CS 4072 - Topics in CS
Process Mining

Lecture # 15

April 12, 2022

Spring 2022

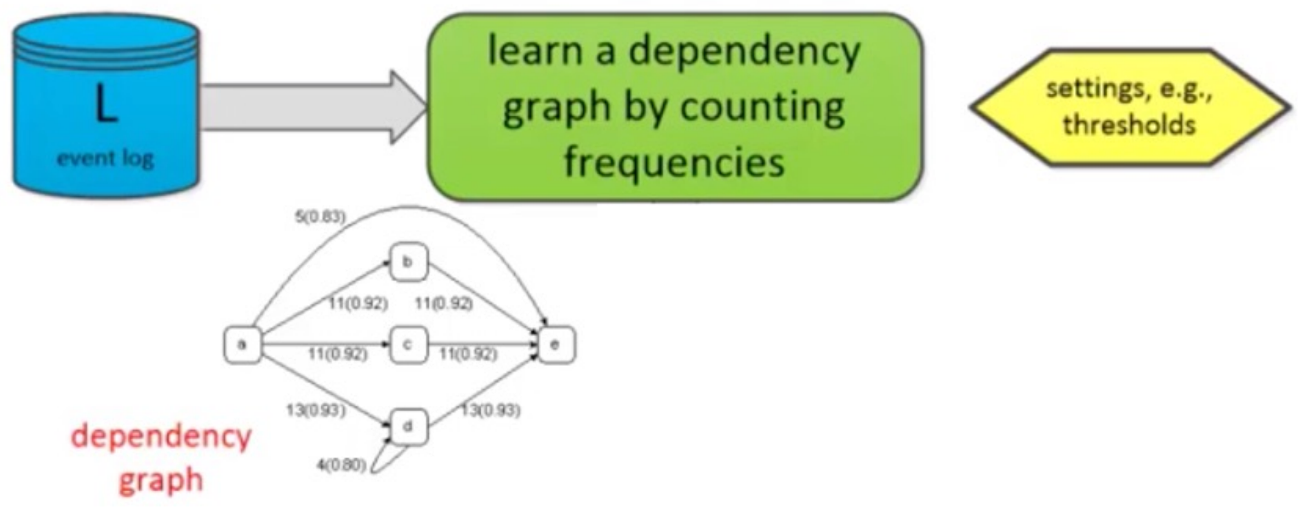
FAST - NUCES, CFD Campus

Dr. Rabia Maqsood

rabia.maqsood@nu.edu.pk

Today’s Topics

- Heuristic Mining
  - Learning Dependency Graph (quick recap)
  - Learning Causal Nets and perform annotations



Learning Dependency Graph

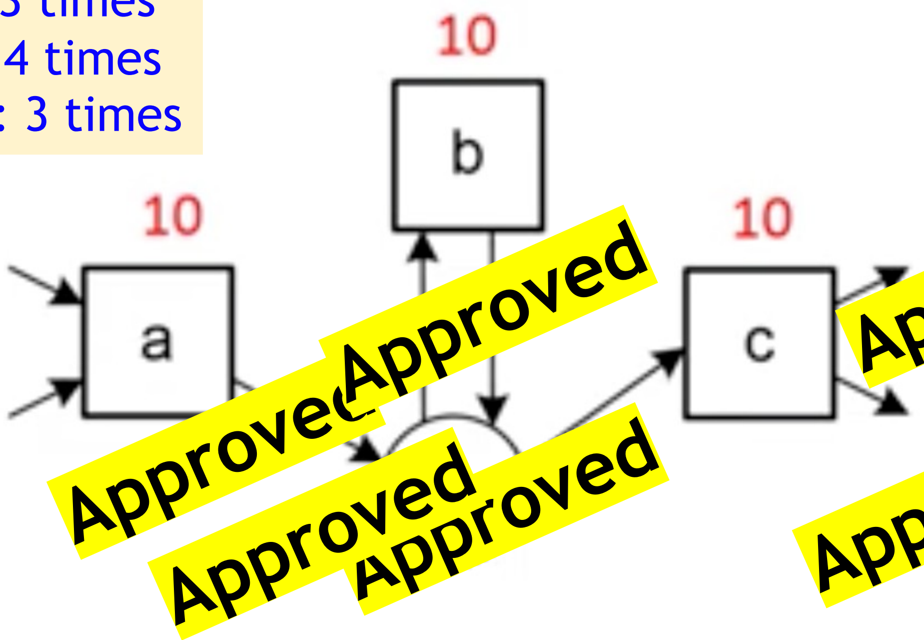
First step in the Heuristic Mining

Loop pattern

$$|a >_L b| = \sum_{\sigma \in L} L(\sigma) \times \left|\{1 \le i < |\sigma| \mid \sigma(i) = a \wedge \sigma(i+1) = b\}\right|$$

$$|a \Rightarrow_L b| = \begin{cases} \dfrac{|a>_L b| - |b>_L a|}{|a>_L b| + |b>_L a| + 1} & \mathit{if}\ a \neq b \\ \dfrac{|a>_L a|}{|a>_L a| + 1} & \mathit{if}\ a = b \end{cases}$$

Assume
ac: 3 times
abc:4 times
abbc: 3 times

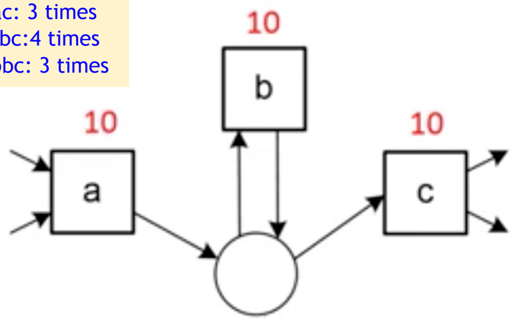


Approved Approved Approved Approved

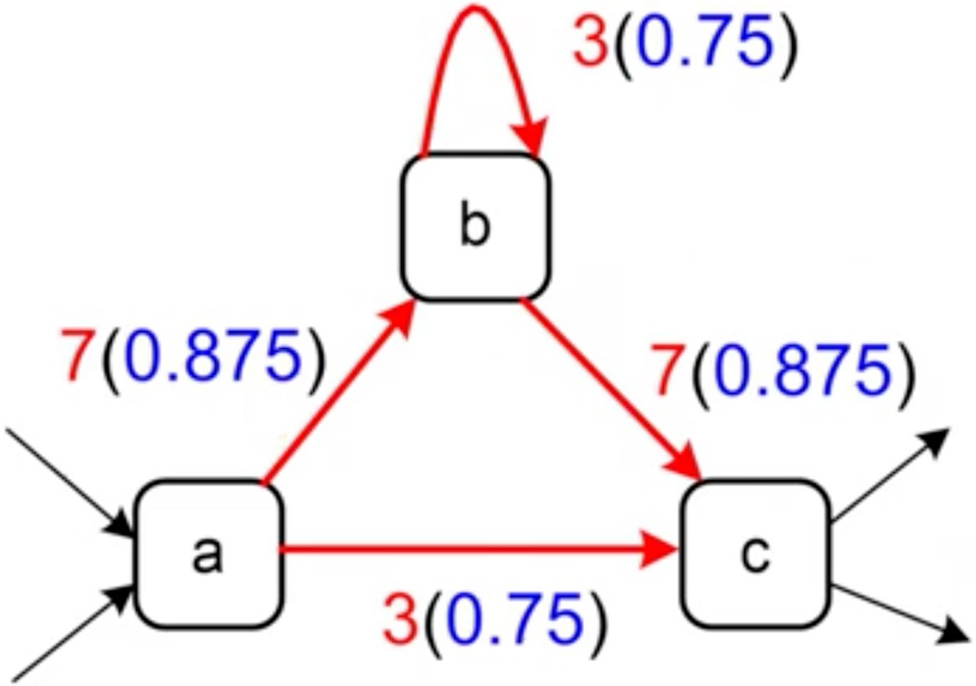
$\|a>_L a\| = 0$	$\|b>_L a\| = 0$	$\|c>_L a\| = 0$
$\|a\Rightarrow_L a\| = 0/1$	$\|b\Rightarrow_L a\| = -7/8$	$\|c\Rightarrow_L a\| = -3/4$
$\|a>_L b\| = 7$	$\|b>_L b\| = 3$	$\|c>_L b\| = 0$
$\|a\Rightarrow_L b\| = 7/8$	$\|b\Rightarrow_L b\| = 3/4$	$\|c\Rightarrow_L b\| = -7/8$
$\|a>_L c\| = 3$	$\|b>_L c\| = 7$	$\|c>_L c\| = 0$
$\|a\Rightarrow_L c\| = 3/4$	$\|b\Rightarrow_L c\| = 7/8$	$\|c\Rightarrow_L c\| = 0/1$

