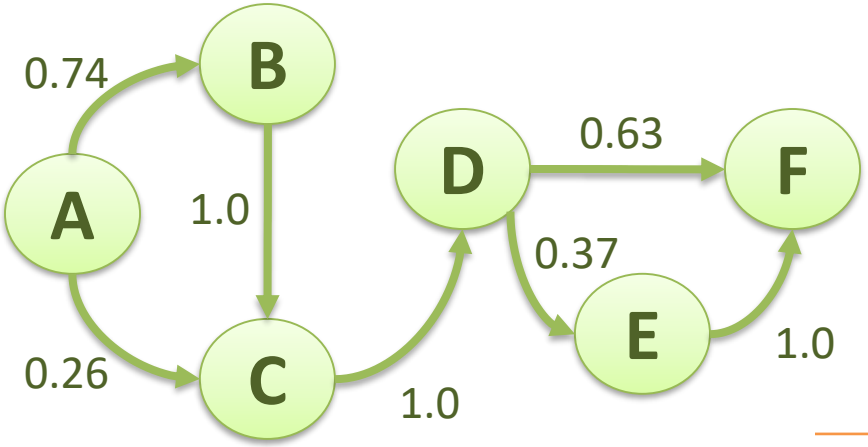
Mineração de processos

- Técnicas avançadas
  - clustering de sequências
  - clustering da rede social

Clustering de sequências



Clustering de sequências

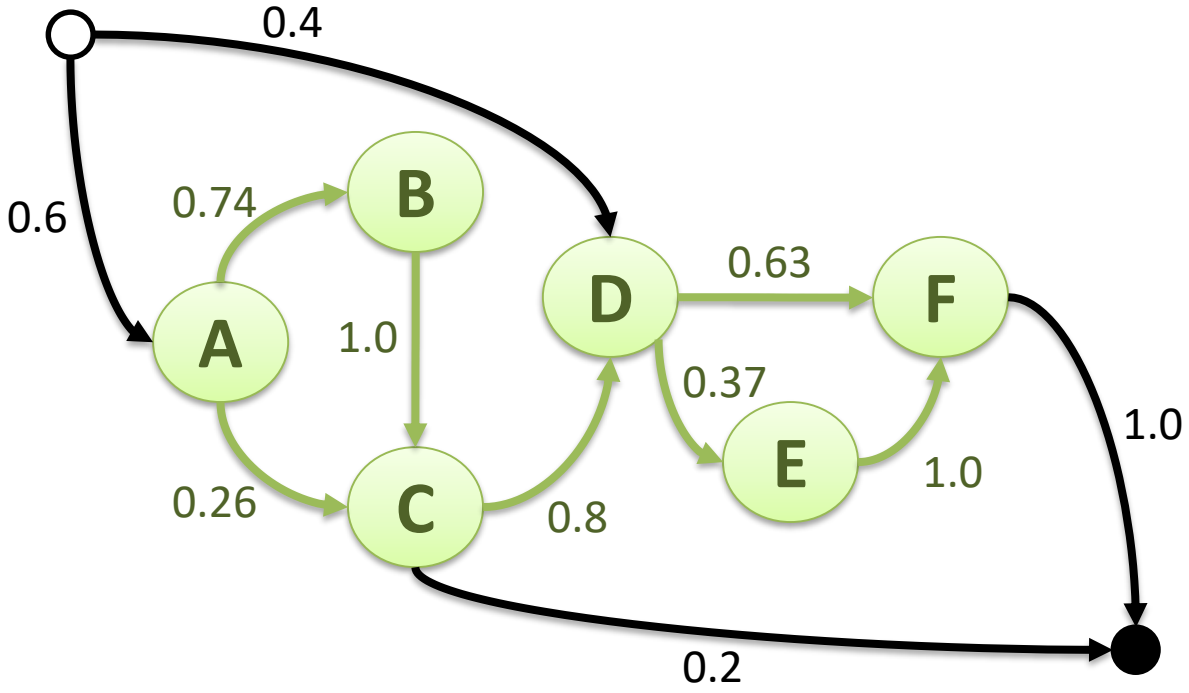


ABCDEF
ACDE
BCDF
CDF
DE
…

	A	B	C	D	E	F
A	-	0.74	0.26	-	-	-
B	-	-	1.0	-	-	-
C	-	-	-	1.0	-	-
D	-	-	-	-	0.37	0.63
E	-	-	-	-	-	1.0
F	-	-	-	-	-	-

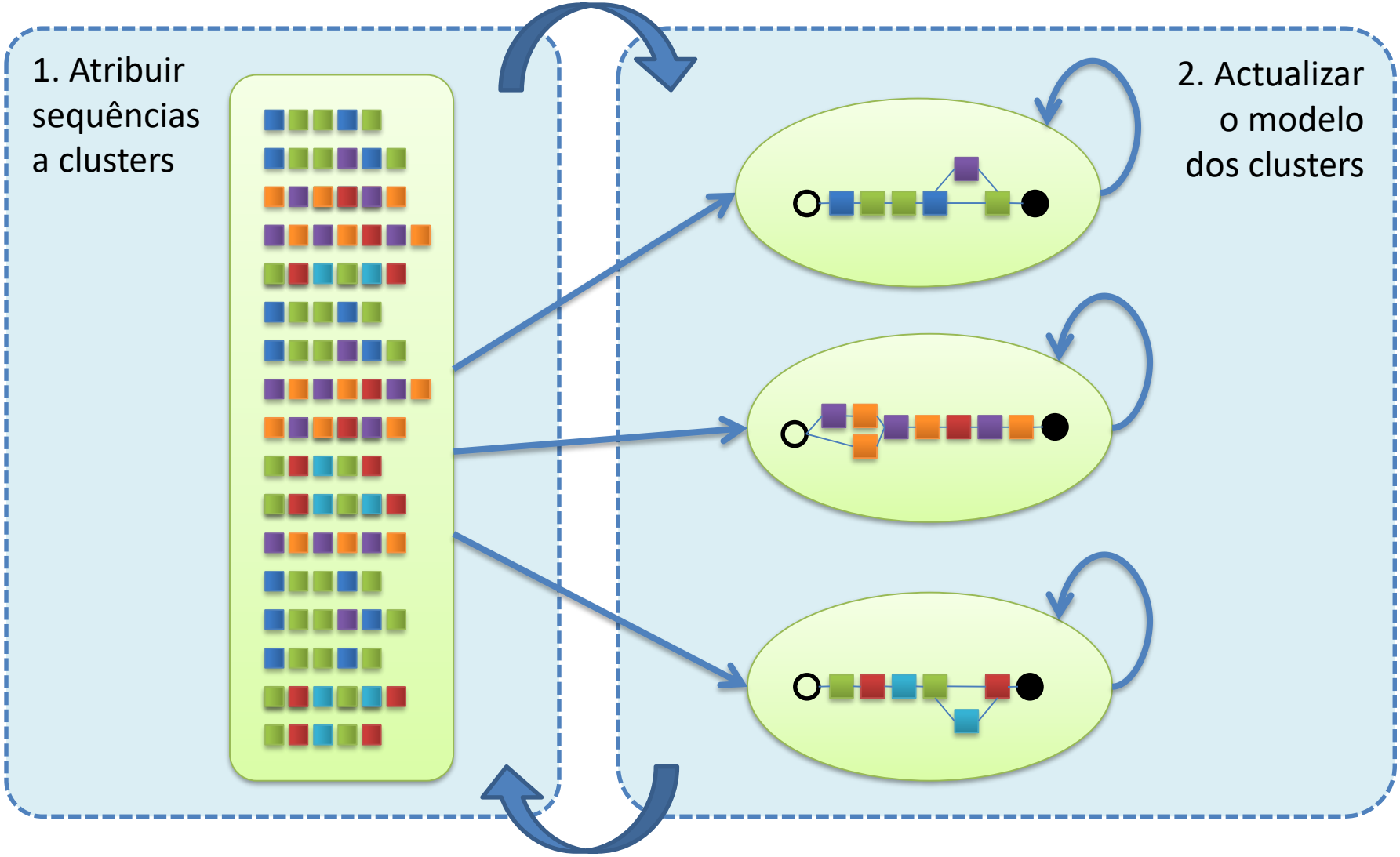
Clustering de sequências

- estados de entrada (○) e de saída (●)



