Karlsruhe Institute of Technology

# Remodularizing Legacy Model Transformations with Automatic Clustering Techniques

**Andreas Rentschler**, Dominik Werle, Qais Noorshams, Lucia Happe, Ralf Reussner

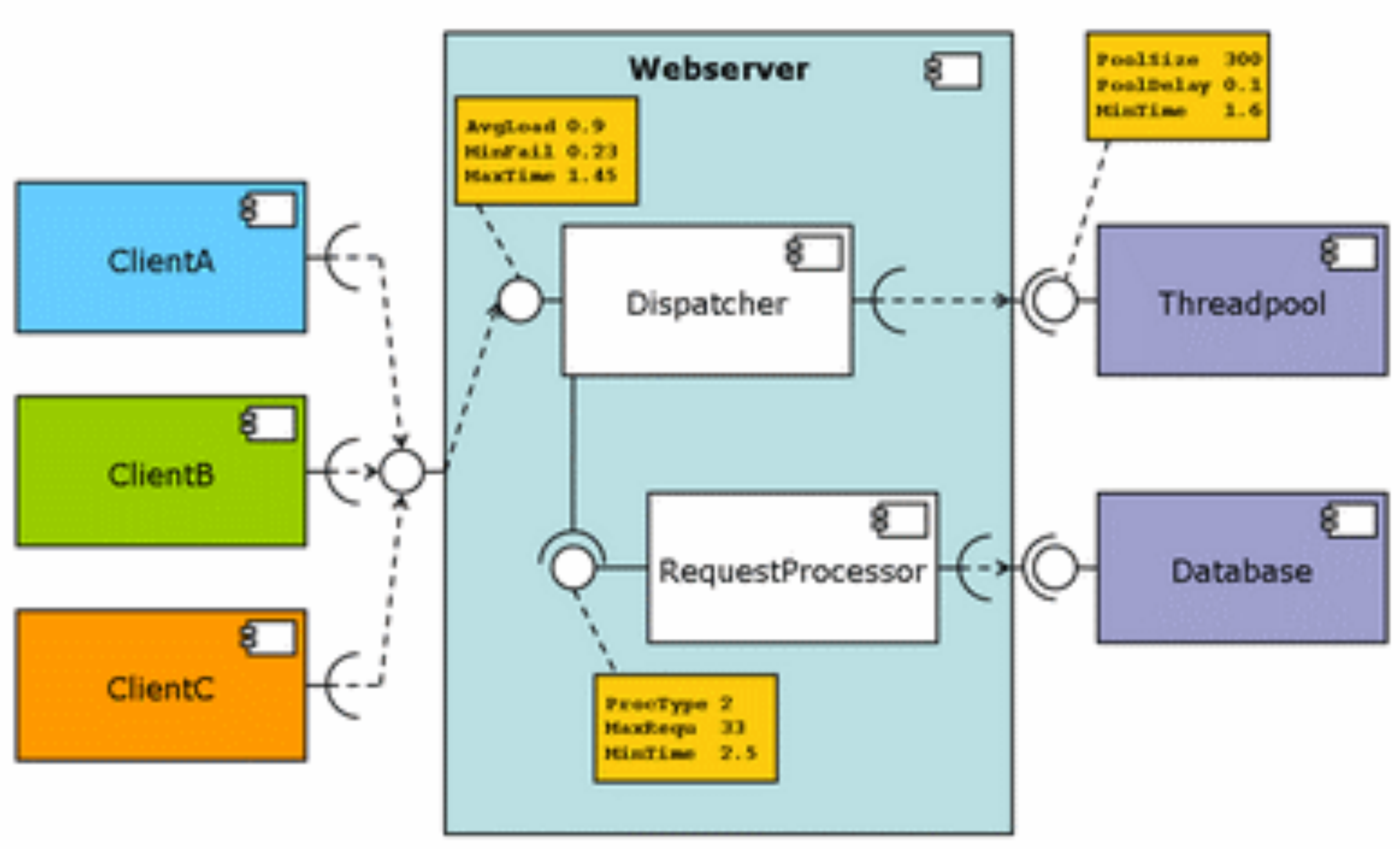
3rd Workshop on the Analysis of Model Transformations
Monday, September 29, 2014

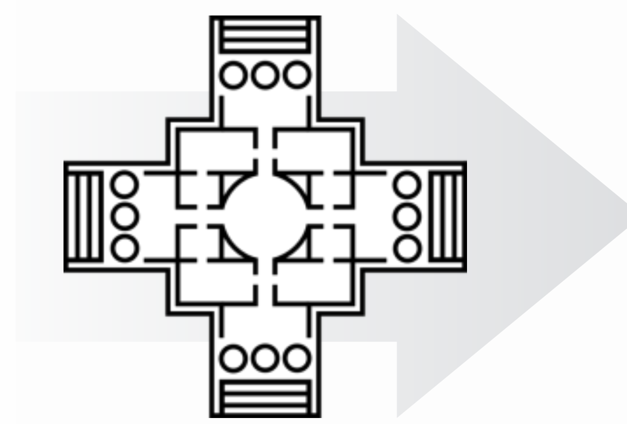
SOFTWARE DESIGN AND QUALITY GROUP
INSTITUTE FOR PROGRAM STRUCTURES AND DATA ORGANIZATION, FACULTY OF INFORMATICS