Included arcs (assuming thresholds >=1 and >=0.5)

Assume
ac: 3 times
abc:4 times
abbc: 3 times



$\vert a >_L a \vert = 0$ $\vert a \Rightarrow_L a \vert = 0/1$	$\vert b >_L a \vert = 0$ $\vert b \Rightarrow_L a \vert = -7/8$	$\vert c >_L a \vert = 0$ $\vert c \Rightarrow_L a \vert = -3/4$
$\vert a >_L b \vert = 7$ $\vert a \Rightarrow_L b \vert = 7/8$	$\vert b >_L b \vert = 3$ $\vert b \Rightarrow_L b \vert = 3/4$	$\vert c >_L b \vert = 0$ $\vert c \Rightarrow_L b \vert = -7/8$
$\vert a >_L c \vert = 3$ $\vert a \Rightarrow_L c \vert = 3/4$	$\vert b >_L c \vert = 7$ $\vert b \Rightarrow_L c \vert = 7/8$	$\vert c >_L c \vert = 0$ $\vert c \Rightarrow_L c \vert = 0/1$



Home Work

- Compute the dependency measures: $|a \Rightarrow_L b|$ and $|d \Rightarrow_L d|$ for the given event log.

$L = [\langle a, e\rangle^{5}, \langle a, b, c, e\rangle^{10}, \langle a, c, b, e\rangle^{10}, \langle a, b, e\rangle^{1}, \langle a, c, e\rangle^{1}, \langle a, d, e\rangle^{10}, \langle a, d, d, e\rangle^{2}, \langle a, d, d, d, e\rangle^{1}]$

$$\boxed{|a >_L b| = \sum_{\sigma \in L} L(\sigma) \times \left|\left\{1 \leq i < |\sigma| \;\middle|\; \sigma(i) = a \wedge \sigma(i+1) = b\right\}\right|}$$

$$\boxed{|a \Rightarrow_L b| = \begin{cases} \dfrac{|a >_L b| - |b >_L a|}{|a >_L b| + |b >_L a| + 1} & \text{if } a \neq b \\[2ex] \dfrac{|a >_L a|}{|a >_L a| + 1} & \text{if } a = b \end{cases}}$$

Solution

L = [⟨a, e⟩^{5}, ⟨a, b, c, e⟩^{10}, ⟨a, c, b, e⟩^{10}, ⟨a, b, e⟩^{1}, ⟨a, c, e⟩^{1}, ⟨a, d, e⟩^{10}, ⟨a, d, d, e⟩^{2}, ⟨a, d, d, d, e⟩^{1}]

$\lvert \Rightarrow_L \rvert$	a	b	c	d	e
a	$\frac{0}{0+1}=0$	$\frac{11-0}{11+0+1}=0.92$	$\frac{11-0}{11+0+1}=0.92$	$\frac{13-0}{13+0+1}=0.93$	$\frac{5-0}{5+0+1}=0.83$
b	$\frac{0-11}{0+11+1}=-0.92$	$\frac{0}{0+1}=0$	$\frac{10-10}{10+10+1}=0$	$\frac{0-0}{0+0+1}=0$	$\frac{11-0}{11+0+1}=0.92$
c	$\frac{0-11}{0+11+1}=-0.92$	$\frac{10-10}{10+10+1}=0$	$\frac{0}{0+1}=0$	$\frac{0-0}{0+0+1}=0$	$\frac{11-0}{11+0+1}=0.92$
d	$\frac{0-13}{0+13+1}=-0.93$	$\frac{0-0}{0+0+1}=0$	$\frac{0-0}{0+0+1}=0$	$\frac{4}{4+1}=0.80$	$\frac{13-0}{13+0+1}=0.93$
e	$\frac{0-5}{0+5+1}=-0.83$	$\frac{0-11}{0+11+1}=-0.92$	$\frac{0-11}{0+11+1}=-0.92$	$\frac{0-13}{0+13+1}=-0.93$	$\frac{0}{0+1}=0$

Example

L = [⟨a, e⟩^{5}, ⟨a, b, c, e⟩^{10}, ⟨a, c, b, e⟩^{10}, ⟨a, b, e⟩^{1}, ⟨a, c, e⟩^{1}, ⟨a, d, e⟩^{10}, ⟨a, d, d, e⟩^{2}, ⟨a, d, d, d, e⟩^{1}]

$\|\Rightarrow_L\|$	a	b	c	d	e
a	$\frac{0}{0+1}=0$	$\boxed{\frac{11-0}{11+0+1}=0.92}$	$\frac{11-0}{11+0+1}=0.92$	$\frac{13-0}{13+0+1}=0.93$	$\frac{5-0}{5+0+1}=0.83$
b	$\frac{0-11}{0+11+1}=-0.92$	$\frac{0}{0+1}=0$	$\frac{10-10}{10+10+1}=0$	$\frac{0-0}{0+0+1}=0$	$\frac{11-0}{11+0+1}=0.92$
c	$\frac{0-11}{0+11+1}=-0.92$	$\frac{10-10}{10+10+1}=0$	$\frac{0}{0+1}=0$	$\frac{0-0}{0+0+1}=0$	$\frac{11-0}{11+0+1}=0.92$
d			$\frac{0-0}{0+0+1}=0$	$\frac{4}{4+1}=0.80$	$\frac{13-0}{13+0+1}=0.93$
e			$\frac{0-11}{0+11+1}=-0$		

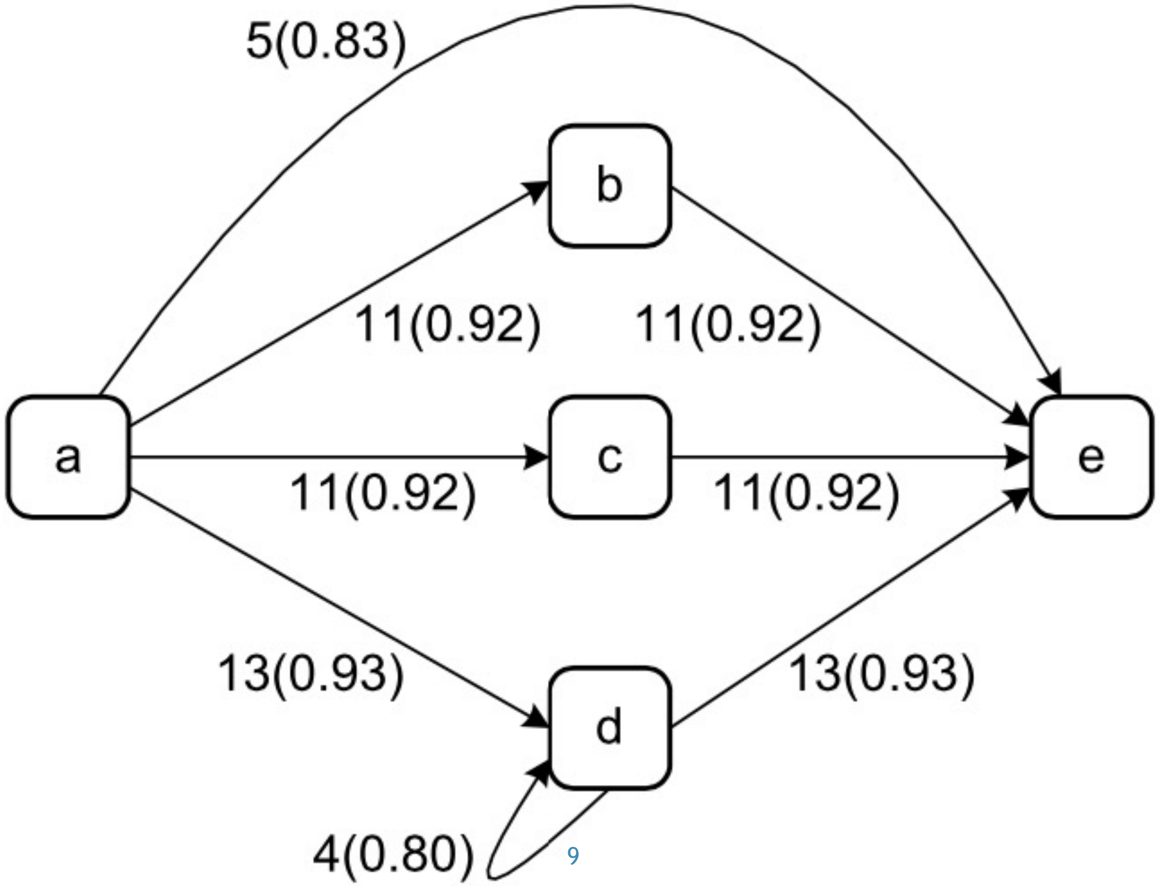
$\|>_L\|$	a	b	c	d	e
a	0	11	11	13	5
b	0	0	10	0	11
c	0	10	0	0	11
d	0	0	0	4	13
e	0	0	0	0	0

