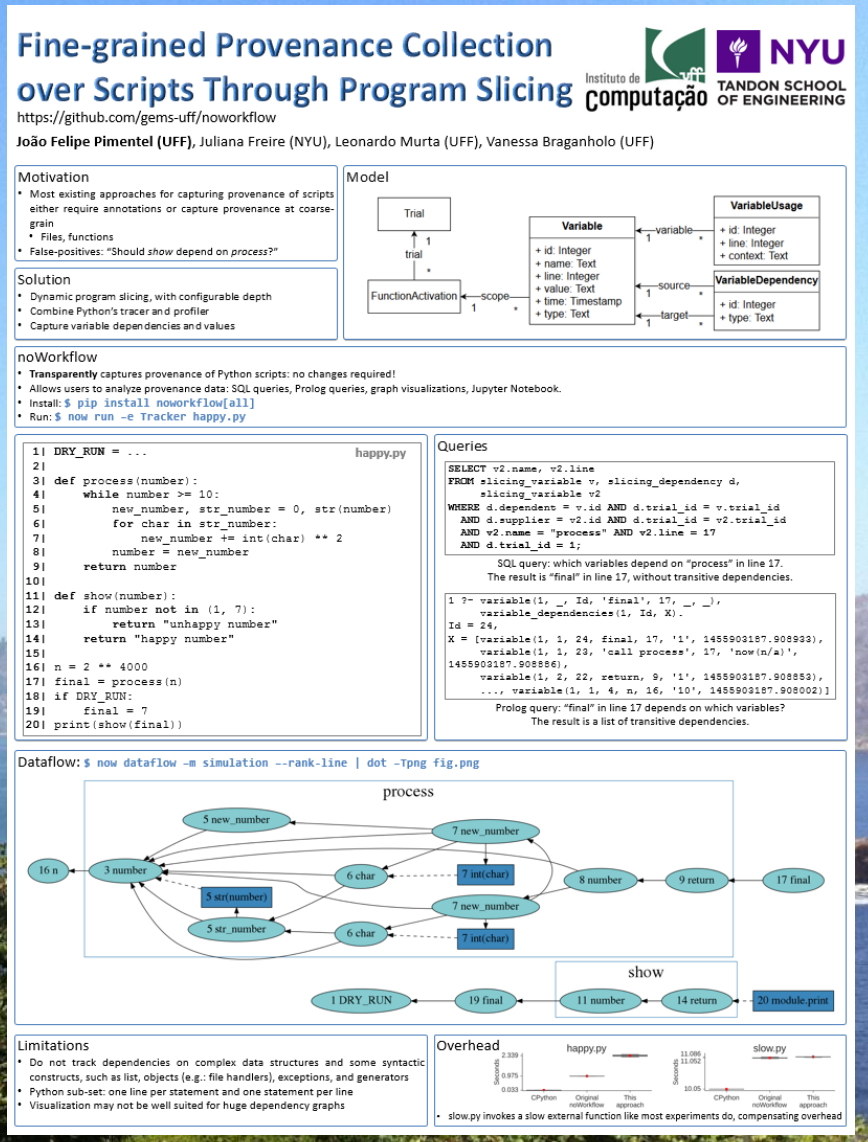
Fine-grained Provenance Collection over Scripts Through Program Slicing



jpimentel@ic.uff.br

github.com/gems-uff/noworkflow

João Felipe Nicolaci Pimentel (UFF),

Juliana Freire (NYU), Leonardo Murta (UFF),

Vanessa Braganholo (UFF)