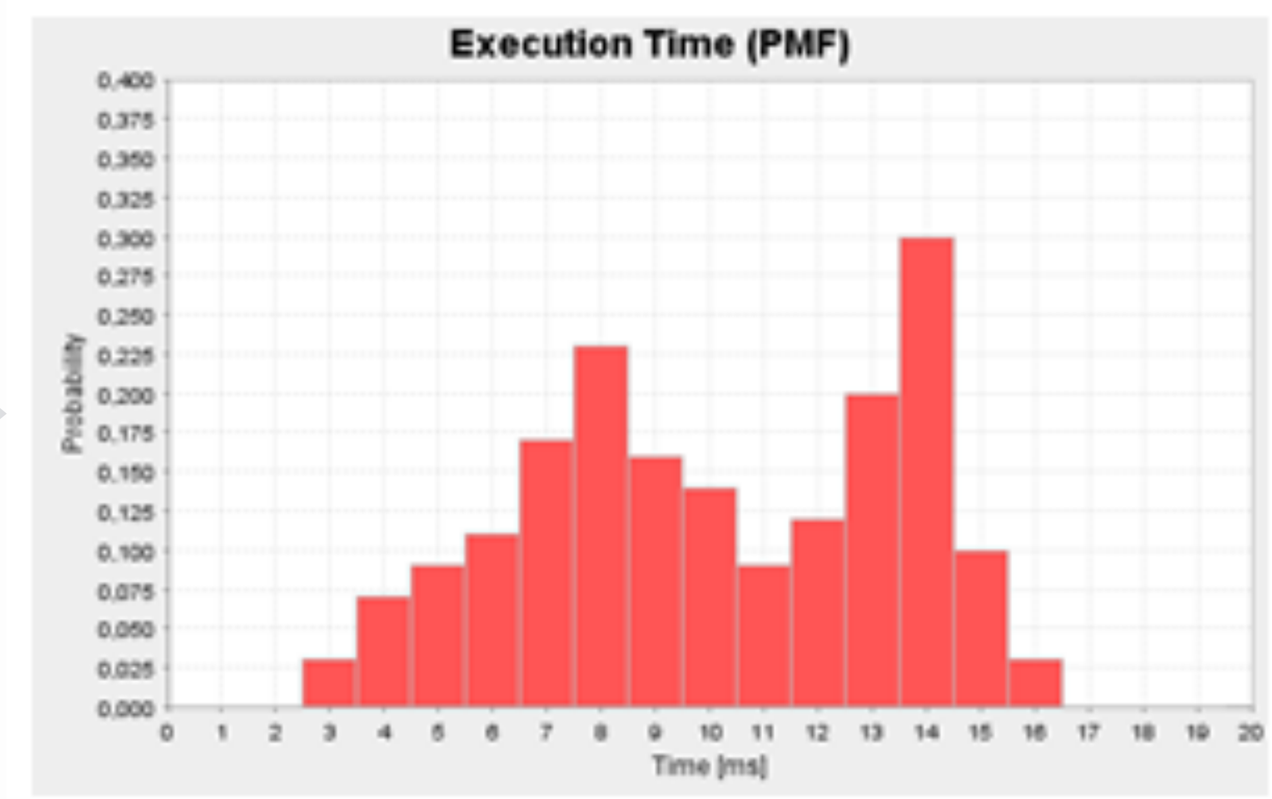
sdq.ipd.kit.edu

Source: pixelio.de

# Model-driven Software Quality Prediction



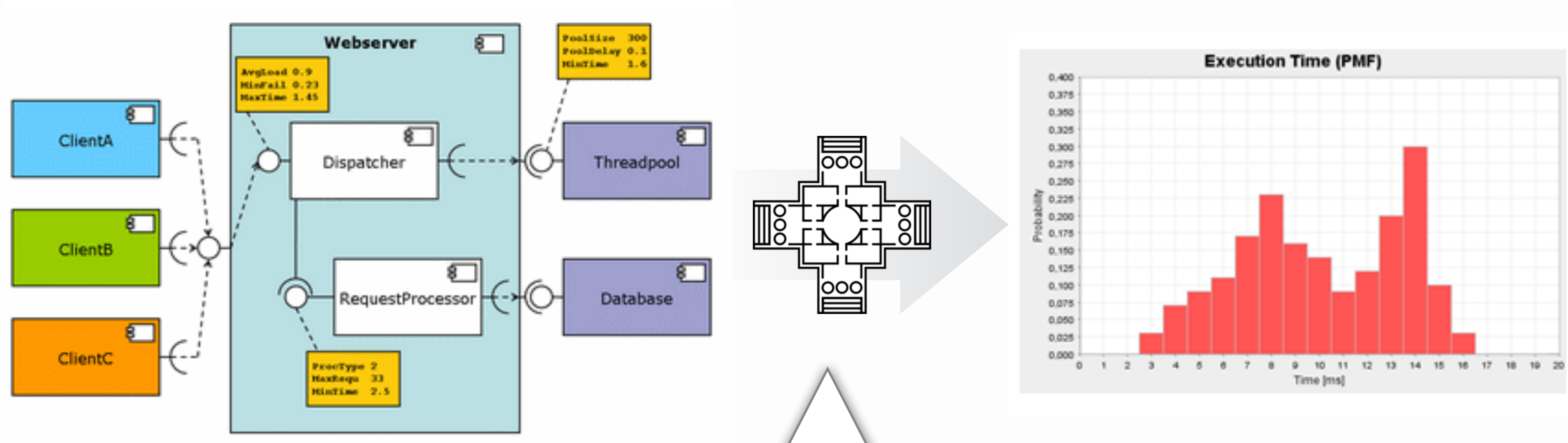


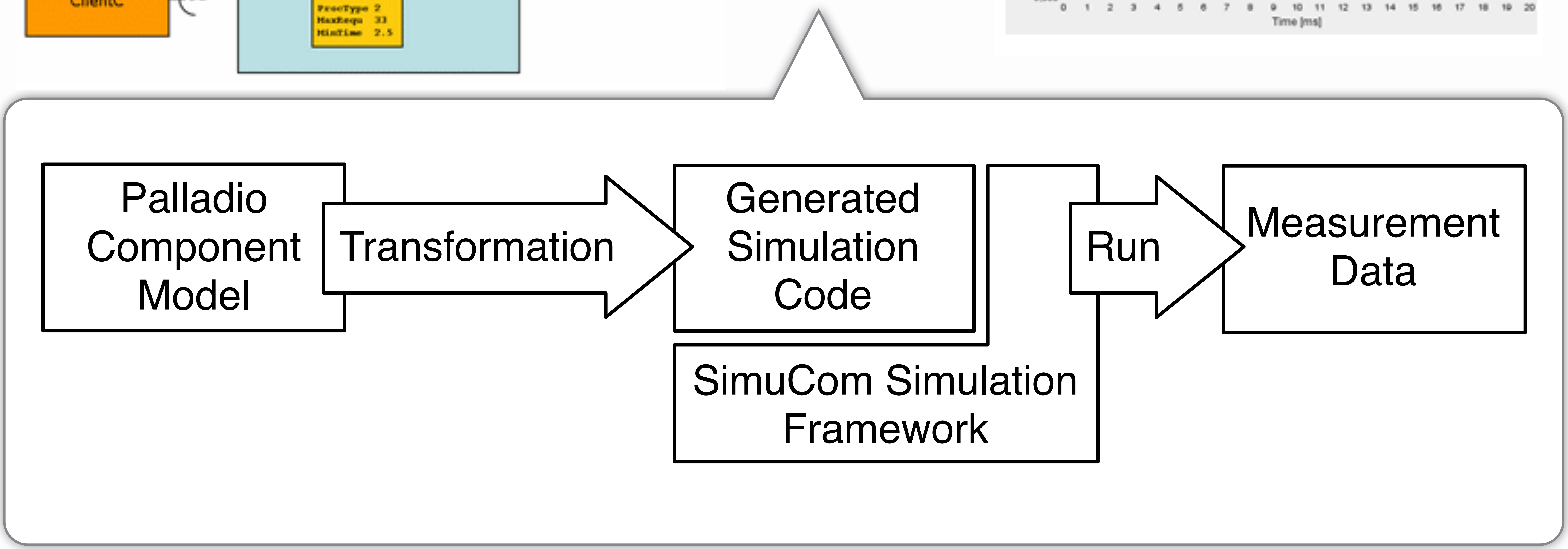

Performance model
of a component-based
software architecture

Performance data:

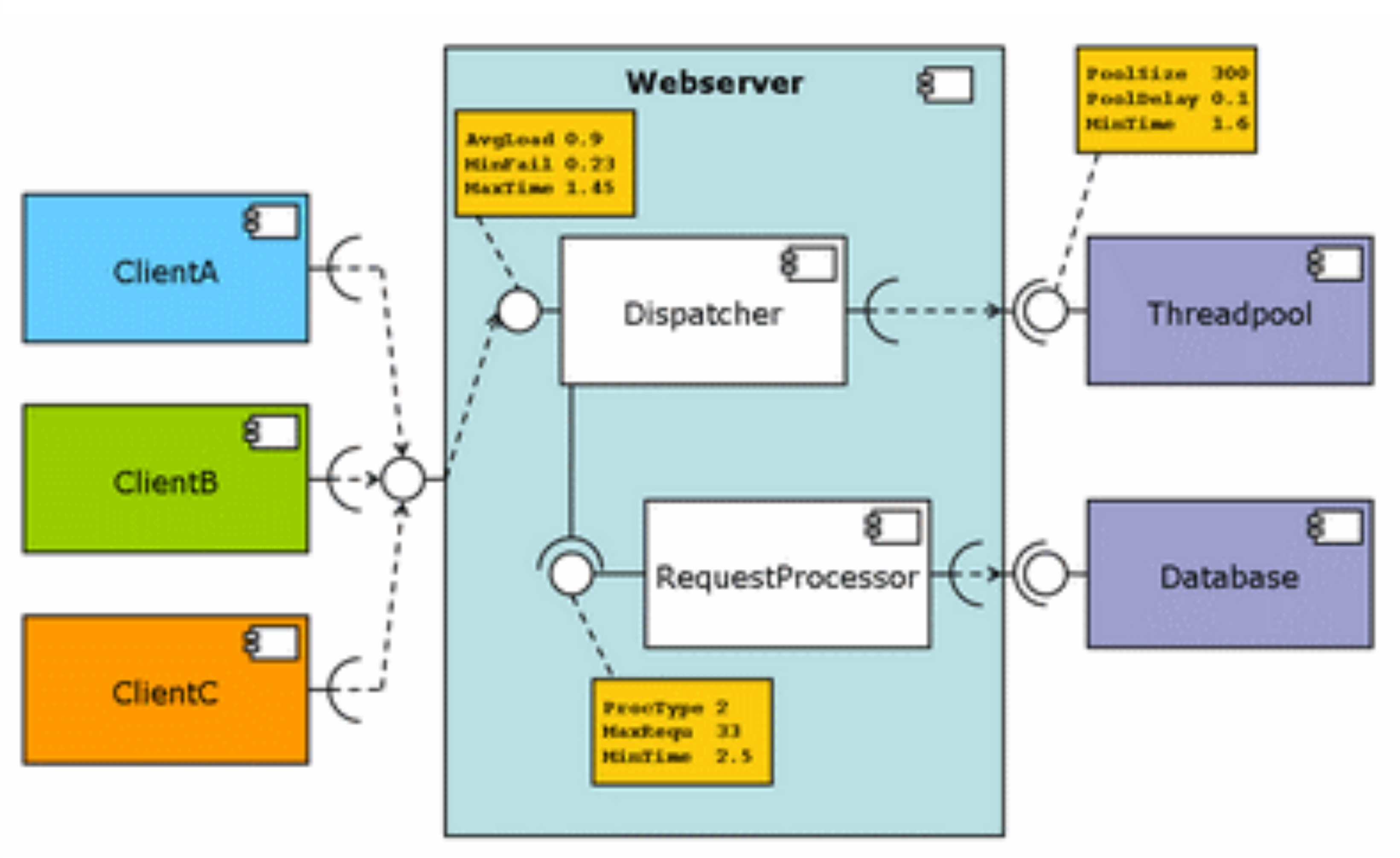
- Execution time
- Throughput
- Resource utilisation

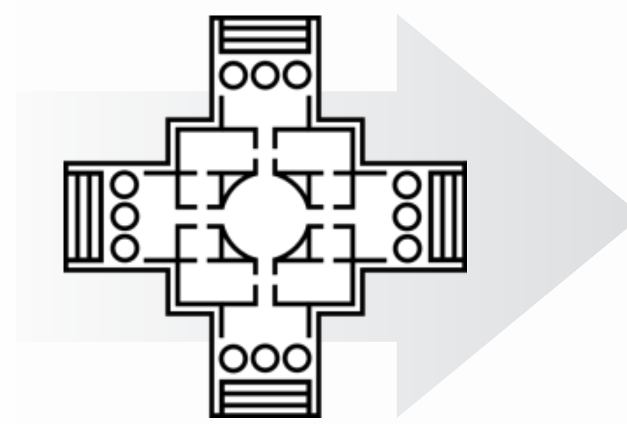
# Model-driven Software Quality Prediction