$$|a \Rightarrow_L b| = \begin{cases} \frac{|a>_L b| - |b>_L a|}{|a>_L b| + |b>_L a| + 1} & \text{if } a \neq b \\ \frac{|a>_L a|}{|a>_L a| + 1} & \text{if } a = b \end{cases}$$

Dependency graph using a lower threshold
(threshold of 2 for |>L| and 0.7 for |⇒L|)

$\vert\Rightarrow_L\vert$	a	b	c	d	e
a	$\frac{0}{0+1}=0$	$\frac{11-0}{11+0+1}=0.92$	$\frac{11-0}{11+0+1}=0.92$	$\frac{13-0}{13+0+1}=0.93$	$\frac{5-0}{5+0+1}=0.83$
b	$\frac{0-11}{0+11+1}=-0.92$	$\frac{0}{0+1}=0$	$\frac{10-10}{10+10+1}=0$	$\frac{0-0}{0+0+1}=0$	$\frac{11-0}{11+0+1}=0.92$
c	$\frac{0-11}{0+11+1}=-0.92$	$\frac{10-10}{10+10+1}=0$	$\frac{0}{0+1}=0$	$\frac{0-0}{0+0+1}=0$	$\frac{11-0}{11+0+1}=0.92$
d	$\frac{0-13}{0+13+1}=-0.93$	$\frac{0-0}{0+0+1}=0$	$\frac{0-0}{0+0+1}=0$	$\frac{4}{4+1}=0.80$	$\frac{13-0}{13+0+1}=0.93$
e	$\frac{0-5}{0+5+1}=-0.83$	$\frac{0-11}{0+11+1}=-0.92$	$\frac{0-11}{0+11+1}=-0.92$	$\frac{0-13}{0+13+1}=-0.93$	$\frac{0}{0+1}=0$

$\vert>_L\vert$	a	b	c	d	e
a	0	11	11	13	5
b	0	0	10	0	11
c	0	10	0	0	11
d	0	0	0	4	13
e	0	0	0	0	0

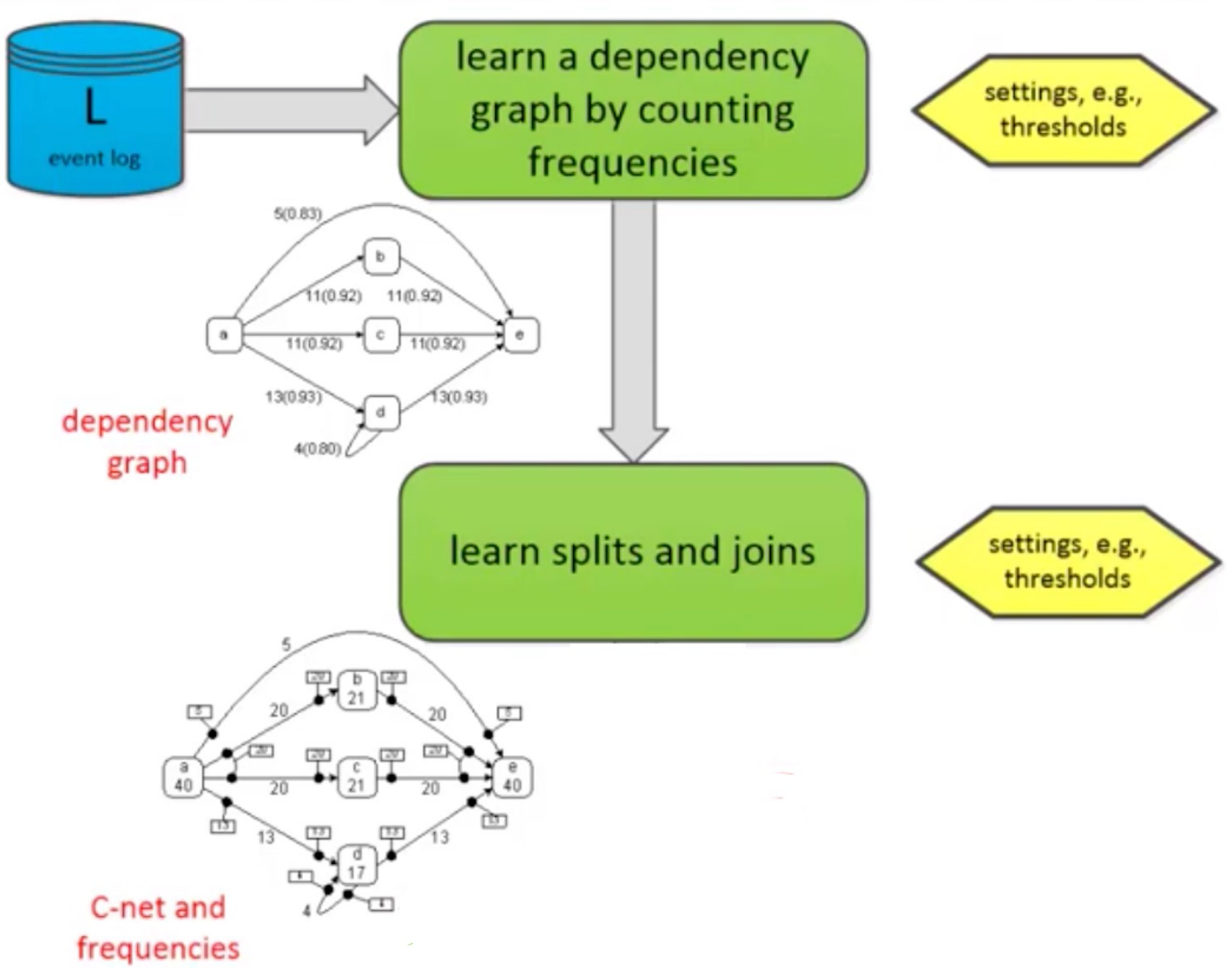


Computing the dependency graph

1. Set thresholds for the minimal number of direct successions and dependency measures
2. Count direct successions
3. Compute dependency measures
4. Draw dependency graph including only arcs that meet both thresholds