Instituto de Computação

NYU TANDON SCHOOL OF ENGINEERING

Provenance of Scripts

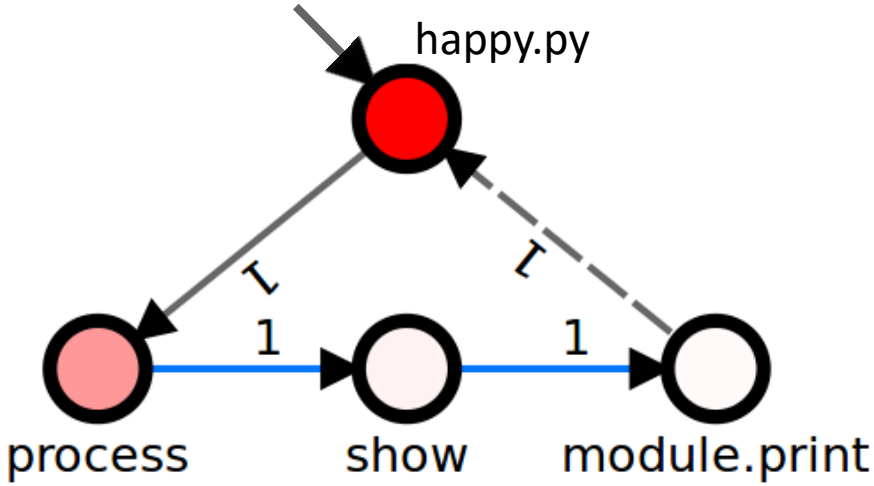
- YesWorkflow
  - McPhillips et al., 2015
- API
  - Bochner; Gude; Schreiber, 2008
- StarFlow
  - Angelino; Yamins; Seltzer, 2010
- RDataTracker
  - Lerner and Boose, 2014
- noWorkflow
  - Murta et al., 2014
- Sumatra
  - Davison, 2012
- LLVM compiler
  - Tariq et al., 2012
- DataTracker
  - Stamatogiannakis et al., 2014
- …

Provenance of Scripts

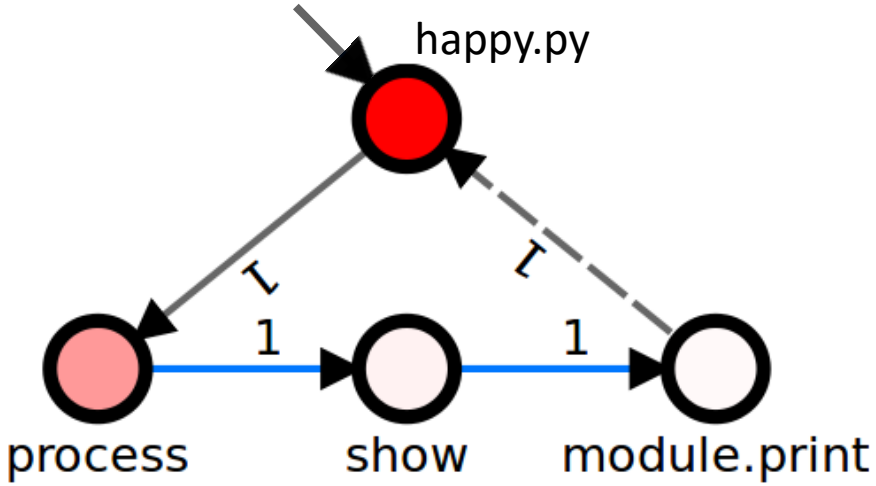
- YesWorkflow
  - McPhillips et al., 2015
- noWorkflow
  - Murta et al., 2014
- Coarse-grained Provenance
  - Files
  - Functions
  - …
- Seltzer, 2010
- RDataTracker
  - Lerner and Boose, 2014
  - Stamatogiannakis et al., 2014
- …

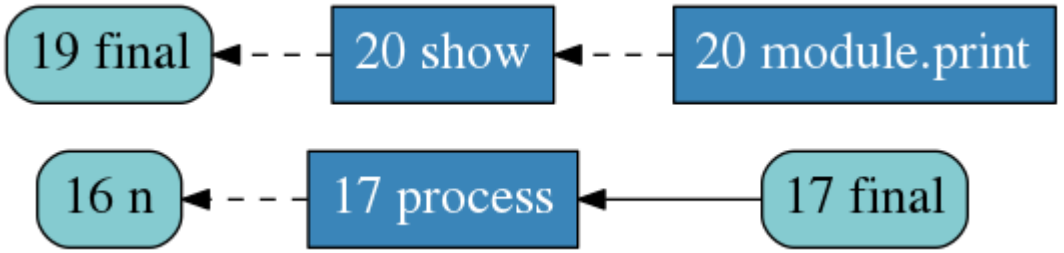
```
 1| DRY_RUN = ...
 2|
 3| def process(number):
 4|     ...
10|
11| def show(number):
12|     ...
15|
16| n = 10
17| final = process(n)
18| if DRY_RUN:
19|    final = 7
20| print(show(final))
```

```
 1| DRY_RUN = ...
 2|
 3| def process(number):
 4|     ...
10|
11| def show(number):
12|     ...
15|
16| n = 10
17| final = process(n)
18| if DRY_RUN:
19|    final = 7
20| print(show(final))
```