○AC●
○ABC●
○ABC●
○ACDF●
○ACDEF●
○ABCDF●
○ABCDEF●
○DF●
○DEF●

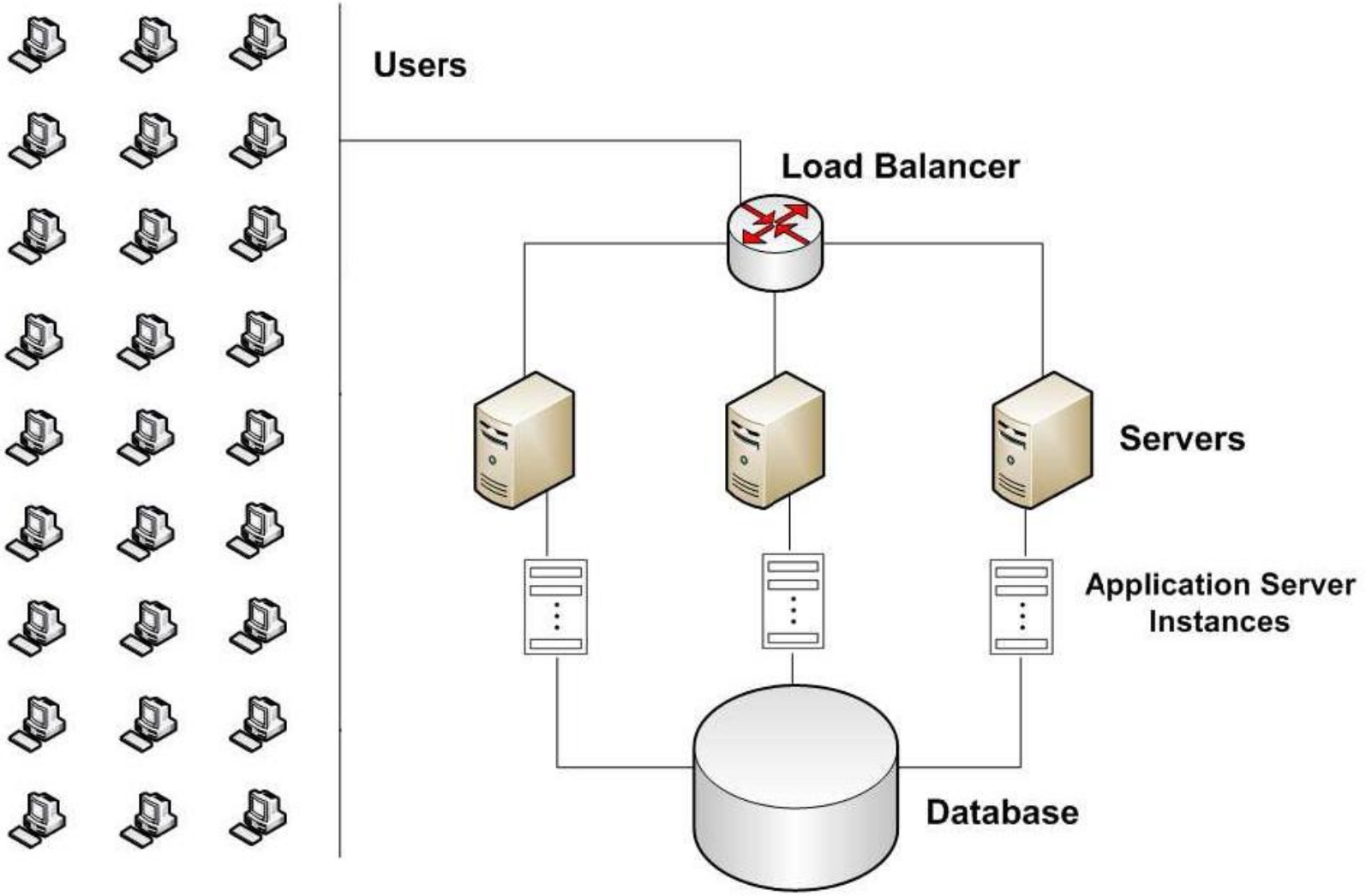
Clustering de sequências



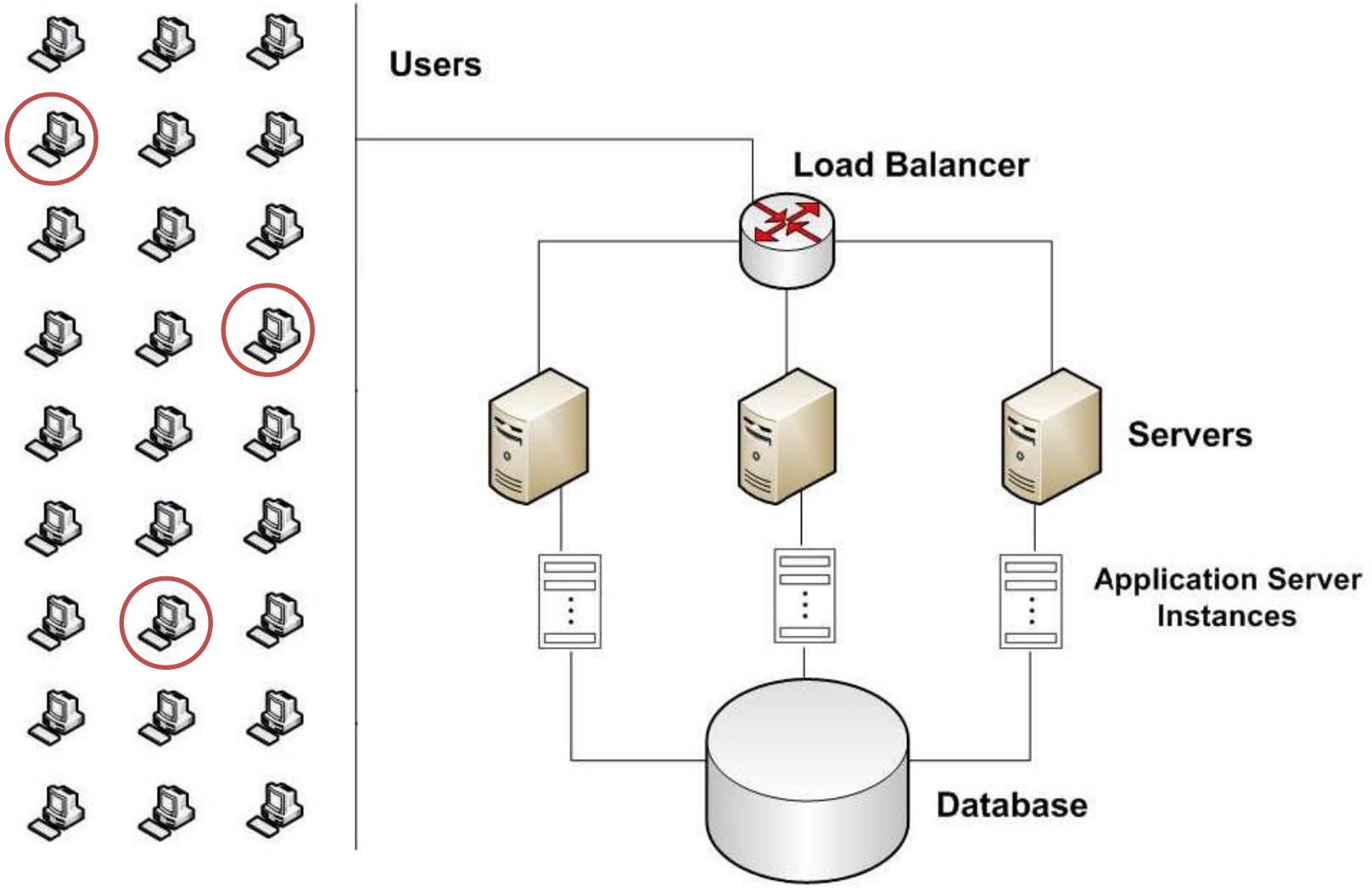
Caso prático II

- Administração pública
  - sistema TI de grande escala
  - replicação, redundância, balanceamento de carga
  - grande número de utilizadores em simultâneo
- Infra-estrutura
  - arquitectura cliente-servidor baseada em Java
  - aplicação Java nos clientes
  - servidor aplicacional de *Enterprise JavaBeans*