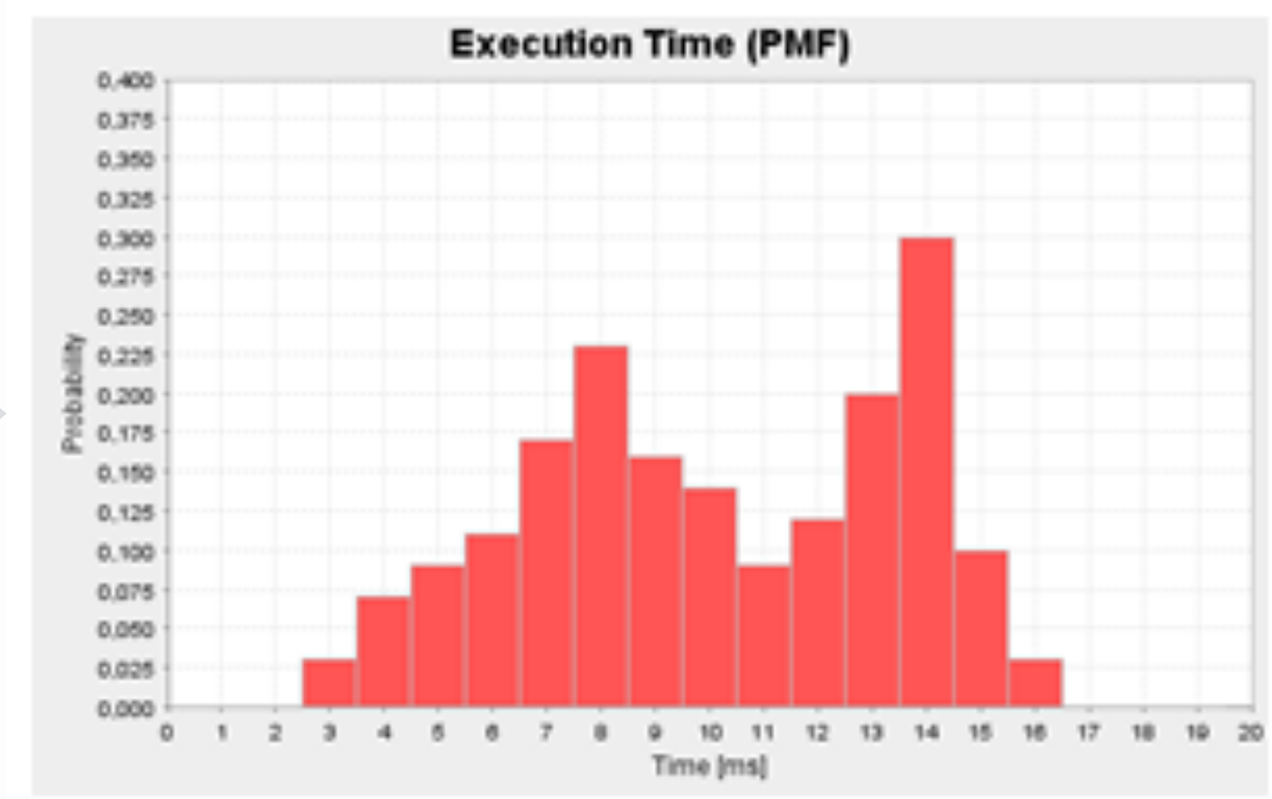


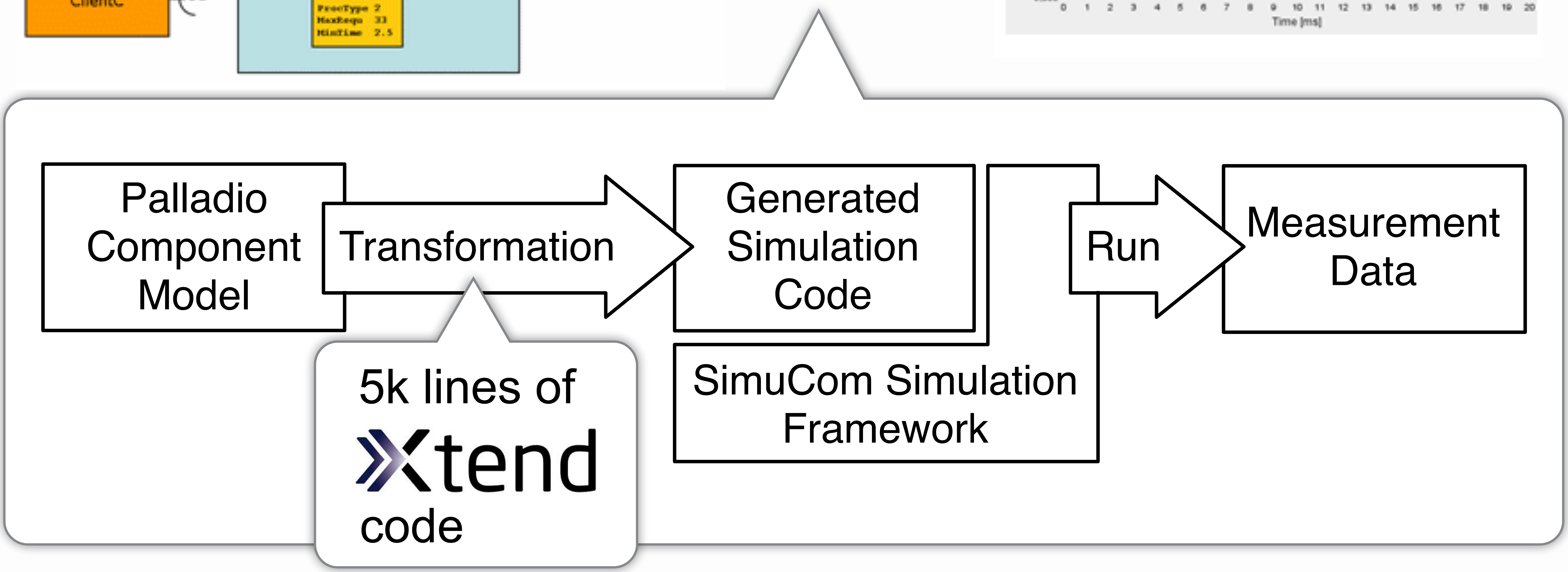



# Model-driven Software Quality Prediction

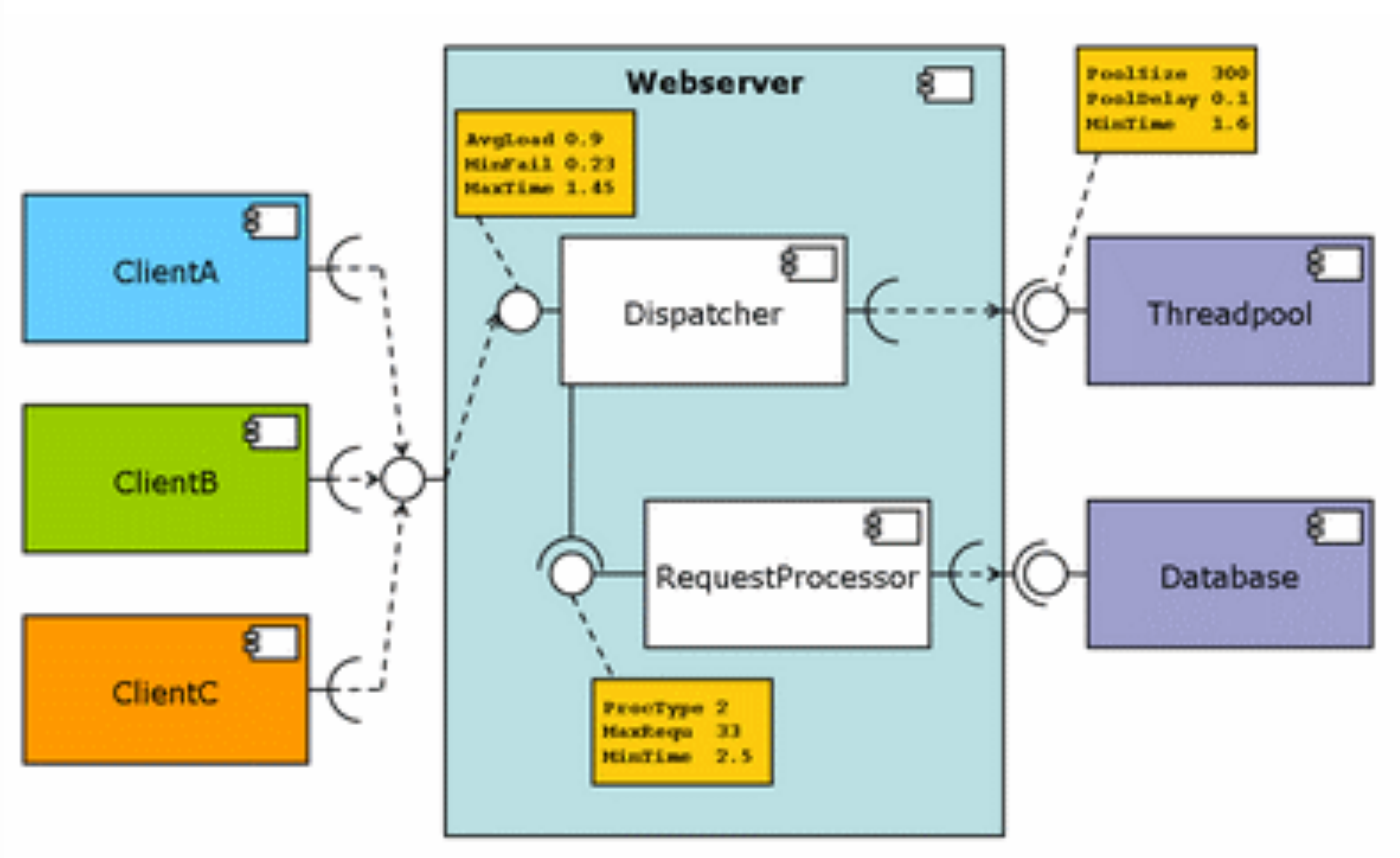


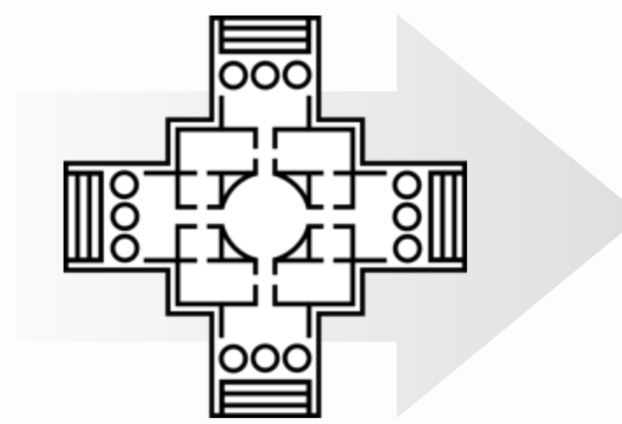