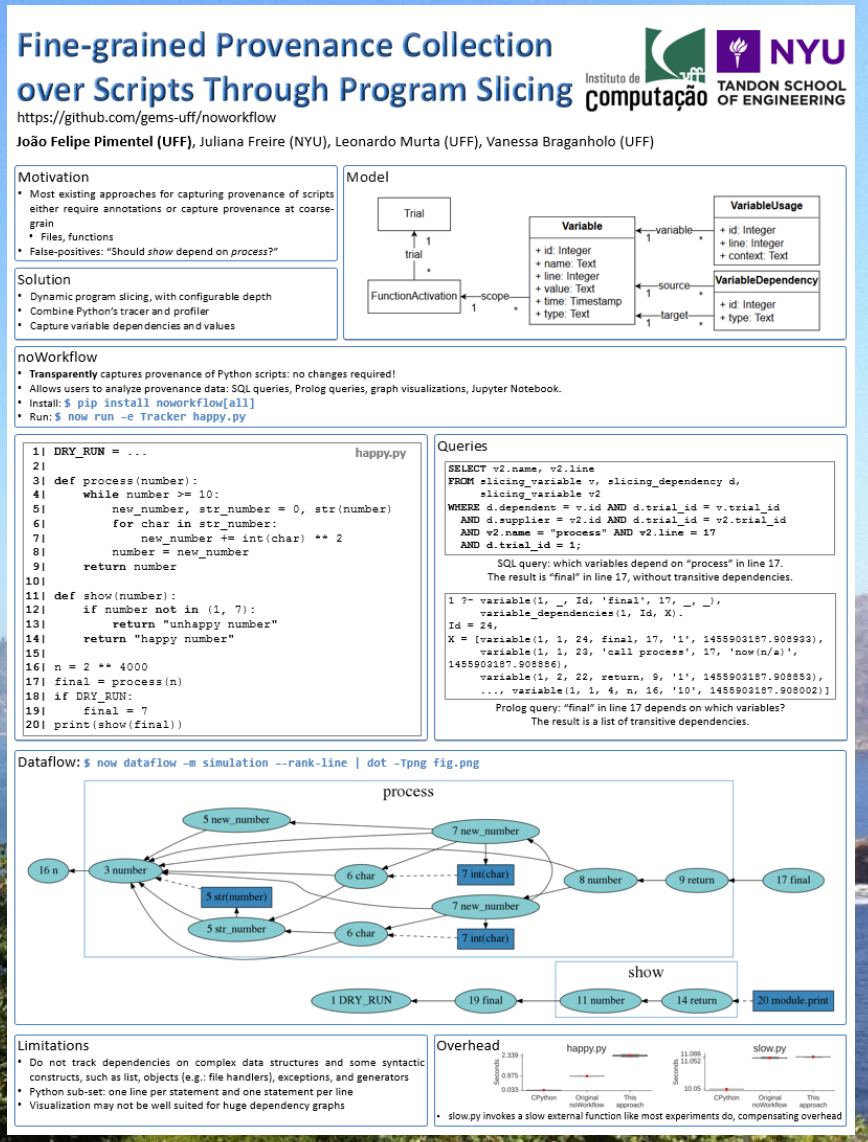


```
 1| DRY_RUN = True
 2|
 3| def process(number):
 4|     ...
10|
11| def show(number):
12|     ...
15|
16| n = 10
17| final = process(n)
18| if DRY_RUN:
19|     final = 7
20| print(show(final))
```





Fine-grained Provenance Collection over Scripts Through Program Slicing



jpimentel@ic.uff.br

github.com/gems-uff/noworkflow

João Felipe Nicolaci Pimentel (UFF),

Juliana Freire (NYU), Leonardo Murta (UFF),

Vanessa Braganholo (UFF)

Instituto de Computação

NYU TANDON SCHOOL OF ENGINEERING