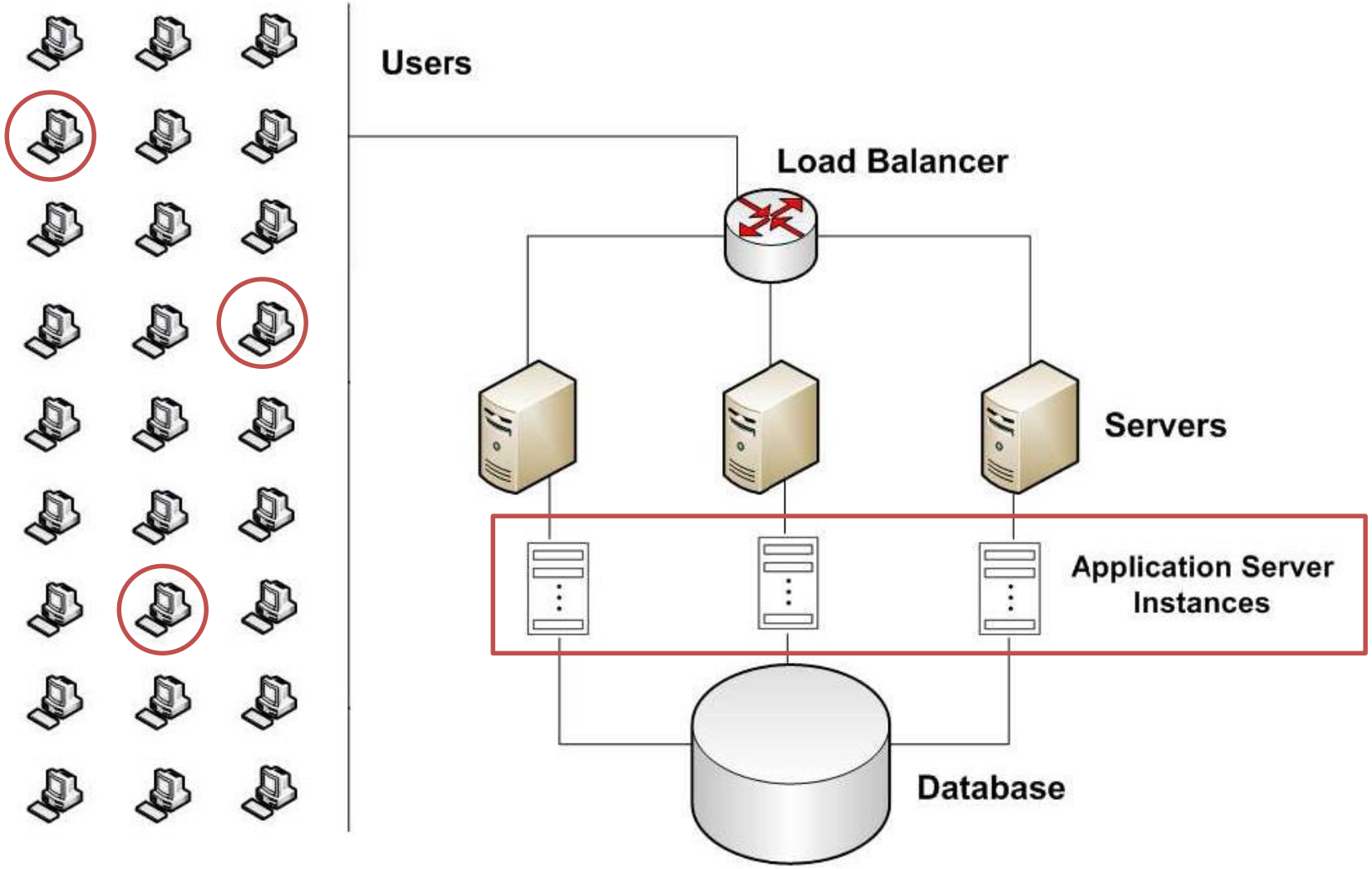
Caso prático II



Caso prático II



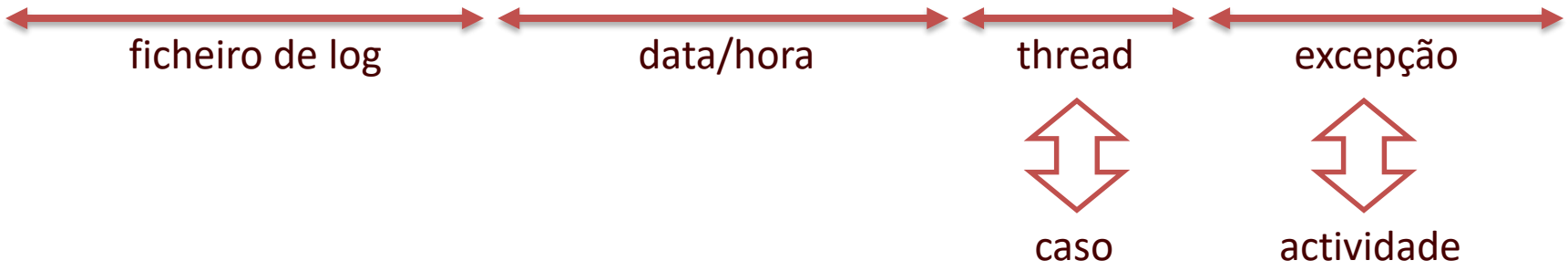
Caso prático II



Caso prático II

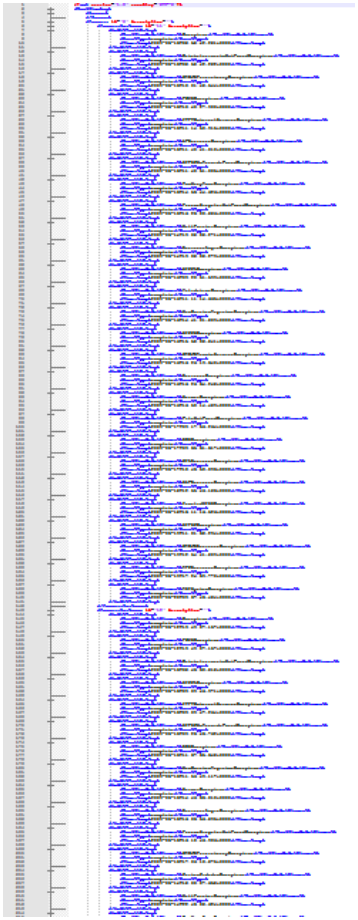
- Excerto do log

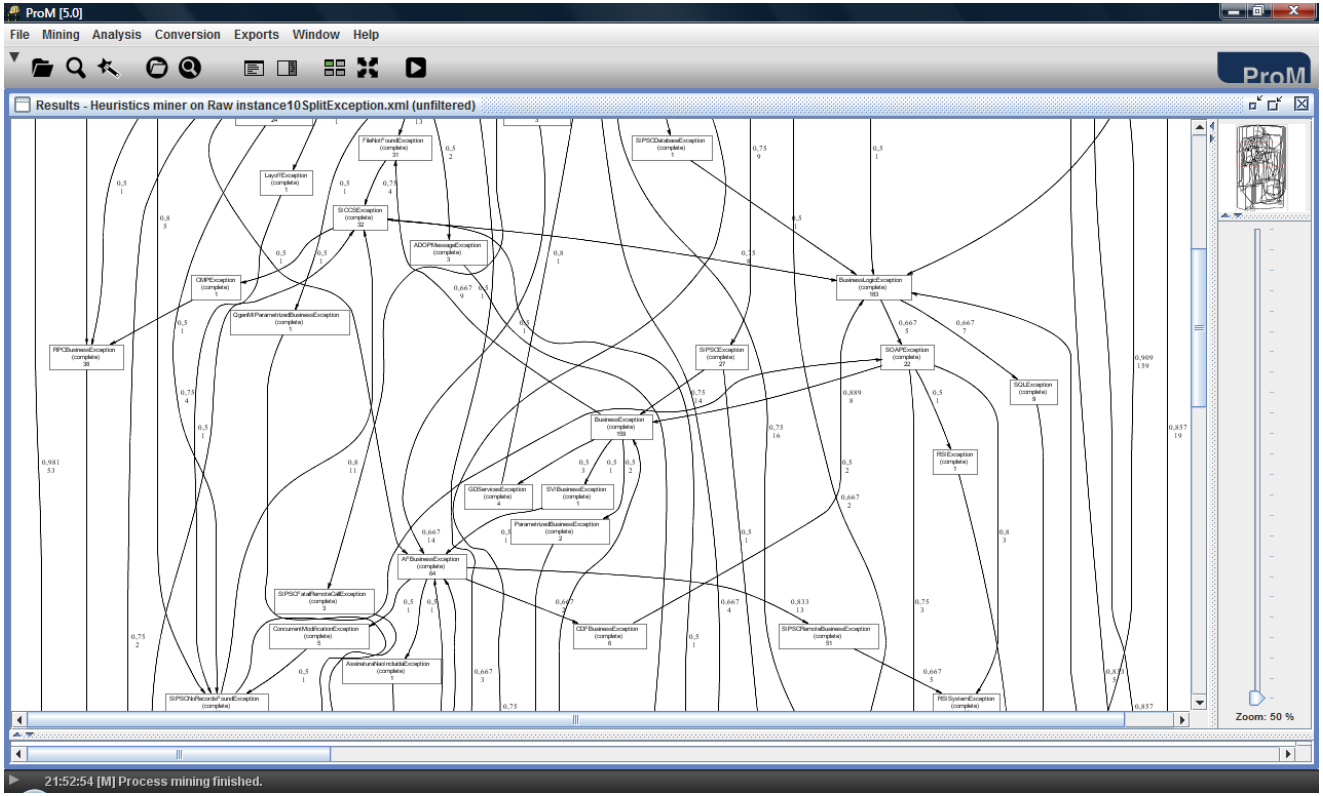
```
server.log_2009-03-13T09-40-29 2009-03-12T19:12:54.135+0000 _ThreadID=23; WSException
server.log_2009-03-13T09-40-29 2009-03-12T19:13:18.145+0000 _ThreadID=23; BusinessException
server.log_2009-03-13T09-40-29 2009-03-12T19:13:18.145+0000 _ThreadID=173; SystemException
server.log_2009-03-13T09-40-29 2009-03-12T19:14:20.189+0000 _ThreadID=21; BusinessException
server.log_2009-03-13T09-40-29 2009-03-12T19:14:20.189+0000 _ThreadID=21; EmptyResultException
server.log_2009-03-13T09-40-29 2009-03-12T19:18:34.128+0000 _ThreadID=17; EmptyResultException
server.log_2009-03-13T09-40-29 2009-03-12T19:18:34.128+0000 _ThreadID=17; NoRecordsFoundException
server.log_2009-03-13T09-40-29 2009-03-12T19:19:00.785+0000 _ThreadID=155; Exception
server.log_2009-03-13T09-40-29 2009-03-12T19:19:00.785+0000 _ThreadID=155; NoRecordsFoundException
server.log_2009-03-13T09-40-29 2009-03-12T19:20:20.339+0000 _ThreadID=410; AFException
server.log_2009-03-13T09-40-29 2009-03-12T19:21:09.291+0000 _ThreadID=172; ExtServiceException
server.log_2009-03-13T09-40-29 2009-03-12T19:21:09.292+0000 _ThreadID=172; BusinessLogicException
```



Caso prático II

- Análise sem clustering de sequências

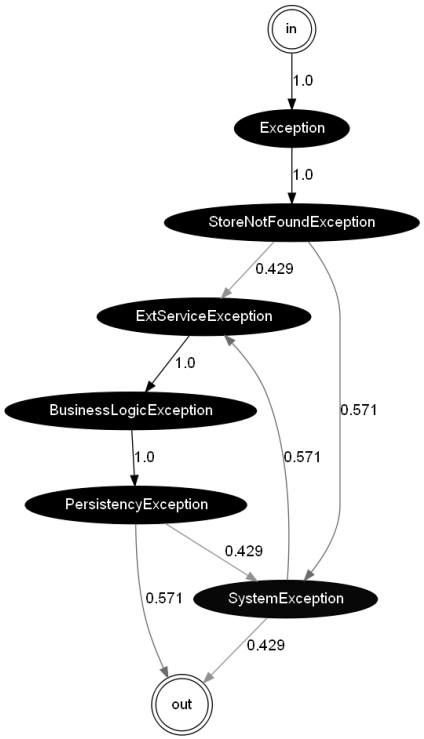




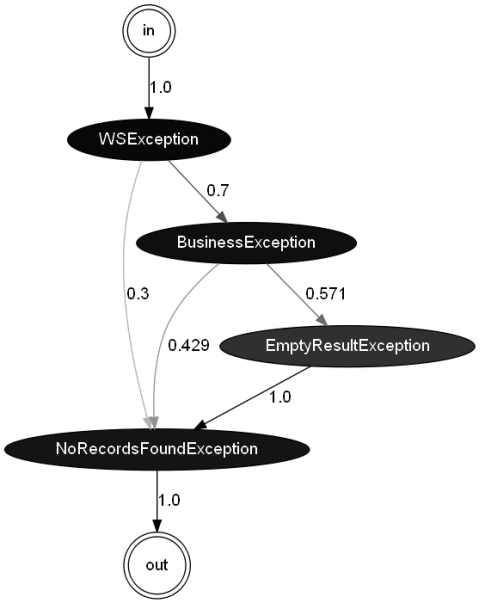
Caso prático II

- Análise com clustering de sequências

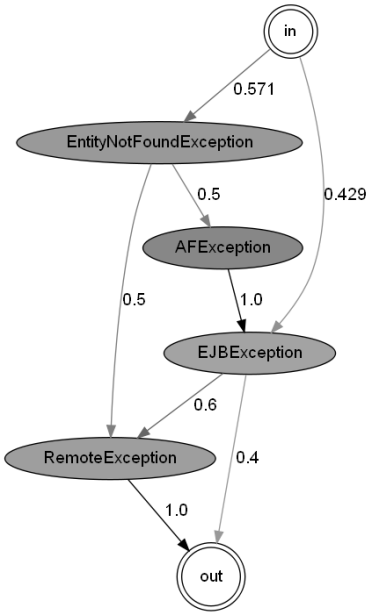
Cluster 1



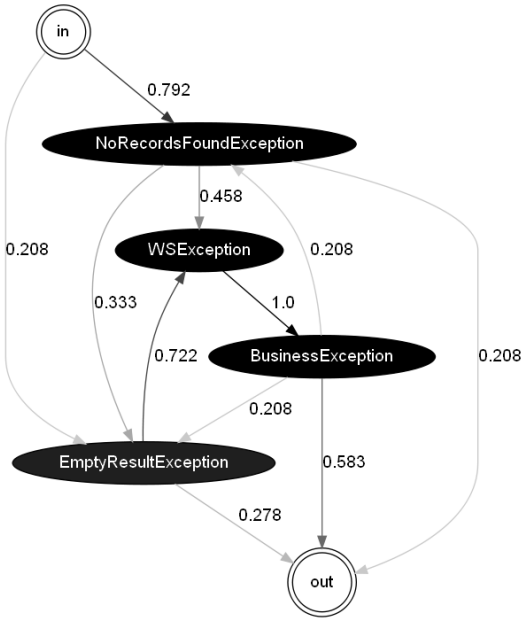
Cluster 2



Cluster 3

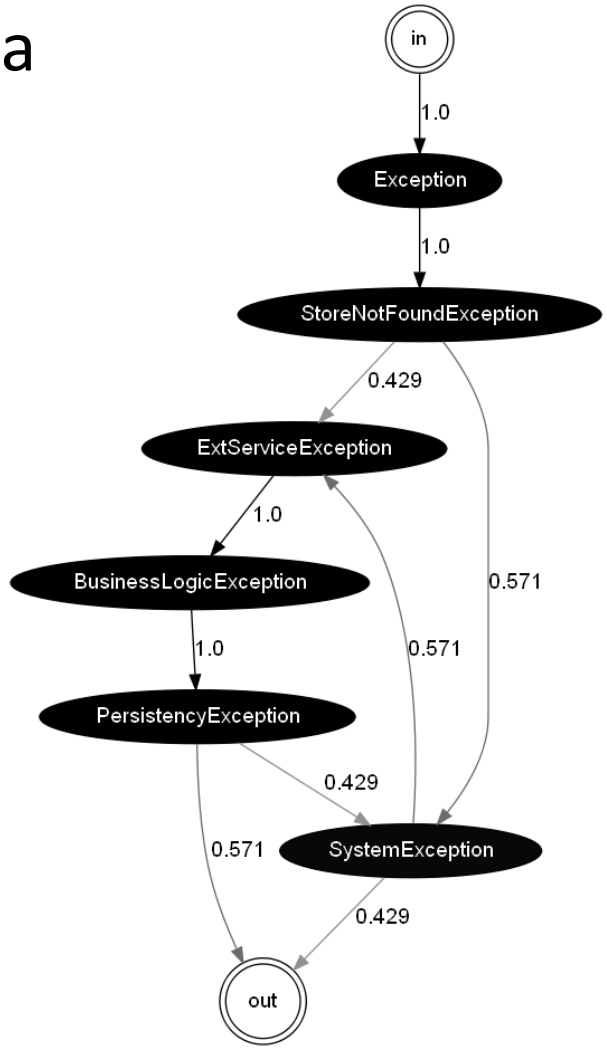


Cluster 4



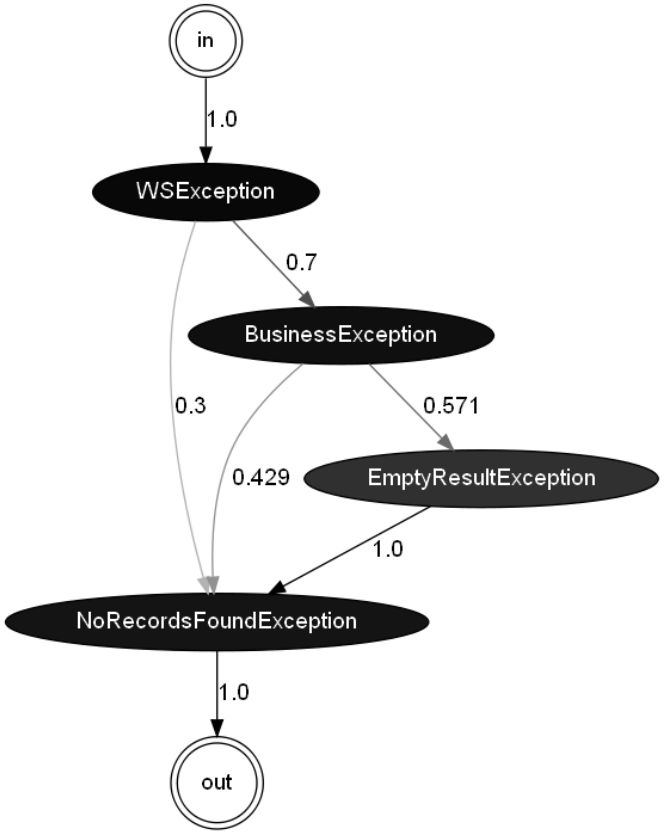
Caso prático II

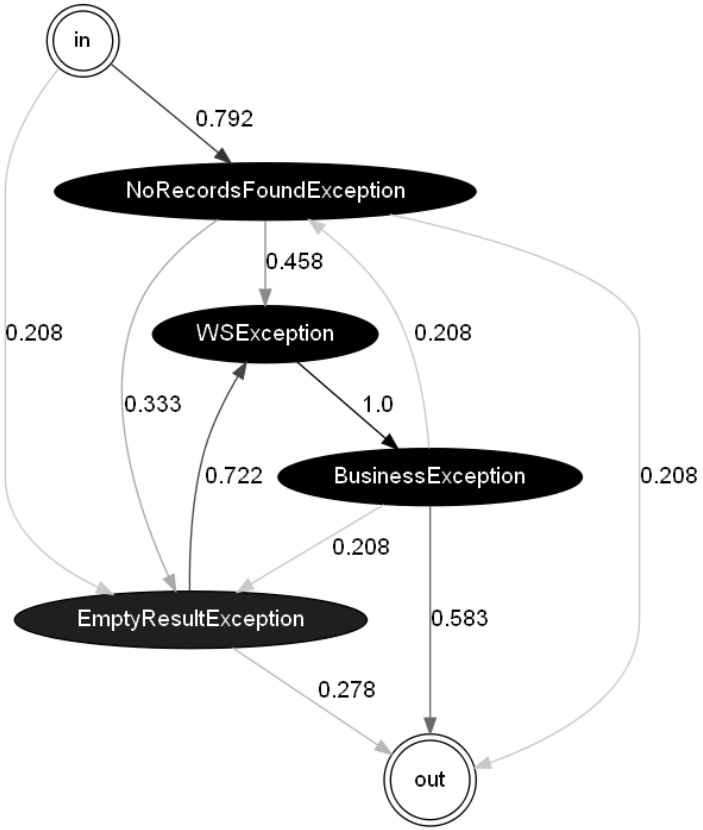
- Cluster 1: item de pesquisa não encontrado