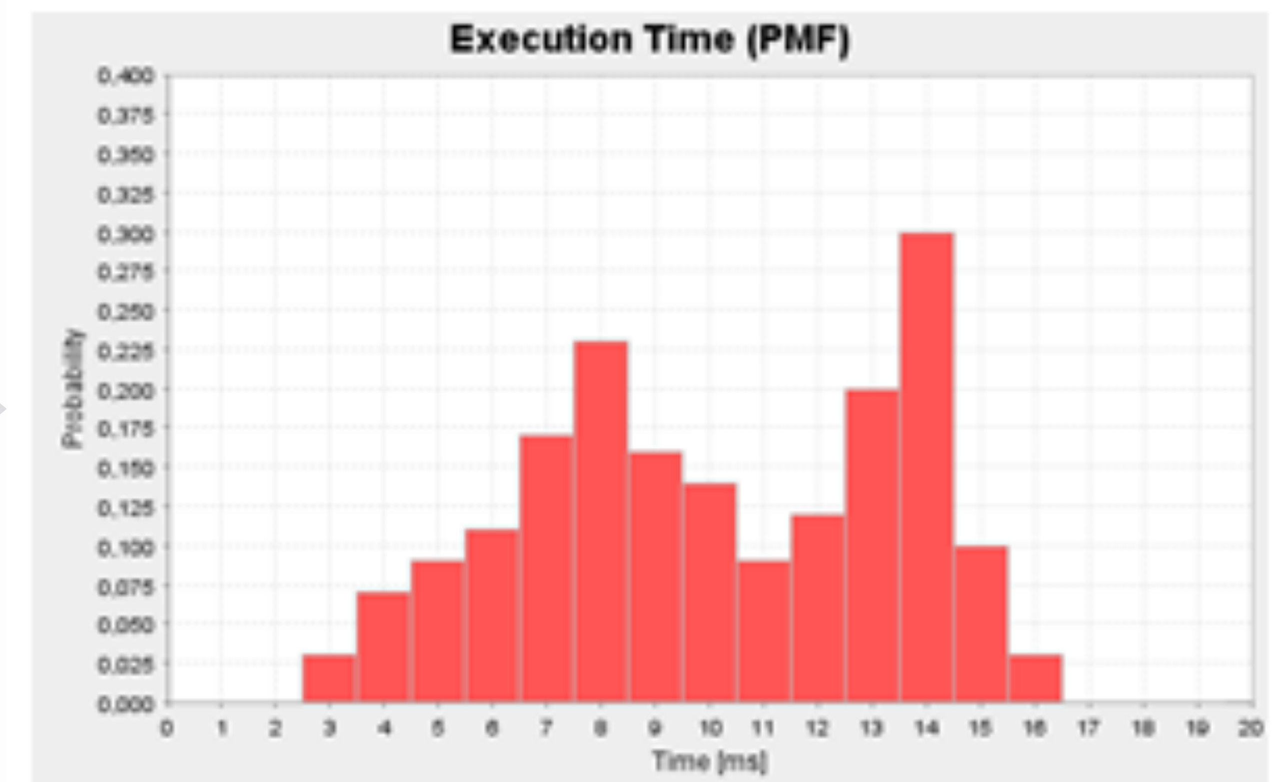


# Model-driven Software Quality Prediction



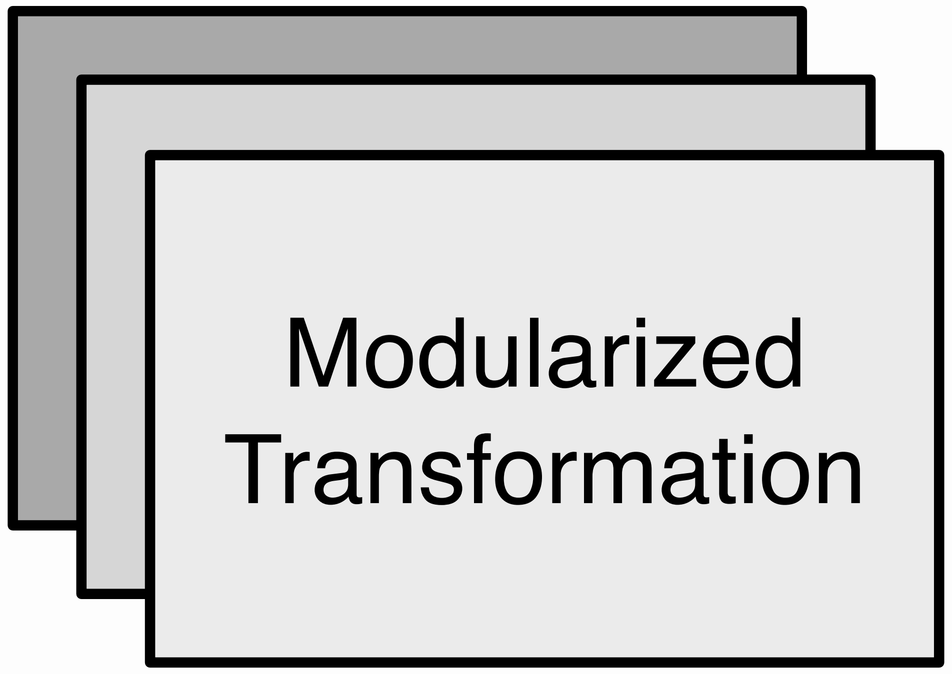



Palladio Component Model → Transformation → Queueing Petri Net Model → Simulate → Measurement Data