Learning Splits/Joins

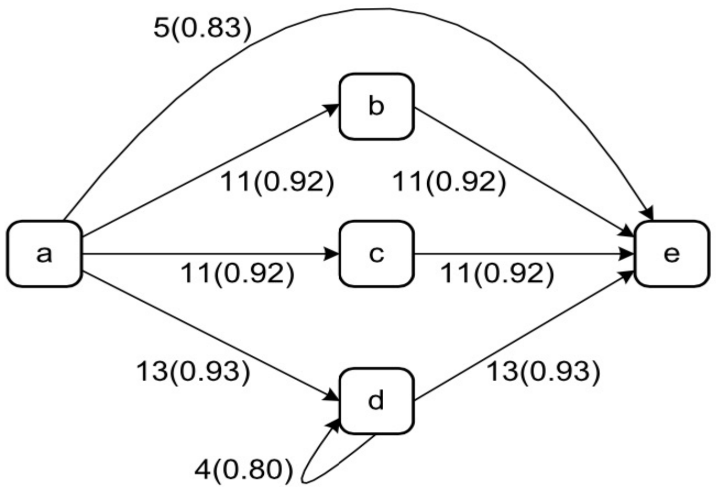
Second step in the Heuristic Mining



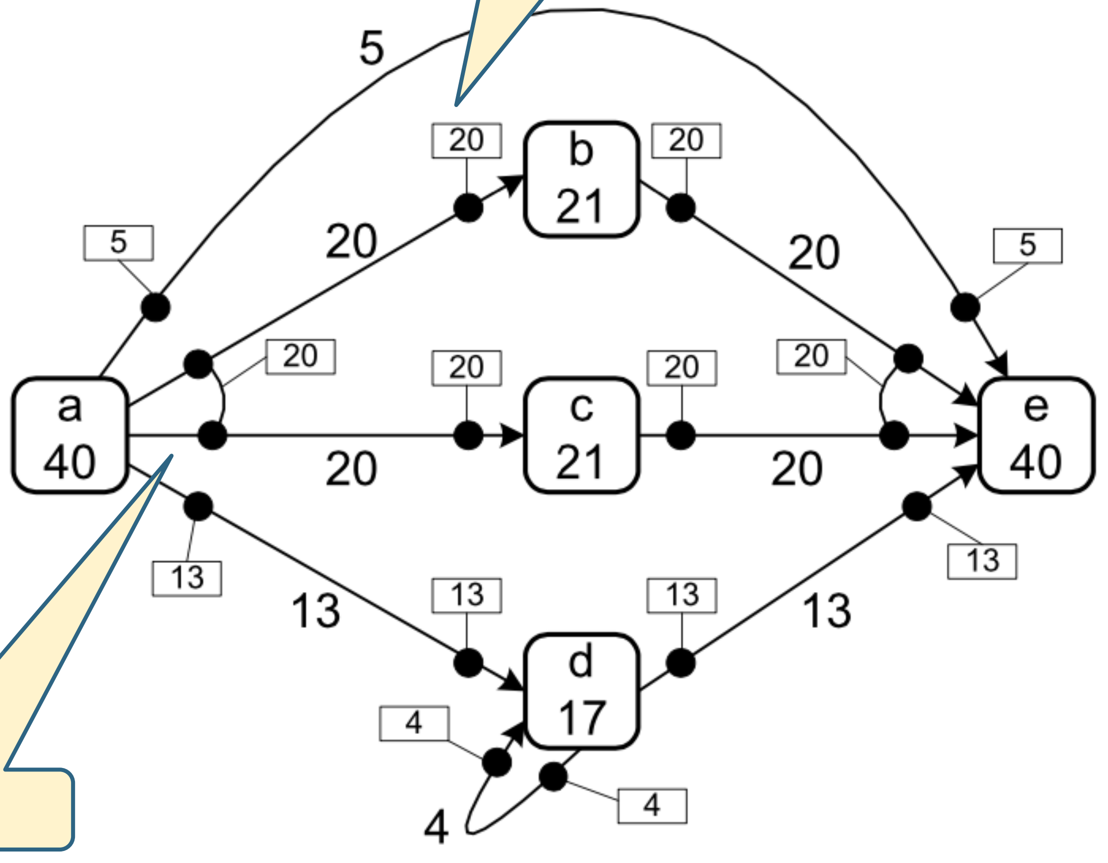
Desired output: a causal net

Splits/joins and frequencies

$L = [\langle a, e\rangle^{5}, \langle a, b, c, e\rangle^{10}, \langle a, c, b, e\rangle^{10}, \langle a, b, e\rangle^{1}, \langle a, c, e\rangle^{1}, \langle a, d, e\rangle^{10}, \langle a, d, d, e\rangle^{2}, \langle a, d, d, d, e\rangle^{1}]$