Caso prático II

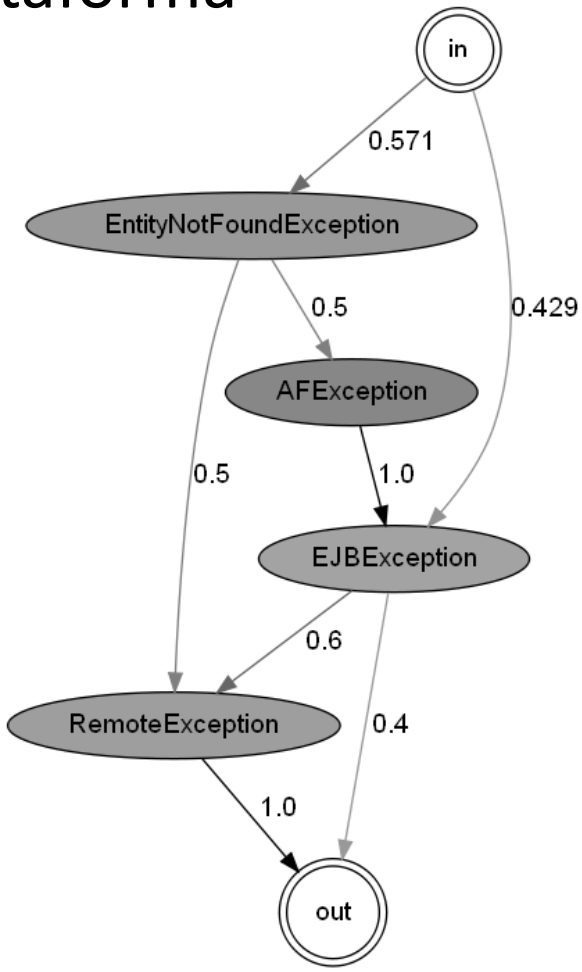
- Clusters 2 e 4: excepções de web services e consultas à base de dados com resultado vazio