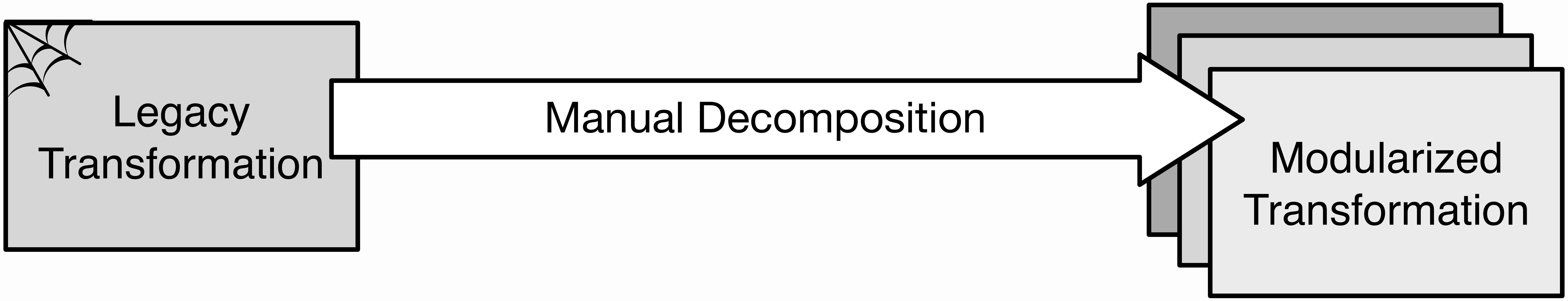
5k lines of **QVT-O** code

# Problem and Overall Approach

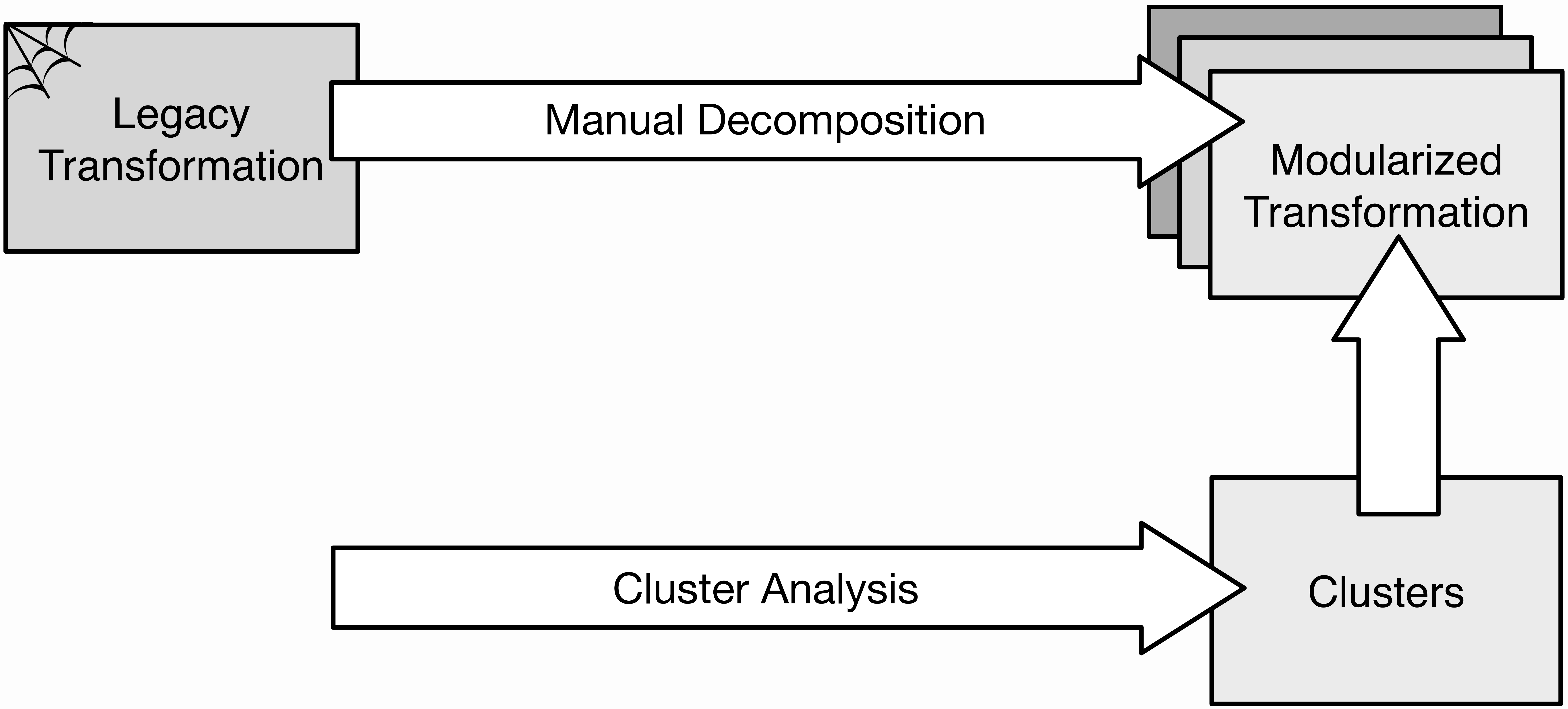
# Problem and Overall Approach

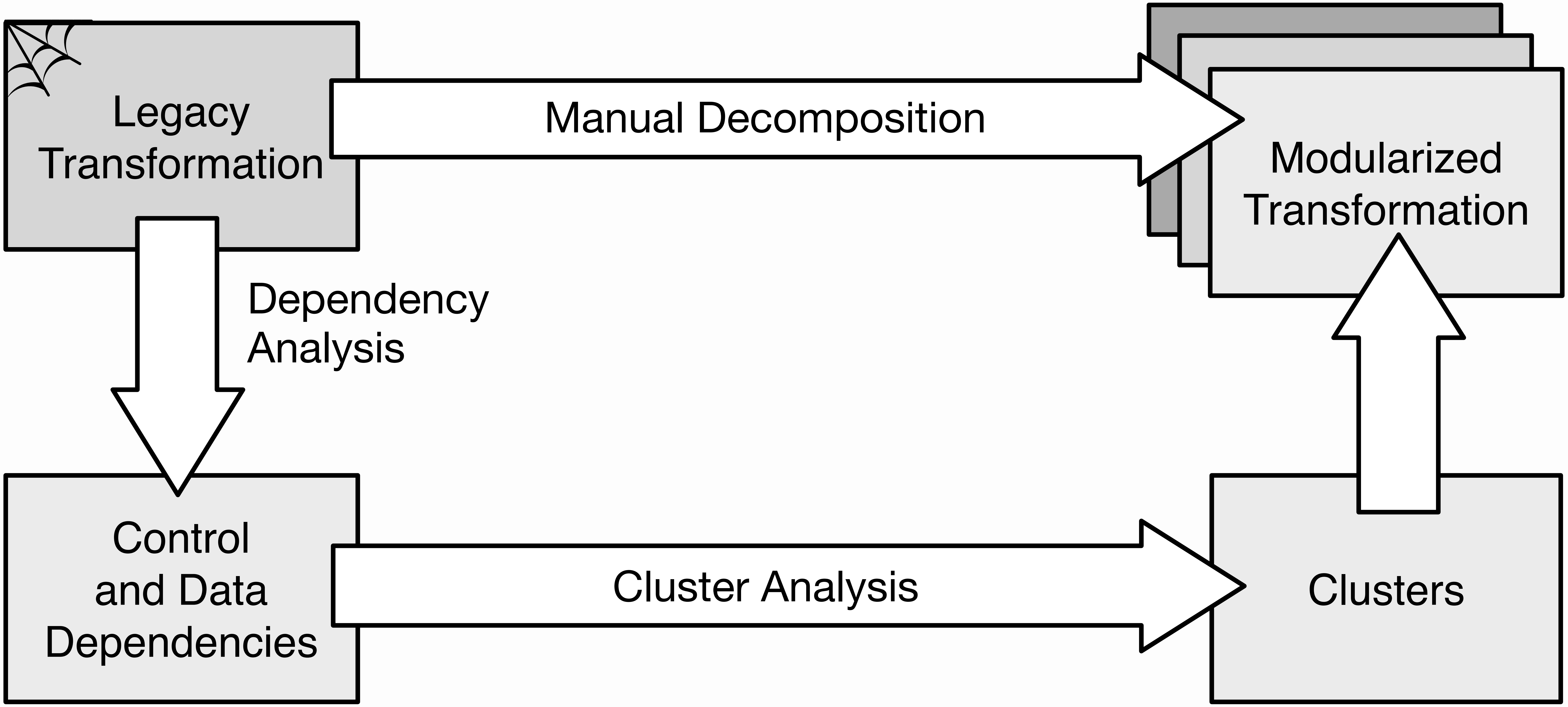
# Problem and Overall Approach



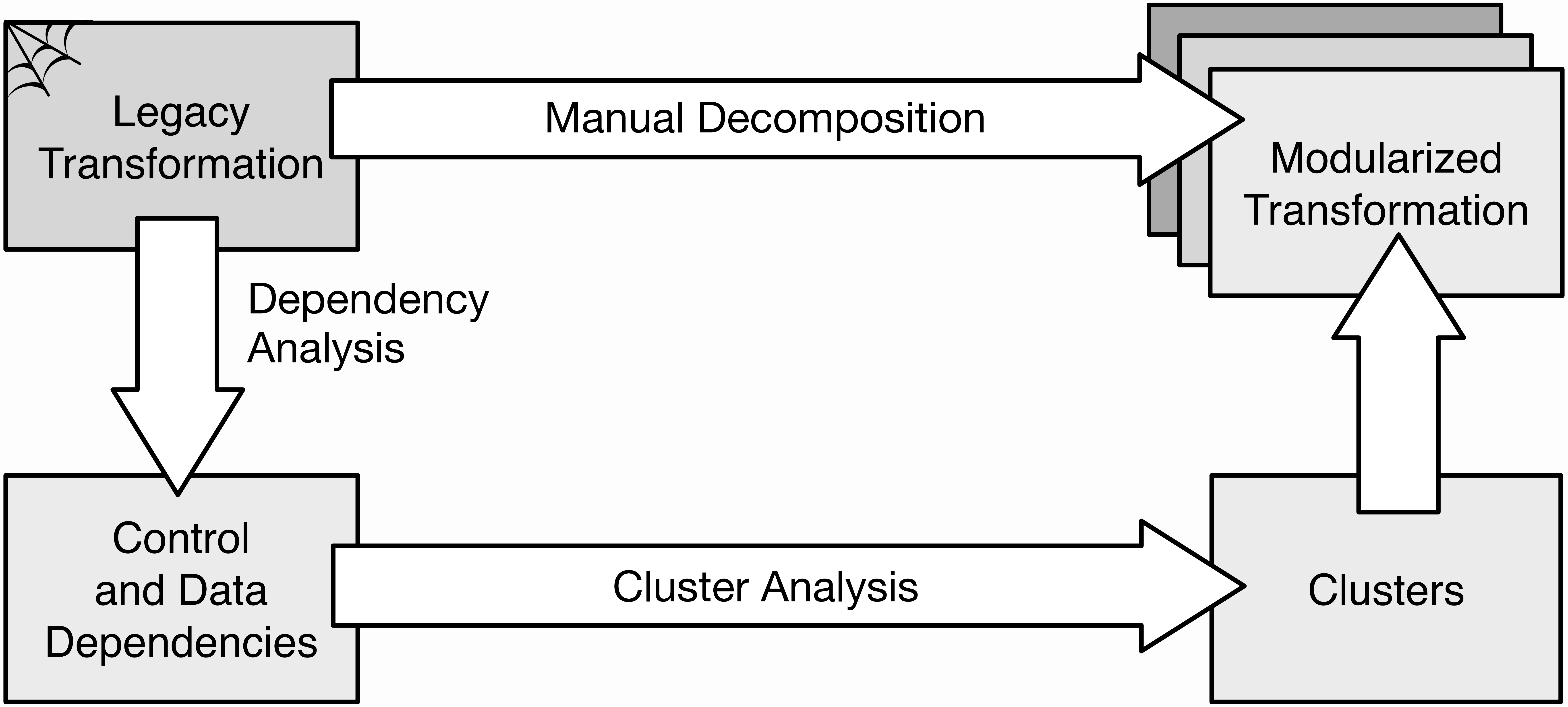
# Problem and Overall Approach



# Problem and Overall Approach

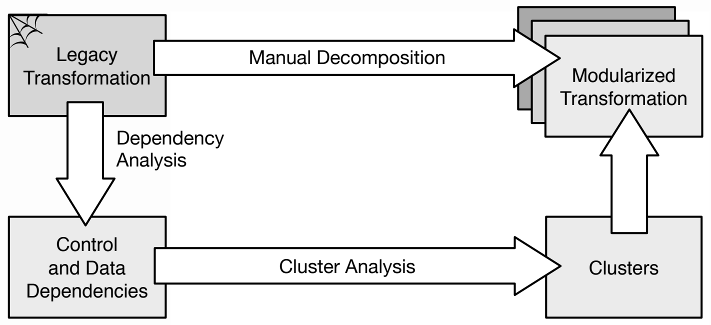


# Problem and Overall Approach

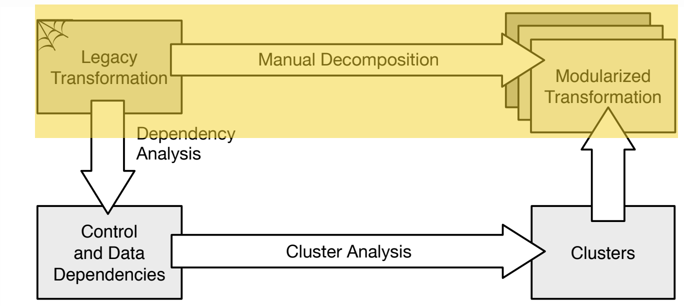



- How can we support typical transformation designs?
- What dependence information is required?

# Problem and Overall Approach




- How can we support typical transformation designs?
- What dependence information is required?