Dependency Graph



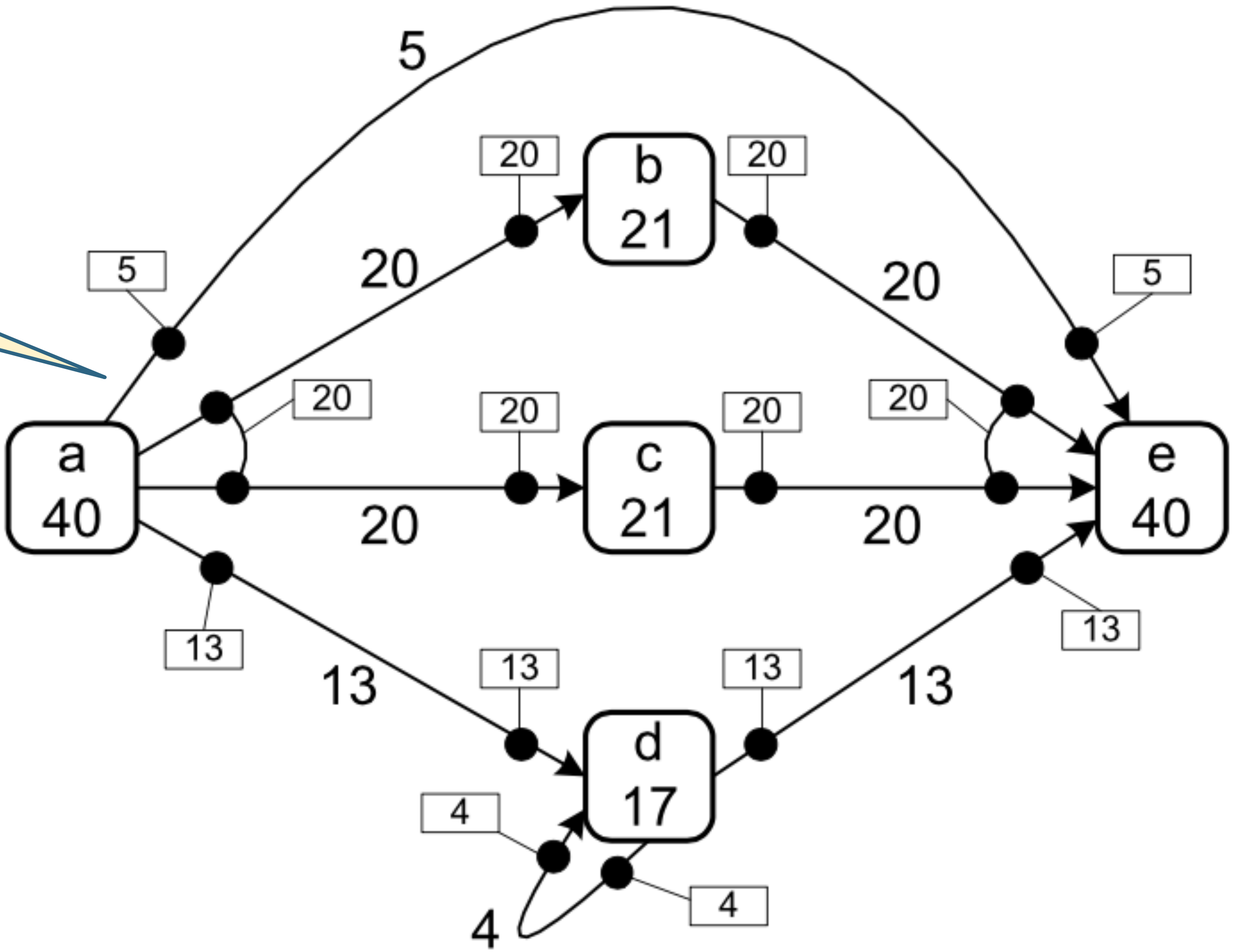
Frequencies

Splits/Joins

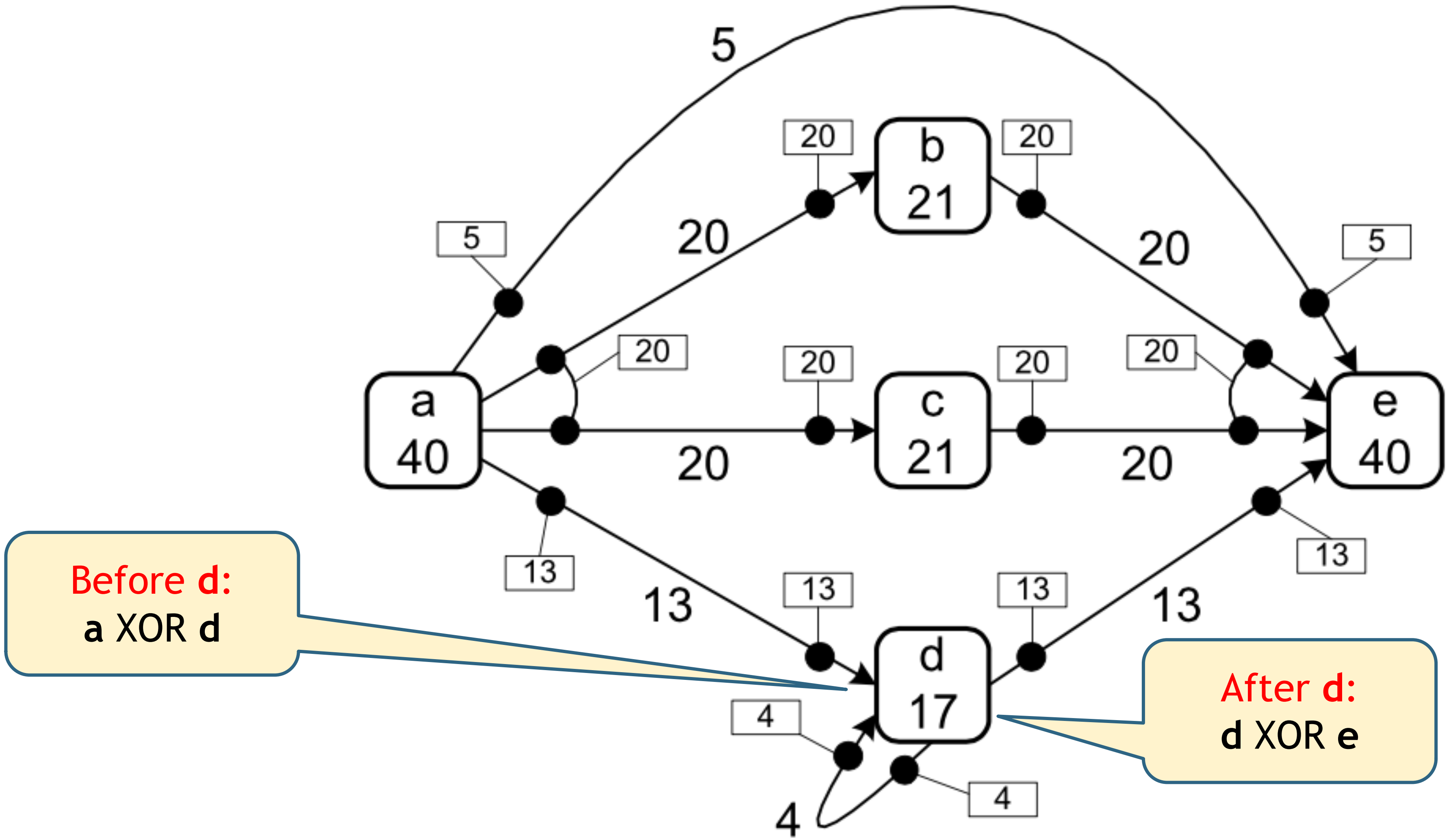
Causal Net

Interpreting Bindings
Example: output bindings of activity **a**

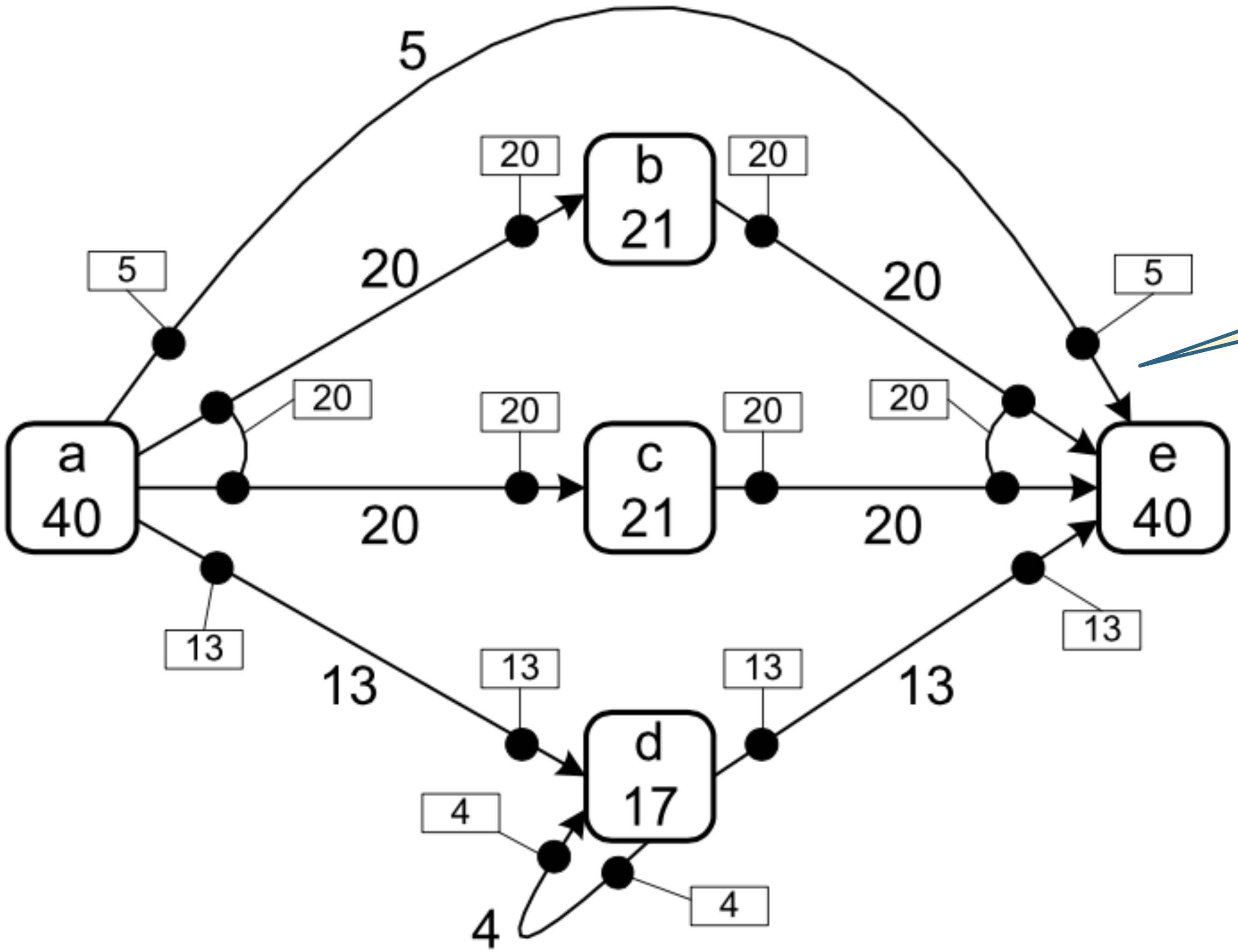
After **a**:
e XOR (**b** AND **c**) XOR **d**