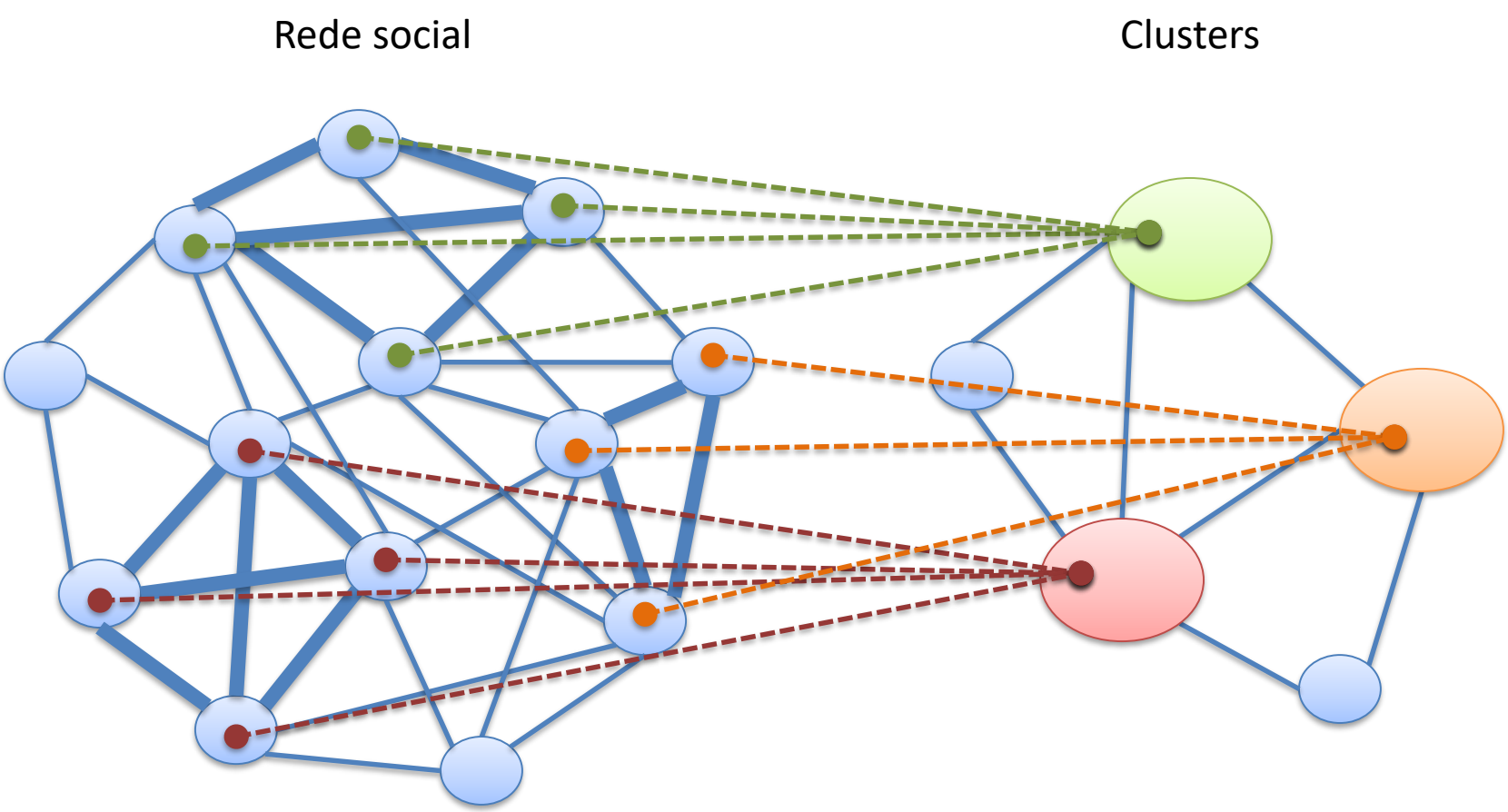


Caso prático II

- Clusters 3: excepções da plataforma em picos de carga

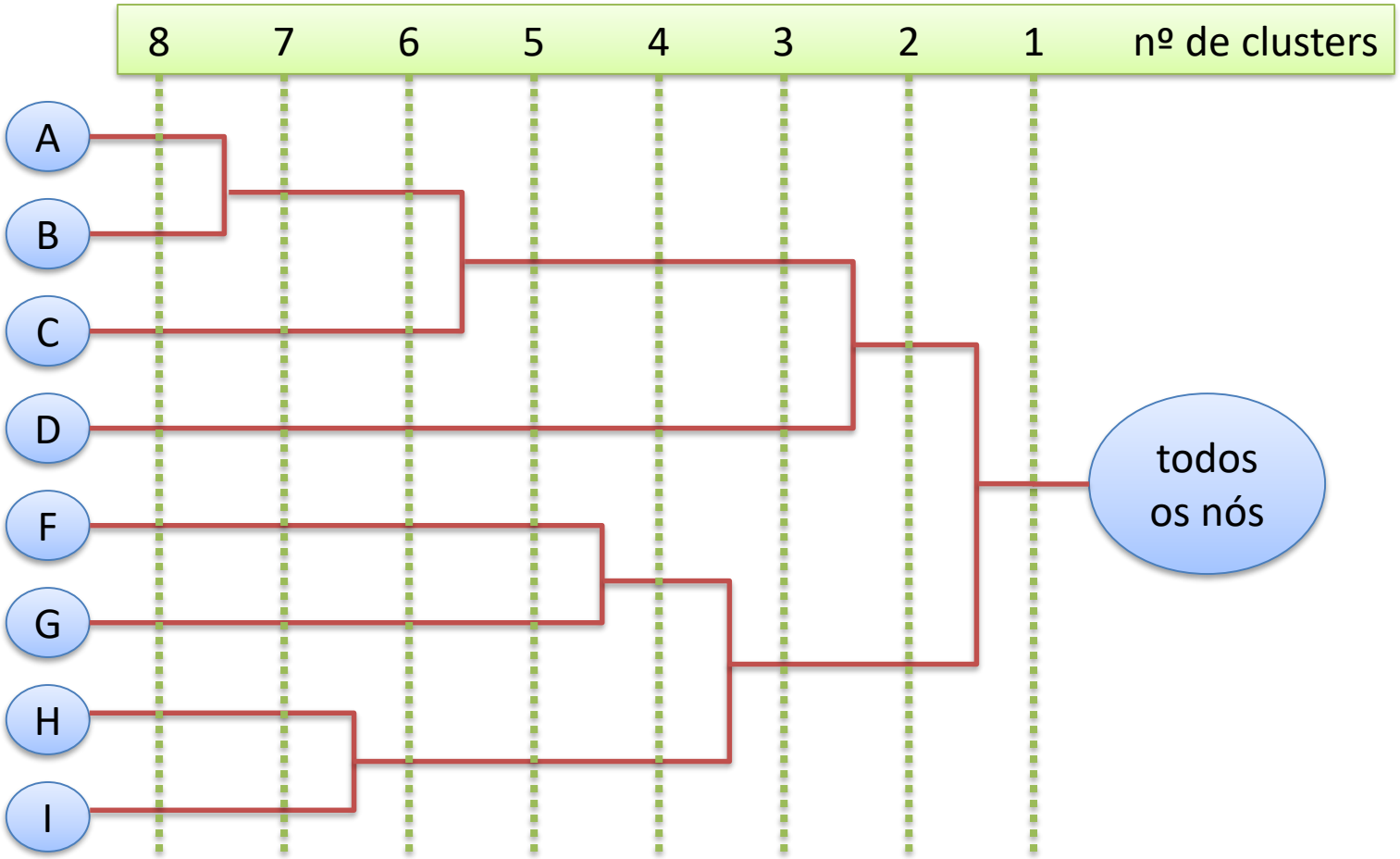


Clustering da rede social



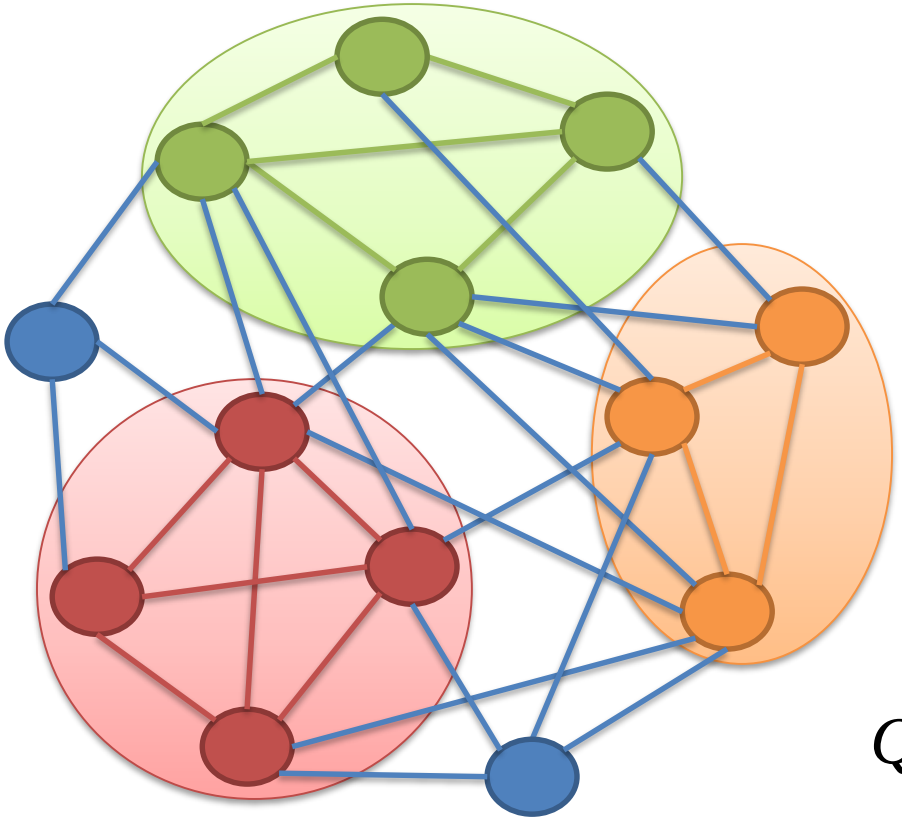
Clustering da rede social

- Clustering hierárquico



Clustering da rede social

- Modularidade



A_{ij} = matriz de adjacências

k_i = grau do nó *i*

m = soma das ligações da rede

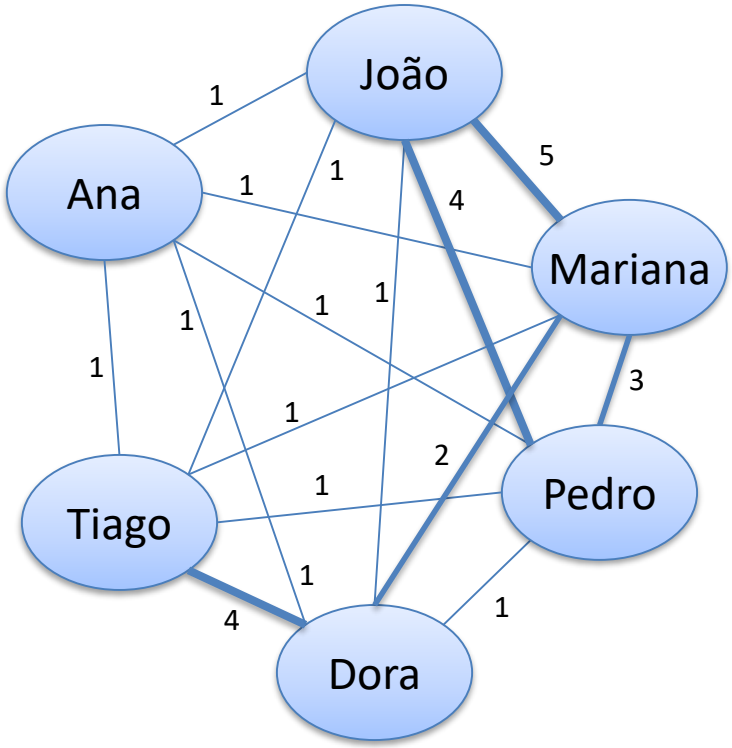
c_i = cluster do nó *i*

Modularidade Q

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \cdot \delta(c_i, c_j)$$

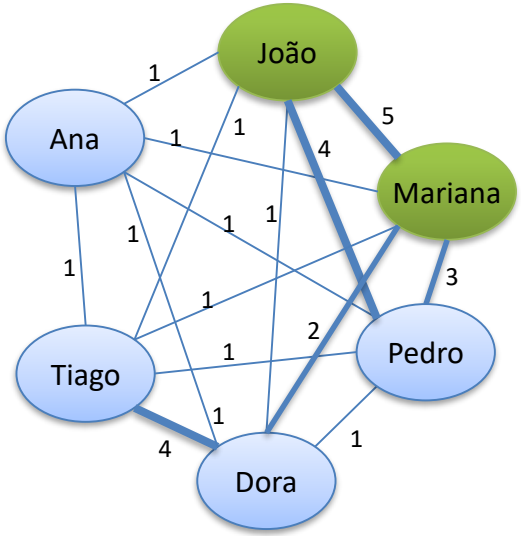
Clustering da rede social

- Exemplo



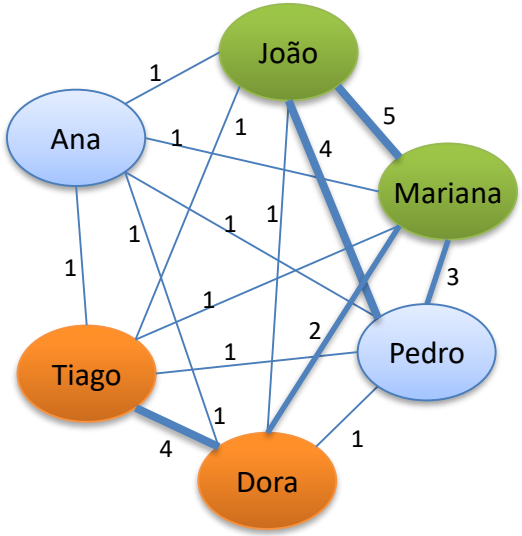
A_{ij}	João	Mariana	Pedro	Dora	Tiago	Ana
João	–	5	4	1	1	1
Mariana	5	–	3	2	1	1
Pedro	4	3	–	1	1	1
Dora	1	2	1	–	4	1
Tiago	1	1	1	4	–	1
Ana	1	1	1	1	1	–

Clustering da rede social



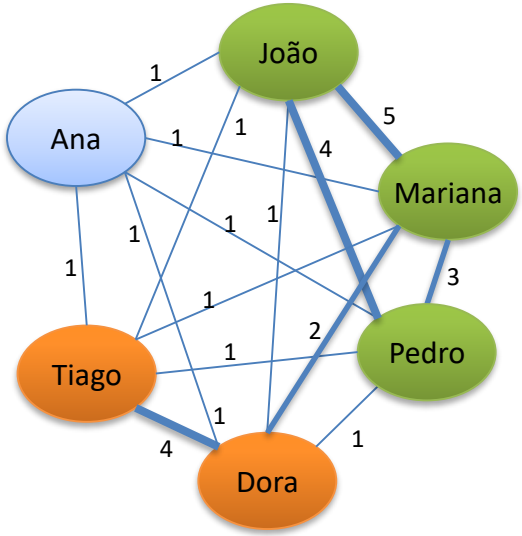
A_{ij}	J	M	P	D	T	A
J	–	5	4	1	1	1
M	5	–	3	2	1	1
P	4	3	–	1	1	1
D	1	2	1	–	4	1
T	1	1	1	4	–	1
A	1	1	1	1	1	–

D = 5.0 Q = 0.087



A_{ij}	J	M	P	D	T	A
J	–	5	4	1	1	1
M	5	–	3	2	1	1
P	4	3	–	1	1	1
D	1	2	1	–	4	1
T	1	1	1	4	–	1
A	1	1	1	1	1	–

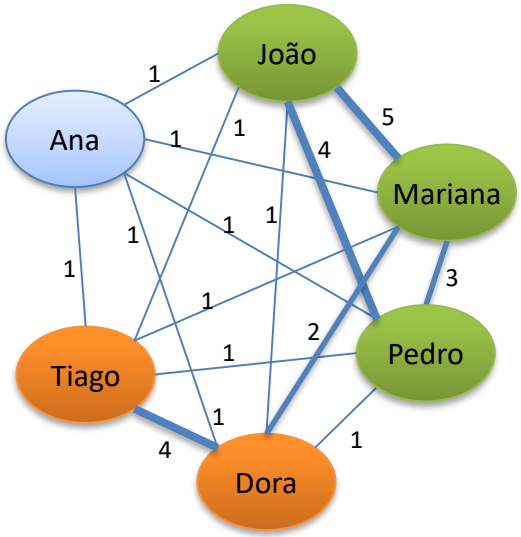
D = 4.0 Q = 0.184



A_{ij}	J	M	P	D	T	A
J	–	5	4	1	1	1
M	5	–	3	2	1	1
P	4	3	–	1	1	1
D	1	2	1	–	4	1
T	1	1	1	4	–	1
A	1	1	1	1	1	–

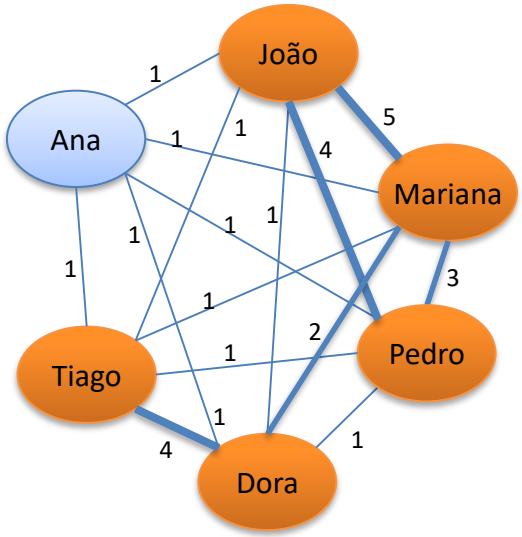
D = 3.5 Q = 0.281

Clustering da rede social



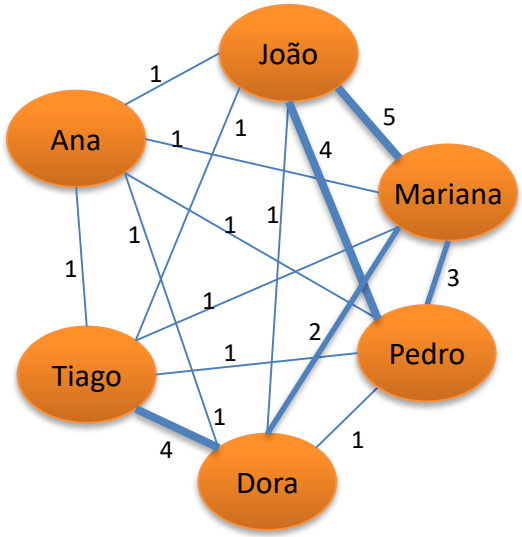
A_{ij}	J	M	P	D	T	A
J	–	5	4	1	1	1
M	5	–	3	2	1	1
P	4	3	–	1	1	1
D	1	2	1	–	4	1
T	1	1	1	4	–	1
A	1	1	1	1	1	–