Interpreting Bindings
Example: input & output bindings of activity **d**



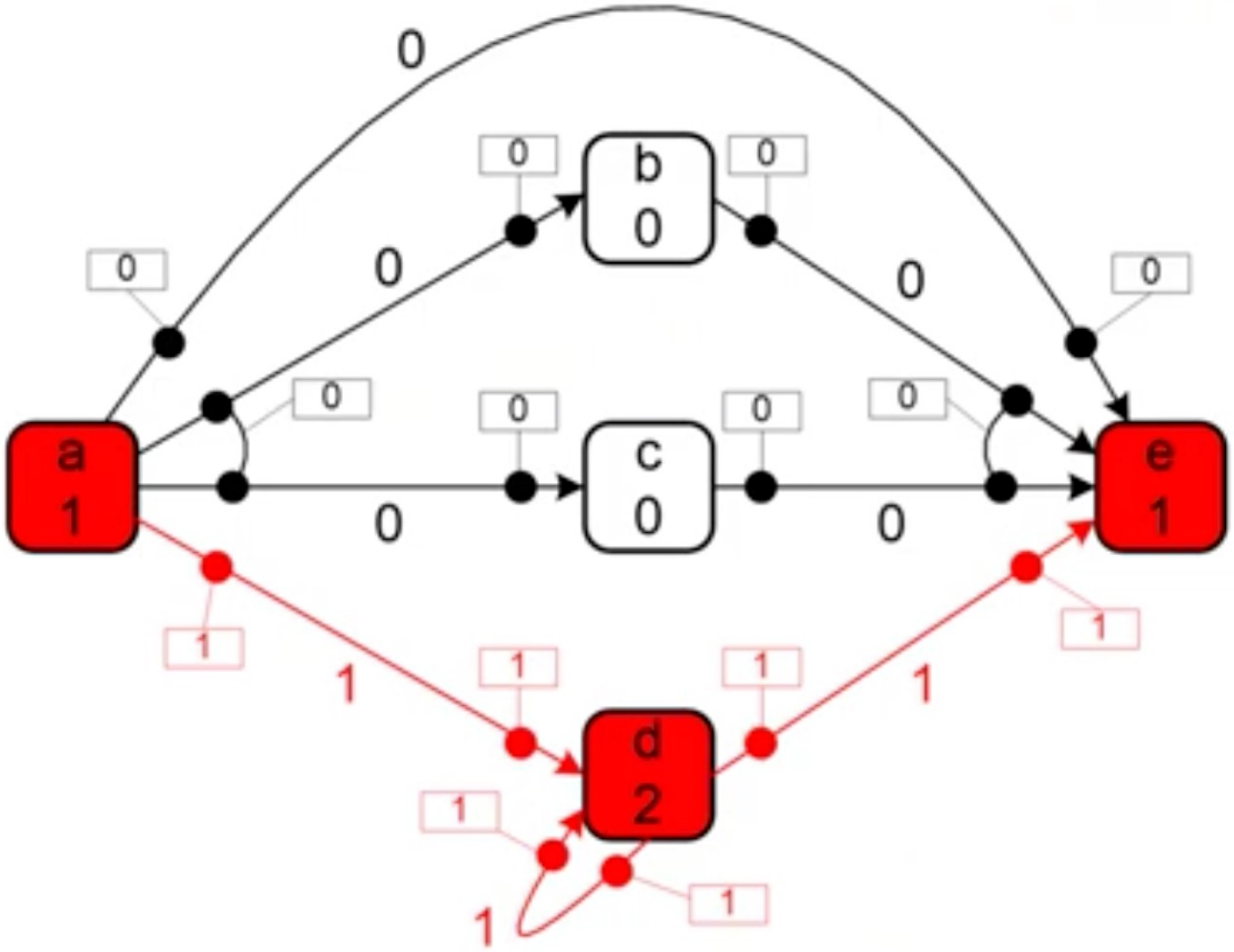
Interpreting Bindings
Example: input bindings of activity **e**



Before **e**:
a XOR (**b** AND **c**) XOR **d**

Example path: adde

L = [⟨a, e⟩^{5}, ⟨a, b, c, e⟩^{10}, ⟨a, c, b, e⟩^{10}, ⟨a, b, e⟩^{1}, ⟨a, c, e⟩^{1}, ⟨a, d, e⟩^{10}, ⟨a, d, d, e⟩^{2}, ⟨a, d, d, d, e⟩^{1}]



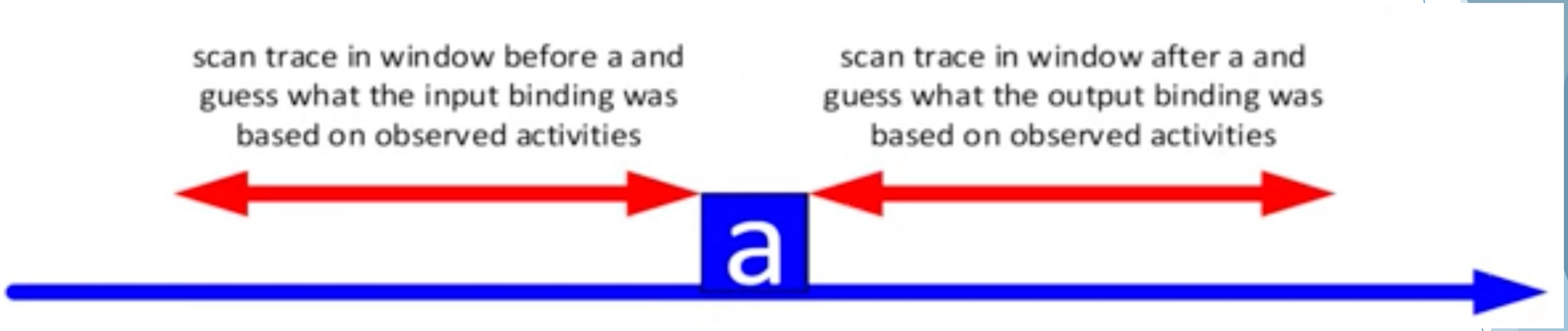
How to discover splits and joins?

- Two classes of approaches:

1. **Heuristics** using a **time window** before and after each activity.
   - By counting sets of input and output activities the bindings can be determined (local decision).
2. **Optimization** approaches based on **replay**.
   - Given a set of possible input and output bindings, one can see whether reality can be **replayed properly**.
   - The set of possible input and output bindings is finite, so a “**best set of bindings**” can be determined using some goal function.

- Many variations are possible!

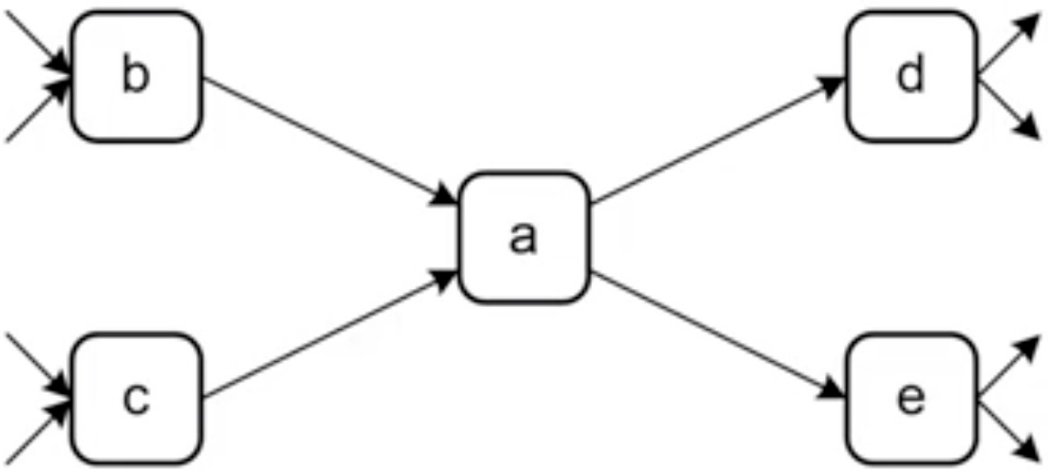
Approach 1: based on heuristics



- Activities have possible inputs and outputs (based on dependency graph).
- Count how often they appear in a window before (for input bindings) and a window after (for output bindings).

Example: window size 4

1....klbgadhek...
2....lkgcahedl...
3....kblgaehdk...
4....klgbadehk...
5....klkcadkeh...



Count frequencies of input and output activities

input bindings

- {b}: 3 times
- {c}: 2 times

output bindings

- {d,e}: 5 times

Add bindings and frequencies

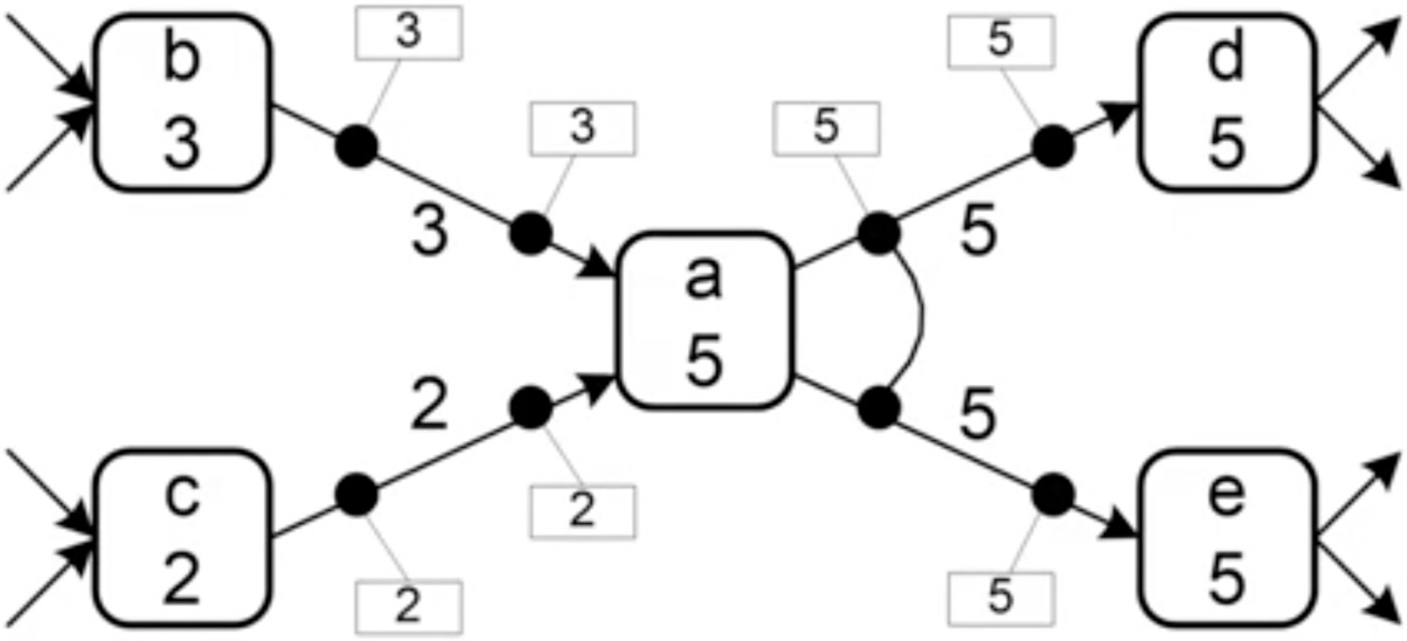
1. ...klbgadhek...
2. ...lkgcahedl...
3. ...kblgaehdk...
4. ...klgbadehk...
5. ...klkcadkeh...

input bindings

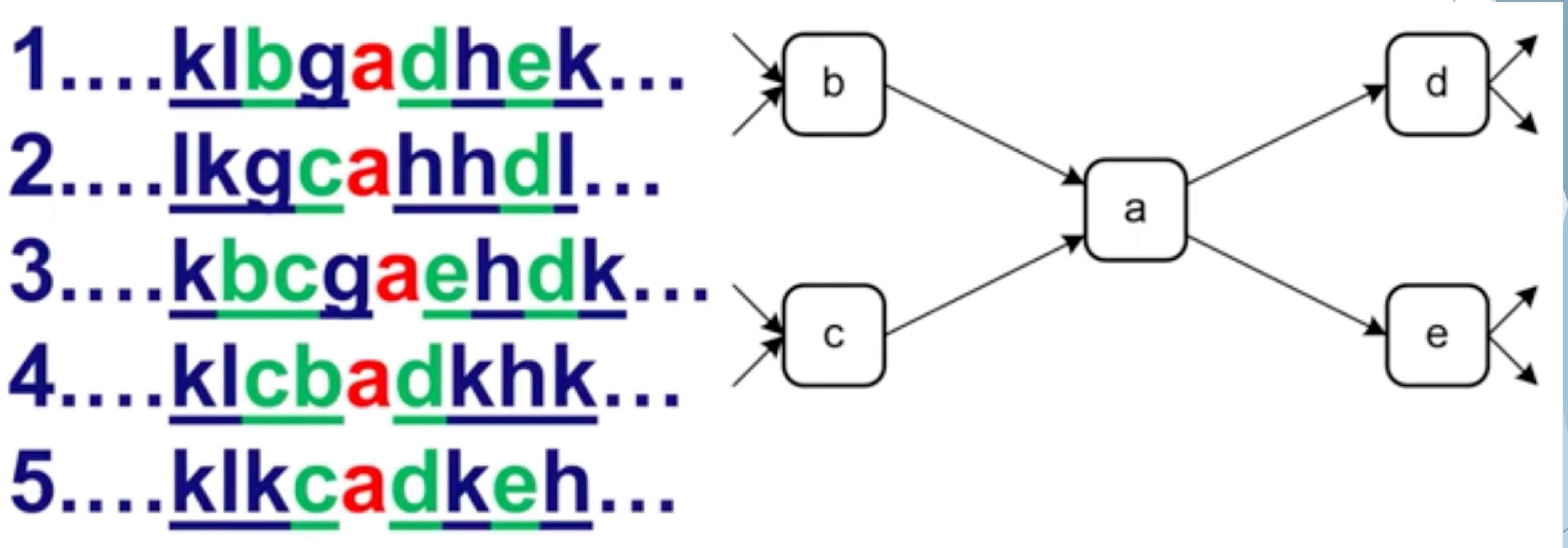
- {b}: 3 times
- {c}: 2 times

output bindings

- {d,e}: 5 times



Another example: window size 4



Count frequencies of input and output activities

input bindings
- {b}: 1 time
- {c}: 2 times
- {b,c}: 2 times

output bindings
- {d}: 2 times
- {d,e}: 3 times

Add bindings and frequencies

Do this yourself!

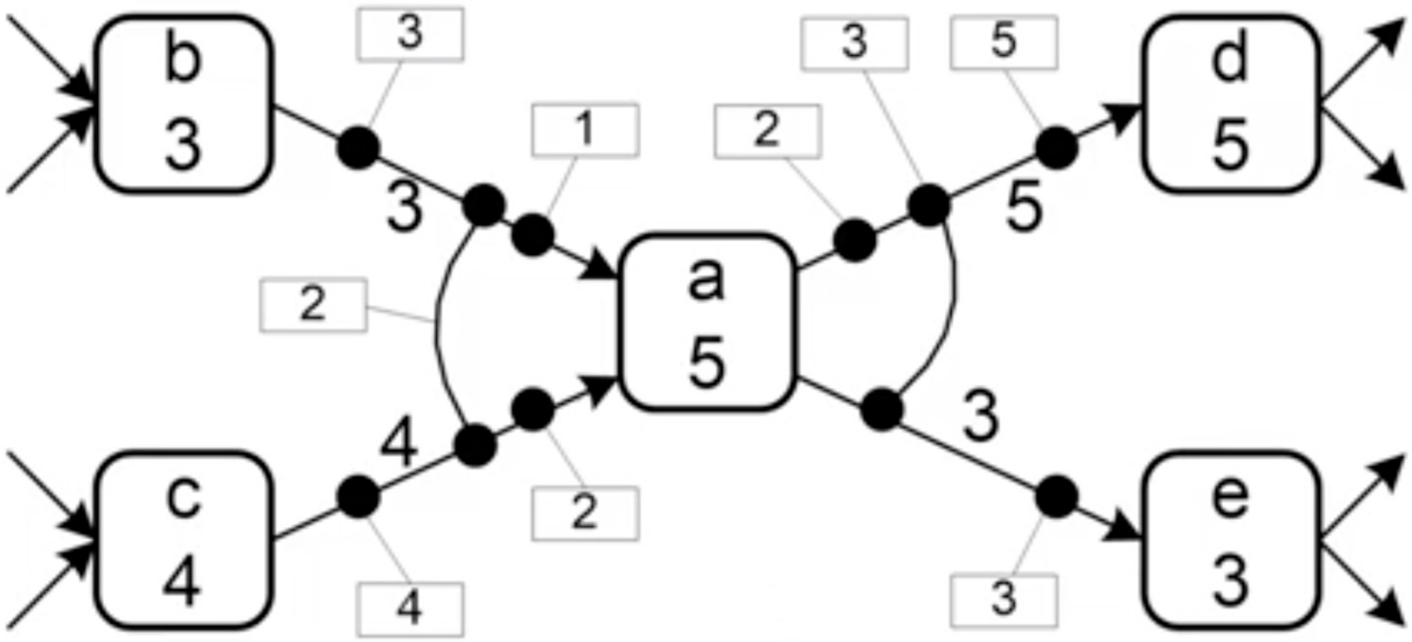
1. ...klbgadhek...
2. ...lkgcahhdl...
3. ...kbcgaehdk...
4. ...klcbadkhk...
5. ...klkcadkeh...