$D = 3.5 \quad Q = 0.281$



A_{ij}	J	M	P	D	T	A
J	–	5	4	1	1	1
M	5	–	3	2	1	1
P	4	3	–	1	1	1
D	1	2	1	–	4	1
T	1	1	1	4	–	1
A	1	1	1	1	1	–

$D = 1.167 \quad Q = 0.162$

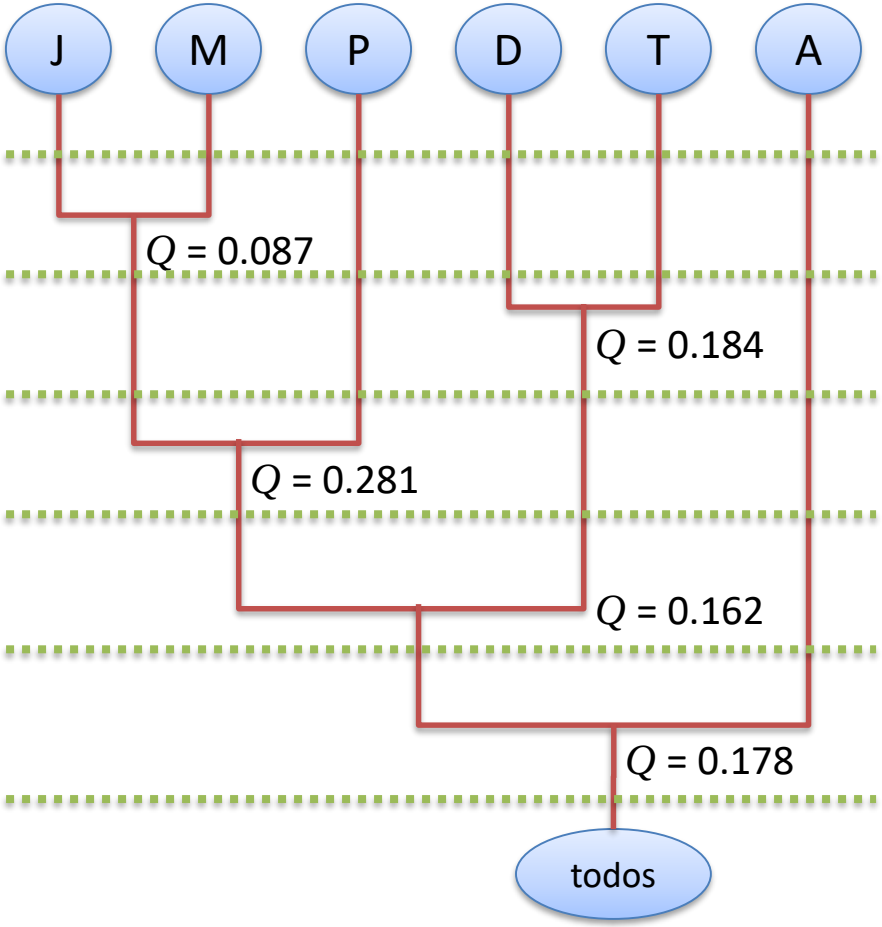


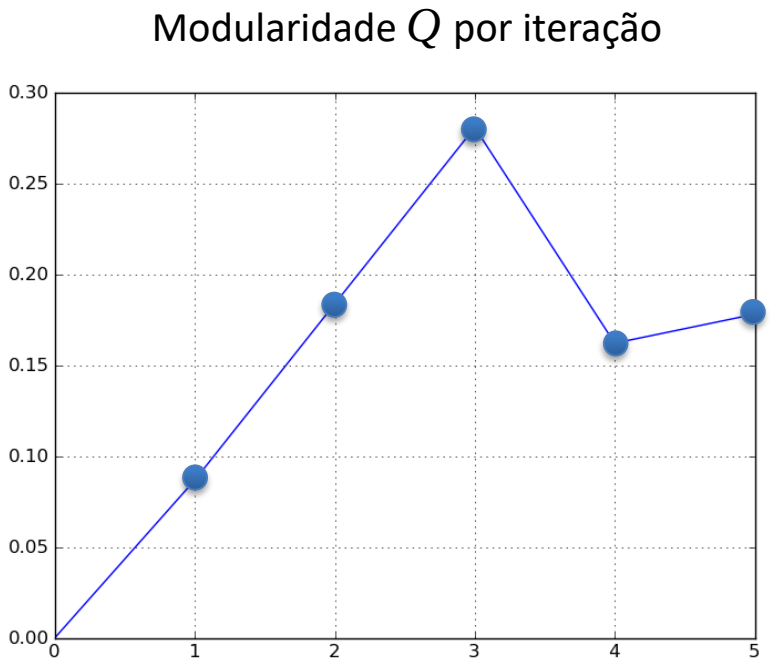
A_{ij}	J	M	P	D	T	A
J	–	5	4	1	1	1
M	5	–	3	2	1	1
P	4	3	–	1	1	1
D	1	2	1	–	4	1
T	1	1	1	4	–	1
A	1	1	1	1	1	–

$D = 1.0 \quad Q = 0.178$

Clustering da rede social

- Exemplo



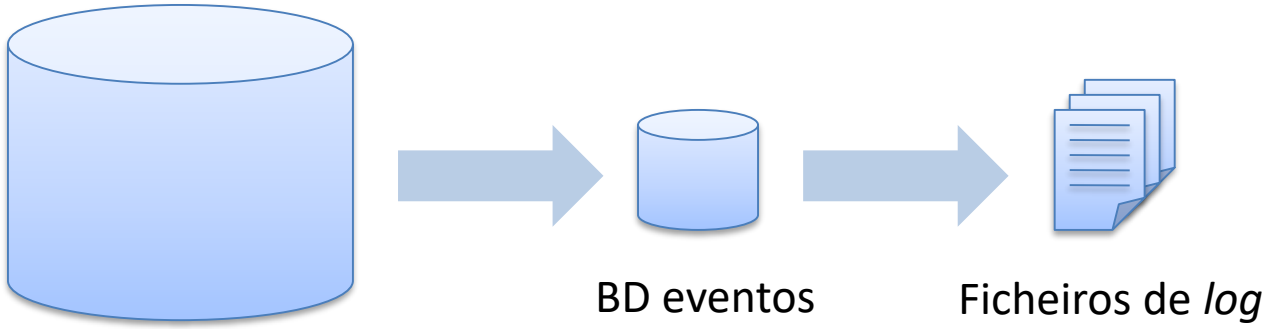


Caso prático III

- Hospital público
  - estudo dos *careflows* da urgência
    - cada paciente é um caso
    - pacientes sujeitos a triagem, consultas, exames, diagnósticos, tratamentos, etc.
    - processo pouco estruturado, objectivo era análise da rede social com a métrica *working together*
  - sistema de informação clínica desenvolvido internamente
    - base de dados com 400 tabelas e sem documentação

Caso prático III

- A partir da base de dados do sistema
  - criada uma base de dados mais pequena
  - extraídos vários ficheiros de *log*
    - desde 12 dias (14 MB)
    - até 6 meses (224 MB)



Caso prático III

- Exemplo de ficheiro MXML

```
<?xml version="1.0" encoding="utf-8"?>
<WorkflowLog>
   <Process id="Emergencia">
      <ProcessInstance id="24125085">
         <AuditTrailEntry>
            <Data>
               <Attribute name="Diagnostico">Tomografia Computorizada</Attribute>
               <Attribute name="Especialidade">Tarefeiros Emergencia</Attribute>
            </Data>
            <WorkflowModelElement>Req. Exame Imagiologia</WorkflowModelElement>
            <EventType>complete</EventType>
            <Timestamp>2009-01-09T05:35:48</Timestamp>
            <Originator>48074</Originator>
         </AuditTrailEntry>
         <AuditTrailEntry>
            ...
```

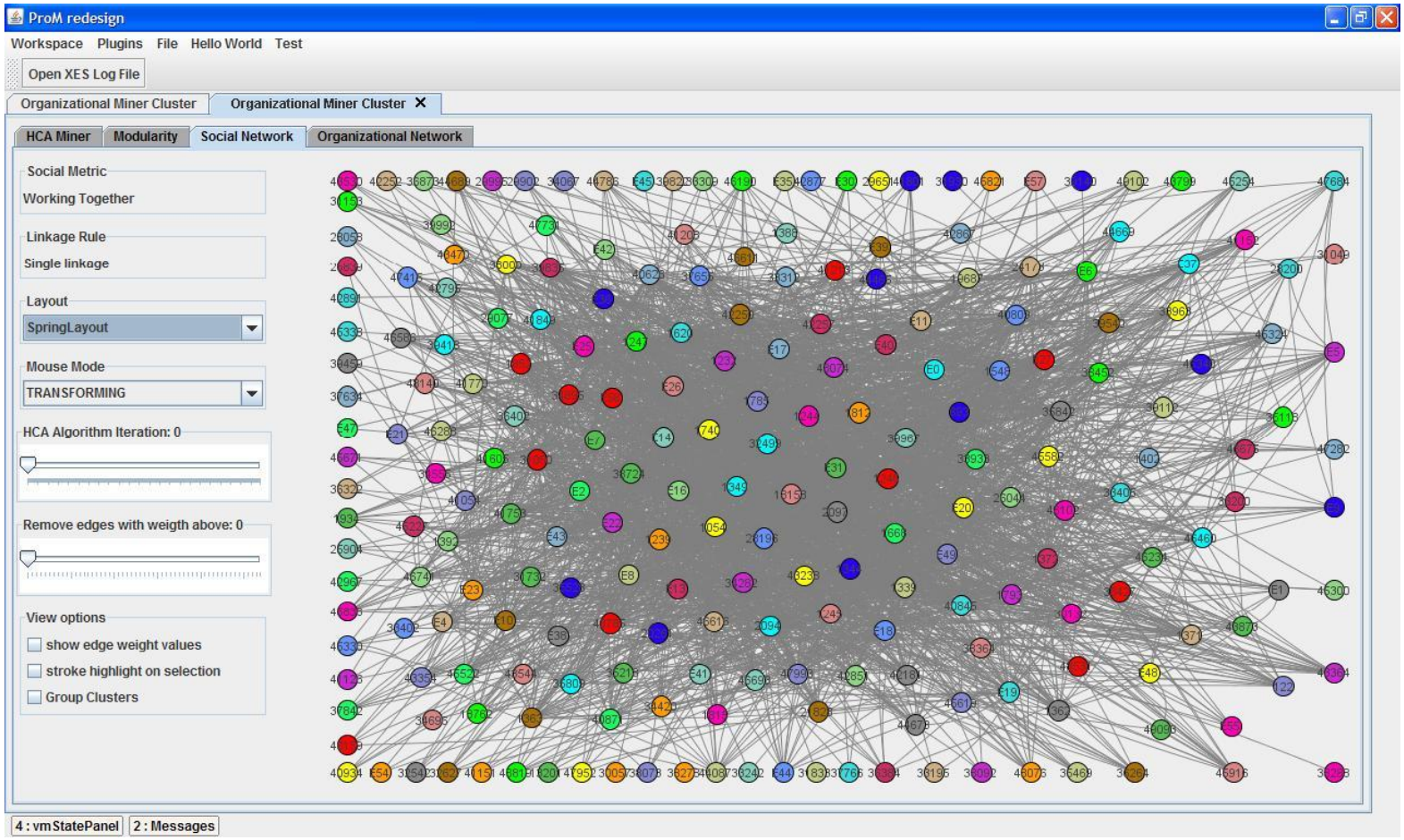
Caso prático III

- Exemplo de ficheiro MXML

```
<?xml version="1.0" encoding="utf-8"?>
<WorkflowLog>
   <Process id="Emergencia">
      <ProcessInstance id="24125085">                                    caso
         <AuditTrailEntry>
            <Data>
               <Attribute name="Diagnostico">Tomografia Computorizada</Attribute>
               <Attribute name="Especialidade">Tarefeiros Emergencia</Attribute>
            </Data>
actividade  <WorkflowModelElement>Req. Exame Imagiologia</WorkflowModelElement>
            <EventType>complete</EventType>
data/hora   <Timestamp>2009-01-09T05:35:48</Timestamp>
utilizador  <Originator>48074</Originator>
         </AuditTrailEntry>
         <AuditTrailEntry>
            ...
```

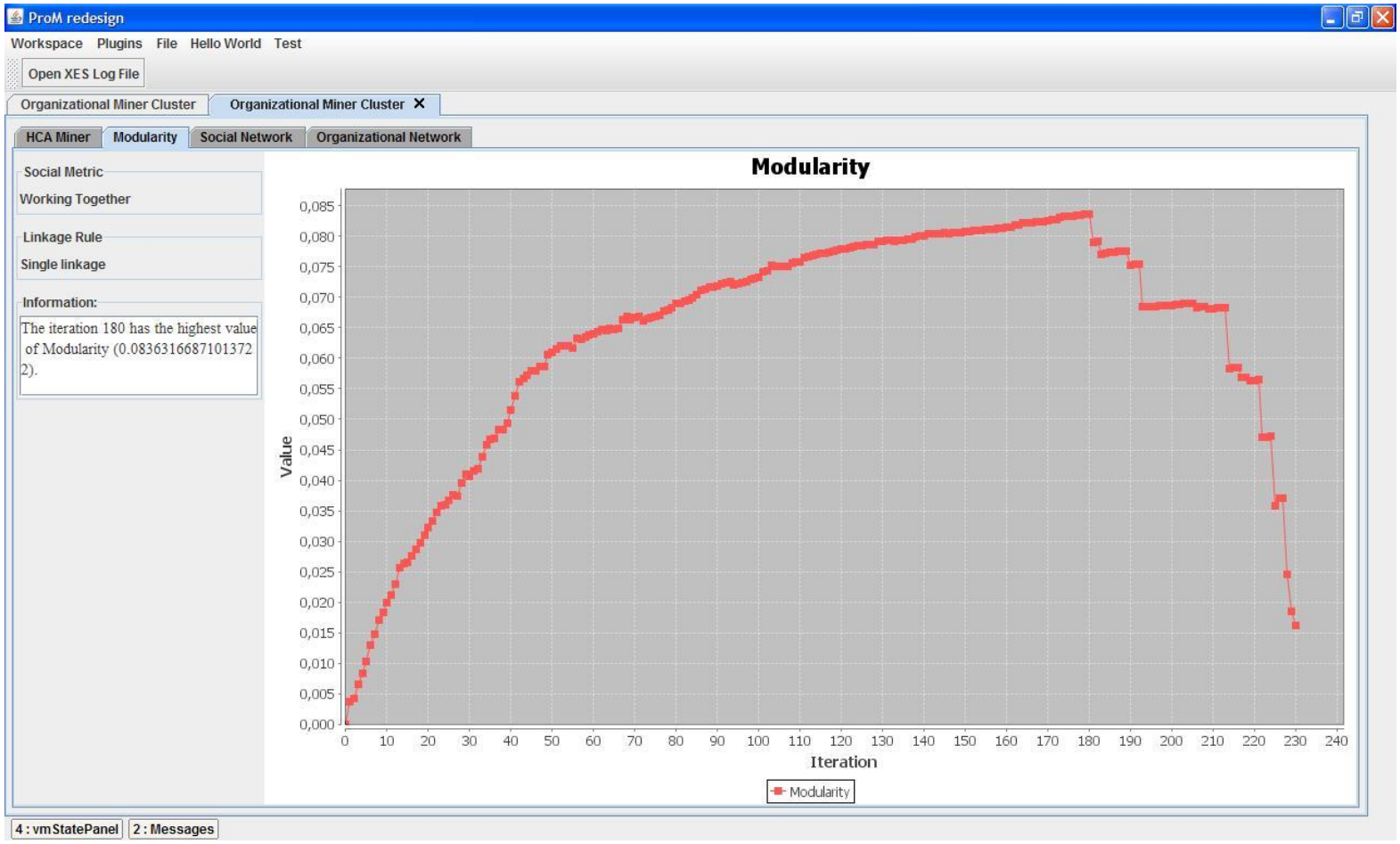
Caso prático III

- Rede social com *log* de 12 dias



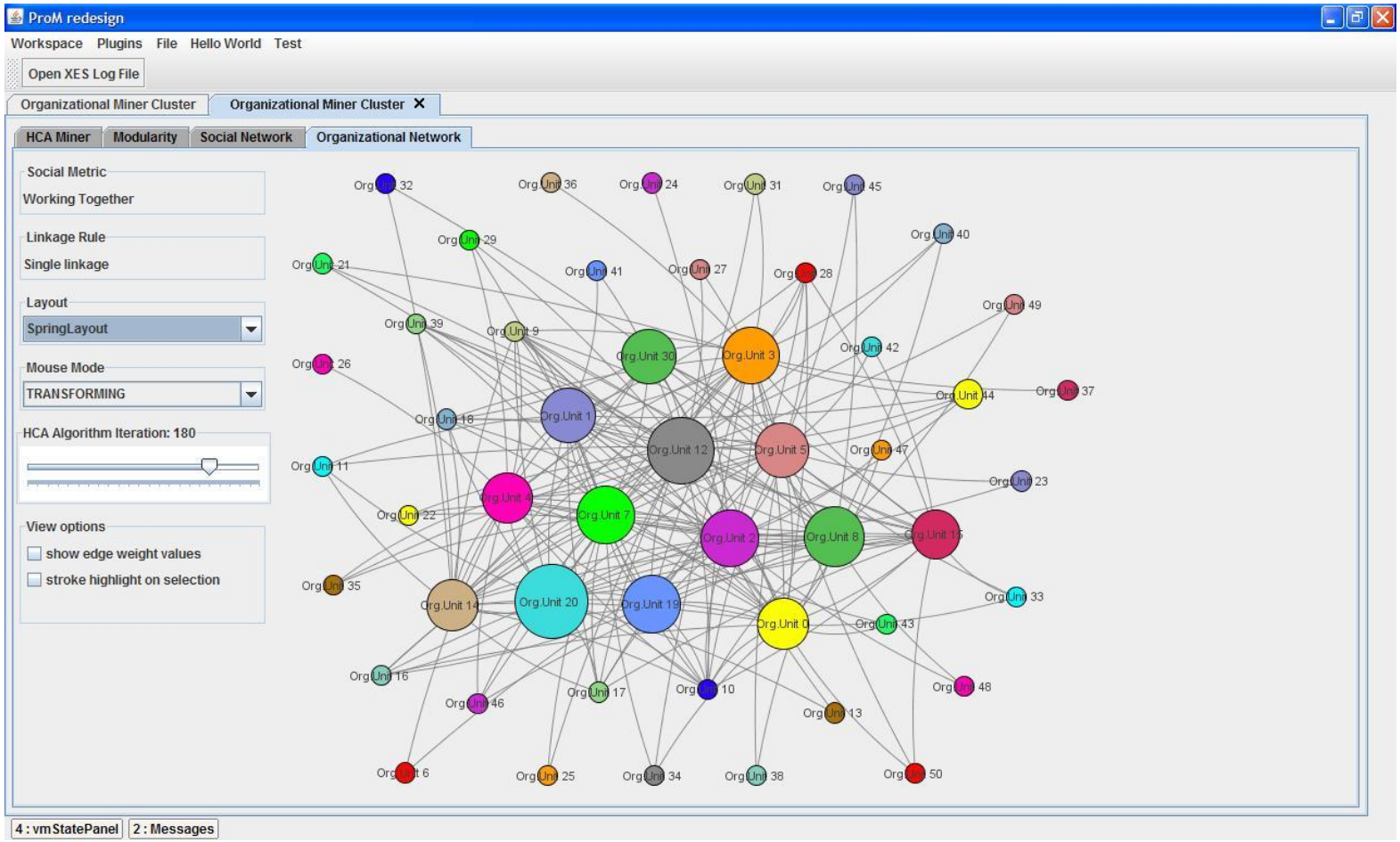
Caso prático III

- Modularidade



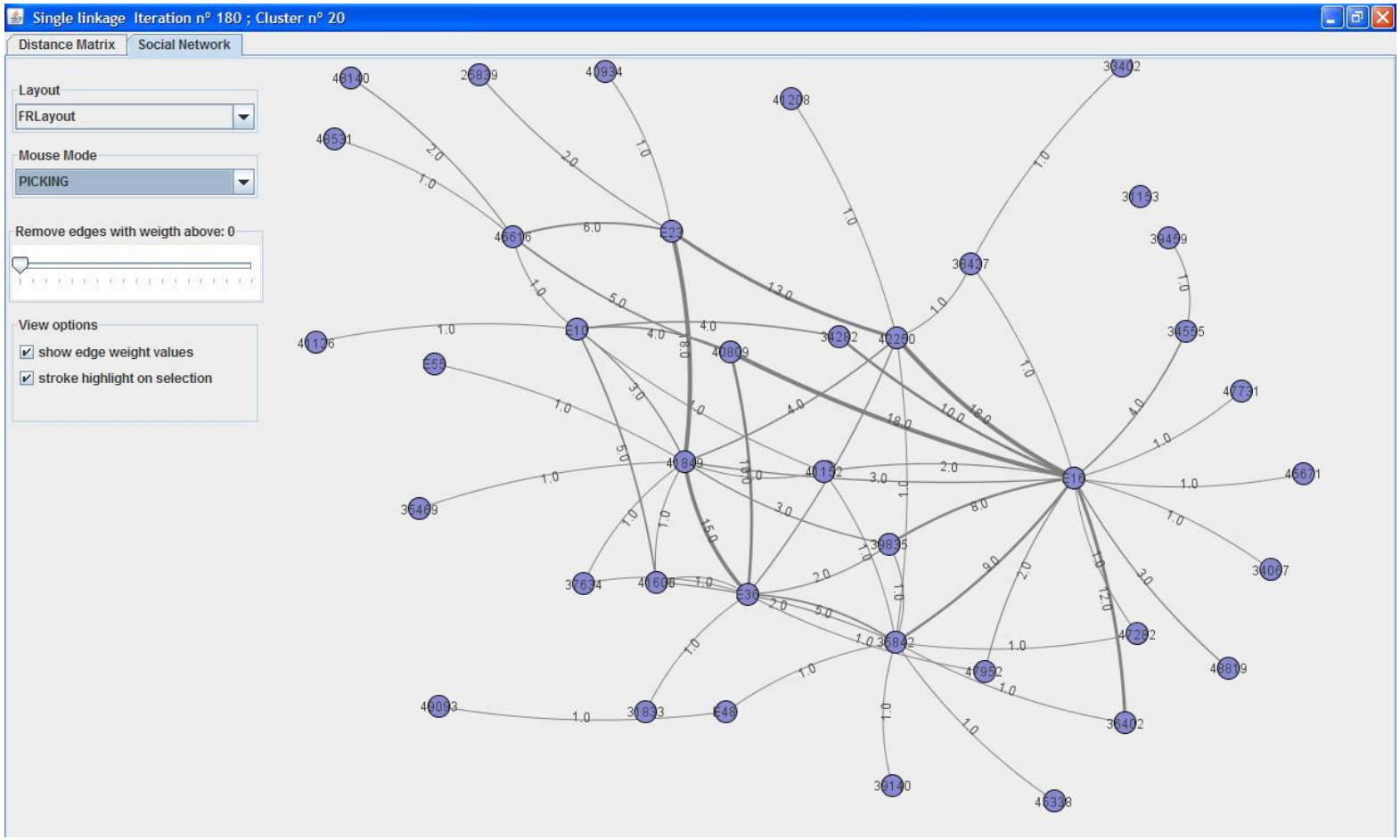
Caso prático III

- Iteração nº 180

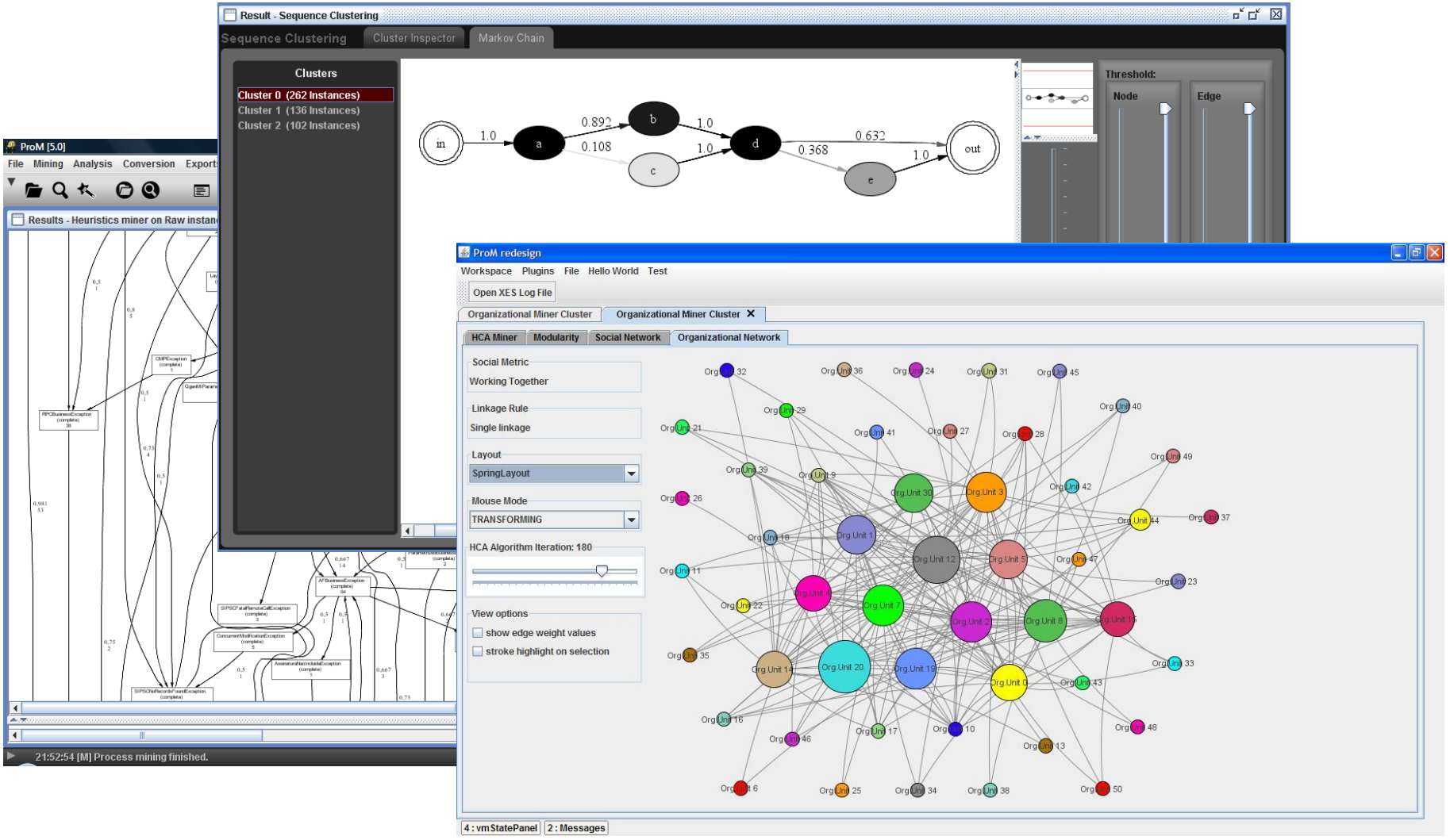


Caso prático III

- Detalhe do cluster nº 20

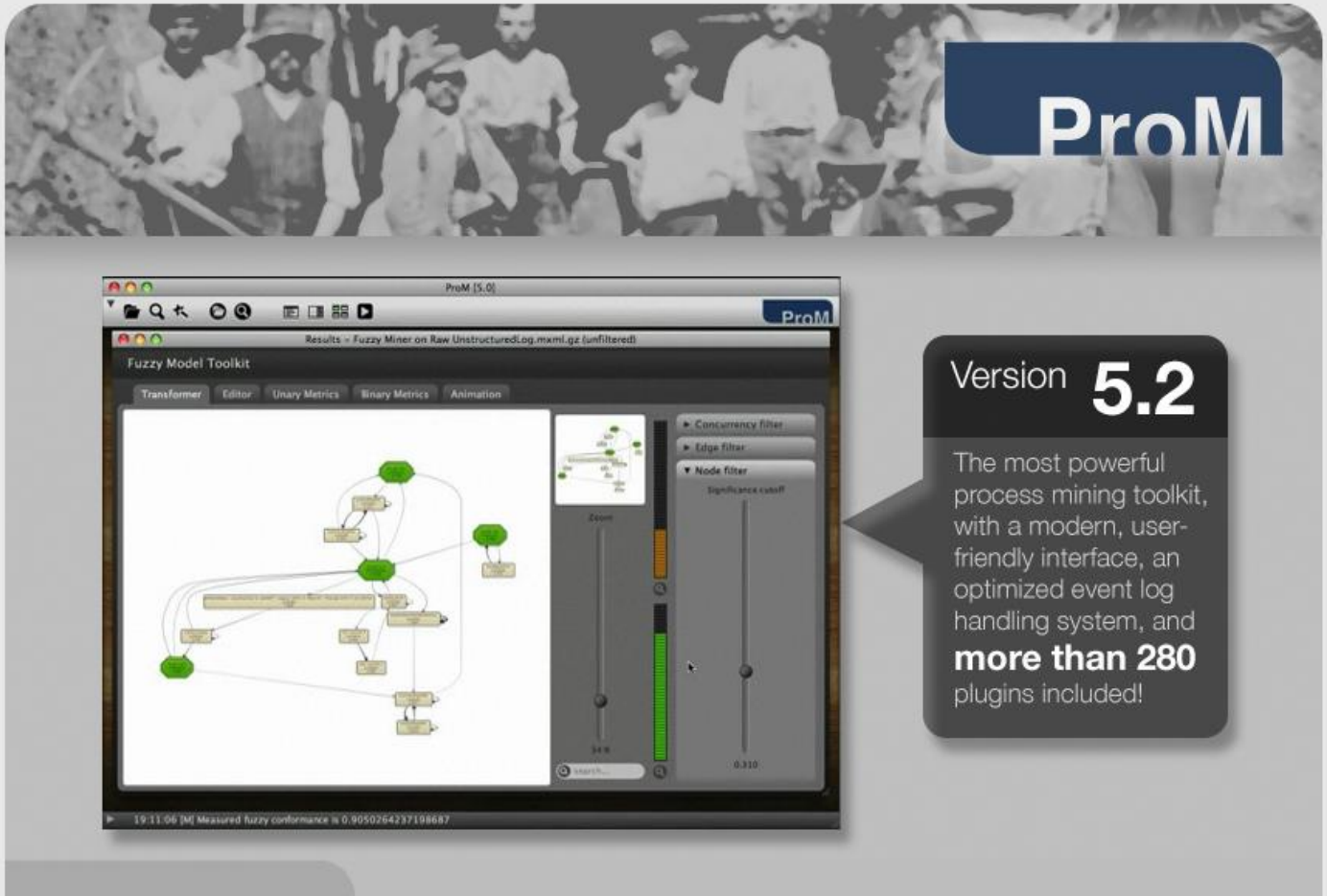


Ferramentas



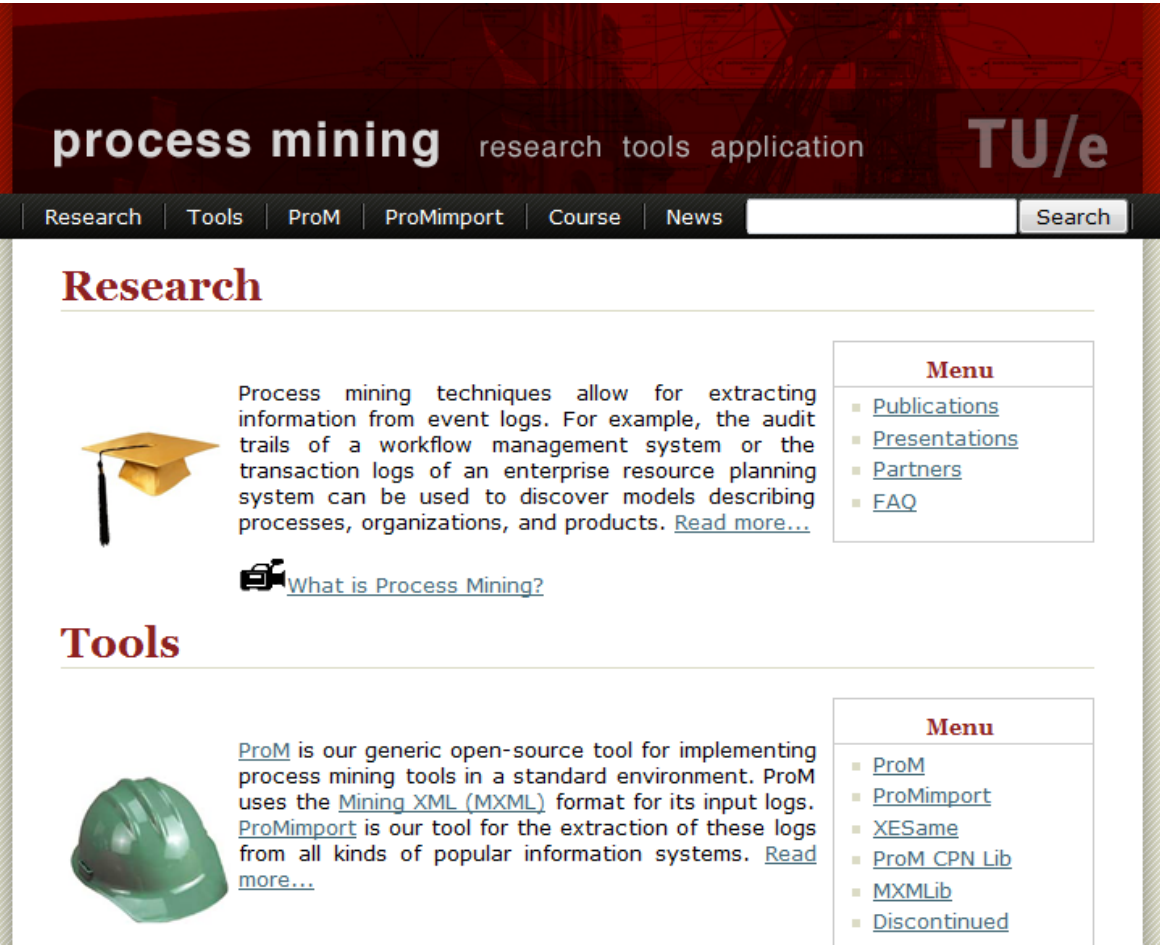
ProM

http://prom.win.tue.nl/tools/prom/ (T.U.Eindhoven)



ProcessMining.org

http://www.processmining.org/ (T.U.Eindhoven)



Algumas referências

- W.M.P. van der Aalst, A.J.M.M. Weijters, L. Maruster, ***Workflow Mining: Discovering Process Models from Event Logs***, IEEE Transactions on Knowledge and Data Engineering, 16(9):1128-1142, 2004
- W.M.P. van der Aalst, B.F. van Dongen, J. Herbst, L. Maruster, G. Schimm, A.J.M.M. Weijters, ***Workflow Mining: A Survey of Issues and Approaches***, Data and Knowledge Engineering, 47(2):237-267, 2003
- W.M.P. van der Aalst, H.A. Reijers, M. Song, ***Discovering Social Networks from Event Logs***, Computer Supported Cooperative work, 14(6):549-593, 2005
- M. Song, W.M.P. van der Aalst, ***Towards comprehensive support for organizational mining***, Decision Support Systems, 46(1):300-317, 2008
- A. Tiwari, A., C.J. Turner, B. Majeed, ***A Review of Business Process Mining: State of the Art and Future Trends***, Business Process Management Journal, 14(1):5-22, 2008
- A. Rozinat, W.M.P. van der Aalst, ***Conformance checking of processes based on monitoring real behavior***, Information Systems, 33(1):64-95, 2008
- G.M. Veiga, D.R. Ferreira, ***Understanding Spaghetti Models with Sequence Clustering for ProM***, 5th Workshop on Business Process Intelligence (BPI 09), September 2009
- D.R. Ferreira, D. Gillblad, ***Discovering Process Models from Unlabelled Event Logs***, 7th International Conference on Business Process Management (BPM 2009), Sep. 2009

Obrigado!

Questões?