input bindings

- {b}: 1 time
- {c}: 2 times
- {b,c}: 2 times

output bindings

- {d}: 2 times
- {d,e}: 3 times



Refinements needed

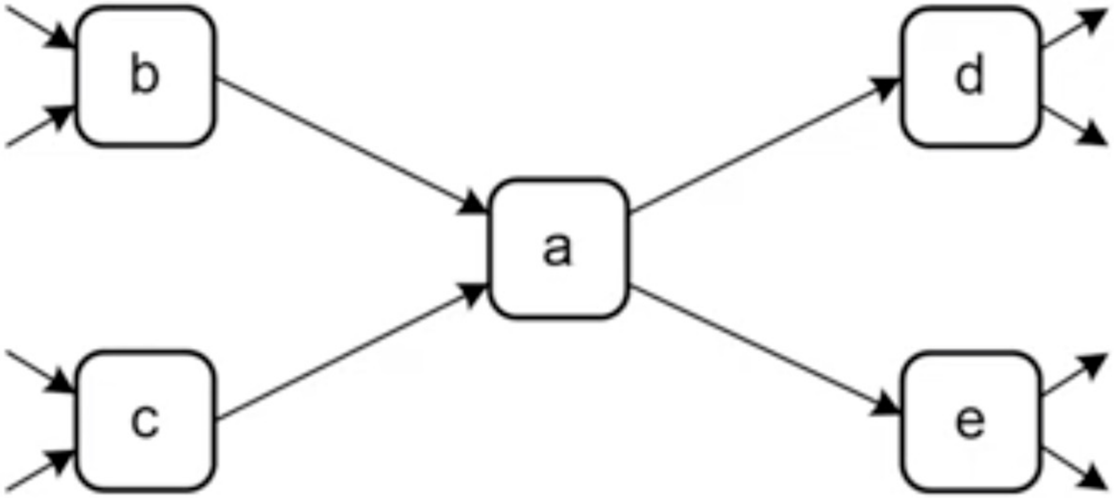
- What if there are no corresponding activities in the input or output window?
- Noise filtering: remove infrequent bindings
- Handling repeating activities (e.g., cut off window size)
- Details are out of scope but be aware of such complications when interpreting results!

Approach 2: optimization problem

- Evaluate all possible activity bindings and take best one.
- Based on the idea that ideally a trace can be **replayed from the initial state to the final state**.
- This can be checked precisely using various replay approaches (will be discussed later).
- Hence, one can use approaches that simply **“try bindings”** exhaustively.

Example: sets of input and output bindings

Each input/output arc needs to be involved in at least one binding.

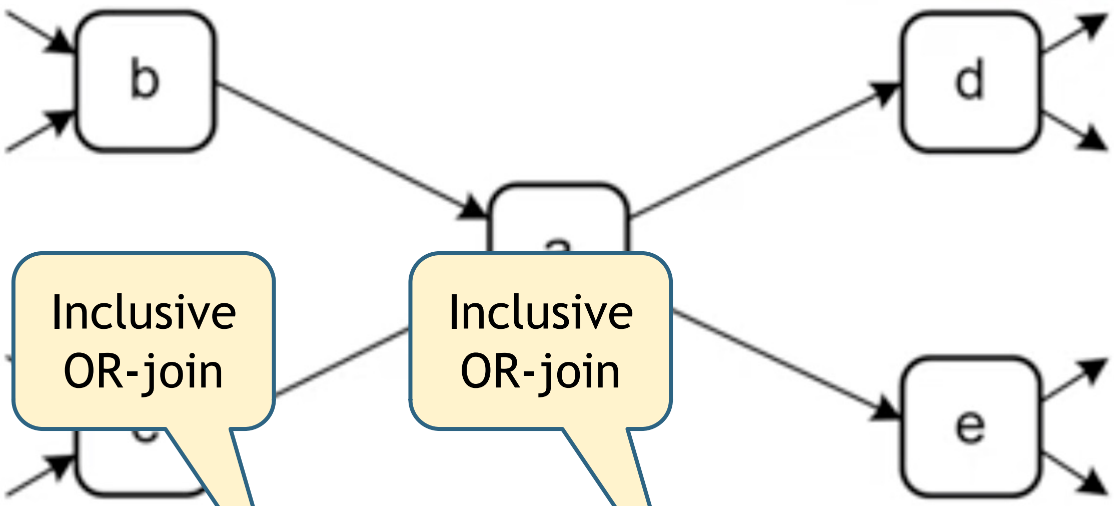


There are

$|\{\ \{\{b,c\}\}\ ,\ \{\{b\},\{c\}\}\ ,\ \{\{b\},\{b,c\}\}\ ,\ \{\{c\},\{b,c\}\}\ ,\ \{\{b\},\{c\},\{b,c\}\}\ \}|$ x
$|\{\ \{\{d,e\}\}\ ,\ \{\{d\},\{e\}\}\ ,\ \{\{d\},\{d,e\}\}\ ,\ \{\{e\},\{d,e\}\}\ ,\ \{\{d\},\{e\},\{d,e\}\}\ \}|$
= 5 x 5 = 25 possible a activities.

Example: sets of input and output bindings

Each input/output arc needs to be involved in at least one binding.



AND-join

XOR-join

Inclusive OR-join

Inclusive OR-join

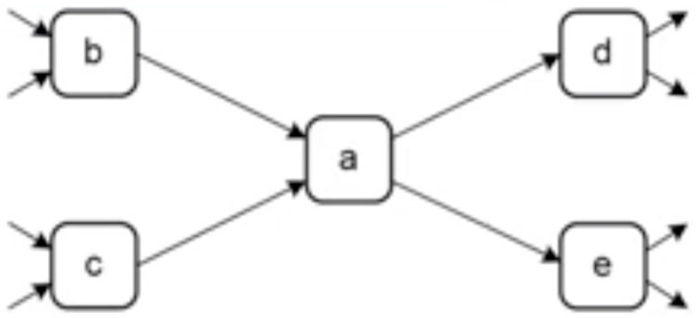
Inclusive OR-join

There are

|{ {{b,c}} , {{b},{c}} , {{b},{b,c}} , {{c},{b,c}} , {{b},{c},{b,c}} }| x
|{ {{d,e}} , {{d},{e}} , {{d},{d,e}} , {{e},{d,e}} , {{d},{e},{d,e}} }|
= 5 x 5 = 25 possible a activities.

Optimization approach

- **For each activity select one of the input-output binding combinations.**
- **One can do this exhaustively and try all combinations.**
- **Evaluation can be done using replay.**
- **Take best one (taking into account fitness, precision, generalization, and simplicity).**



|{ {{b,c}} ,{{b},{c}} , {{b},{b,c}} , {{c},{b,c}} , {{b},{c},{b,c}} }| x
|{ {{d,e}} , {{d},{e}} , {{d},{d,e}} , {{e},{d,e}} , {{d},{e},{d,e}} }|
= 5 x 5 = 25 possible a activities

Possible refinements

- If too time consuming:
  - Randomize
  - Use a genetic algorithm

Reading Material

- Chapter 7: Aalst