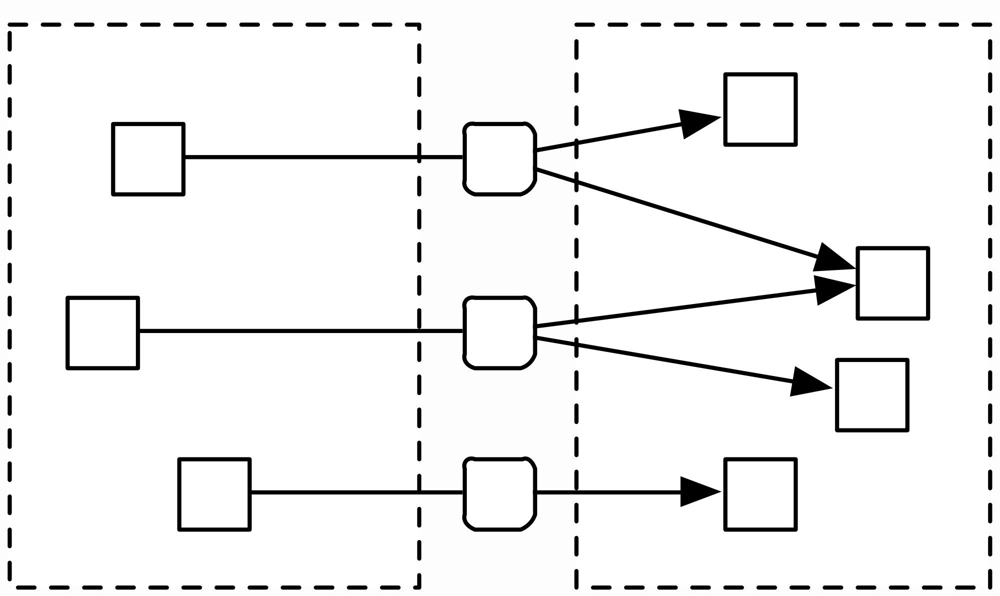

# Design Rules

What’s makes model transformations different from GPL programs?

- Data-centric operations
- Data is hierarchically structured
- Data models extrinsically defined

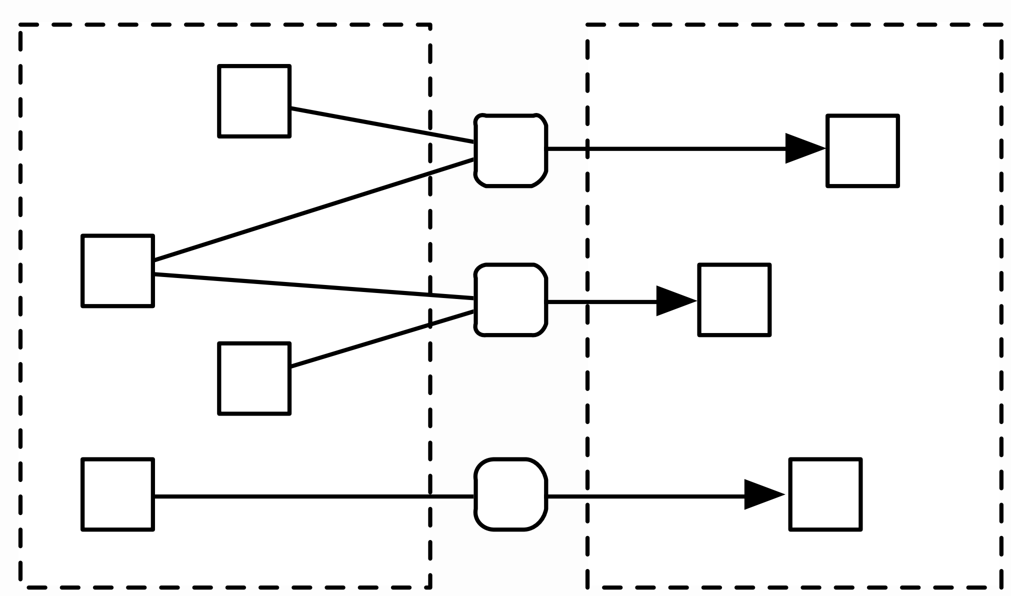
Common decompositional styles [Lawley04]:

Source-driven



one-to-many mappings

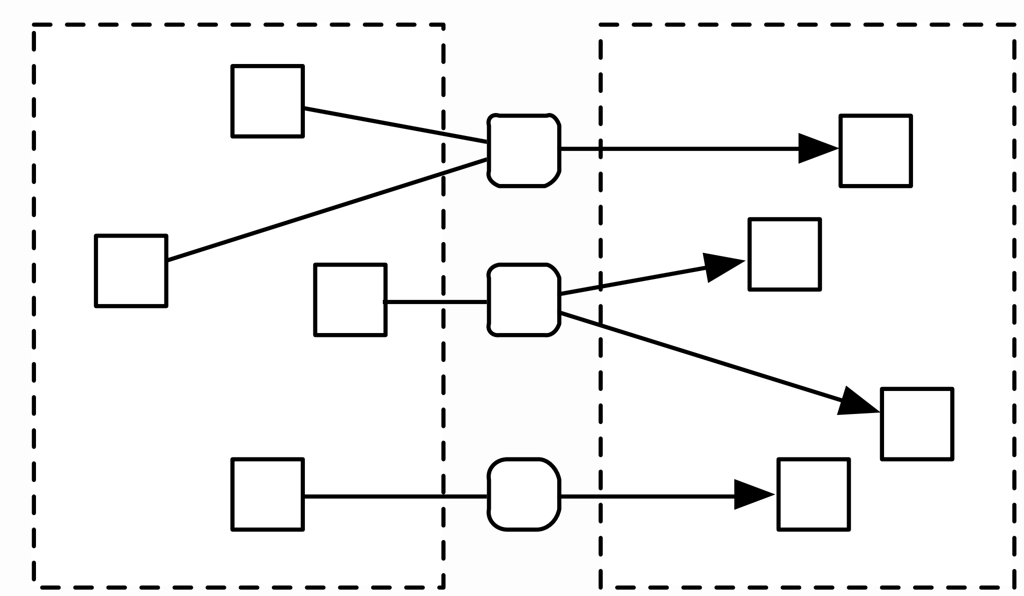
Target-driven



many-to-one mappings,

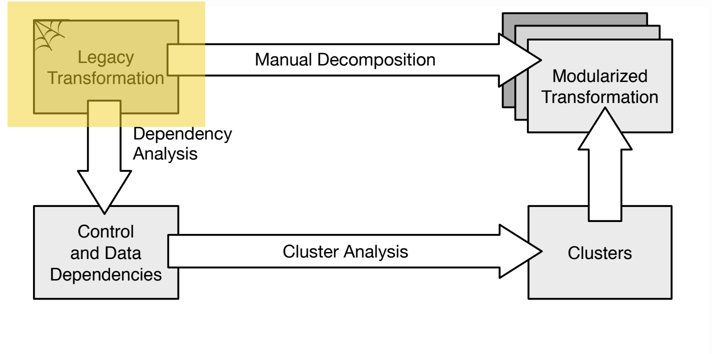
M2T templates

Aspect-driven

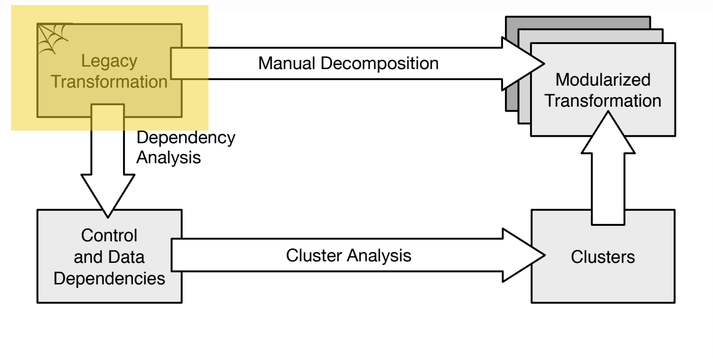


a mixture of both

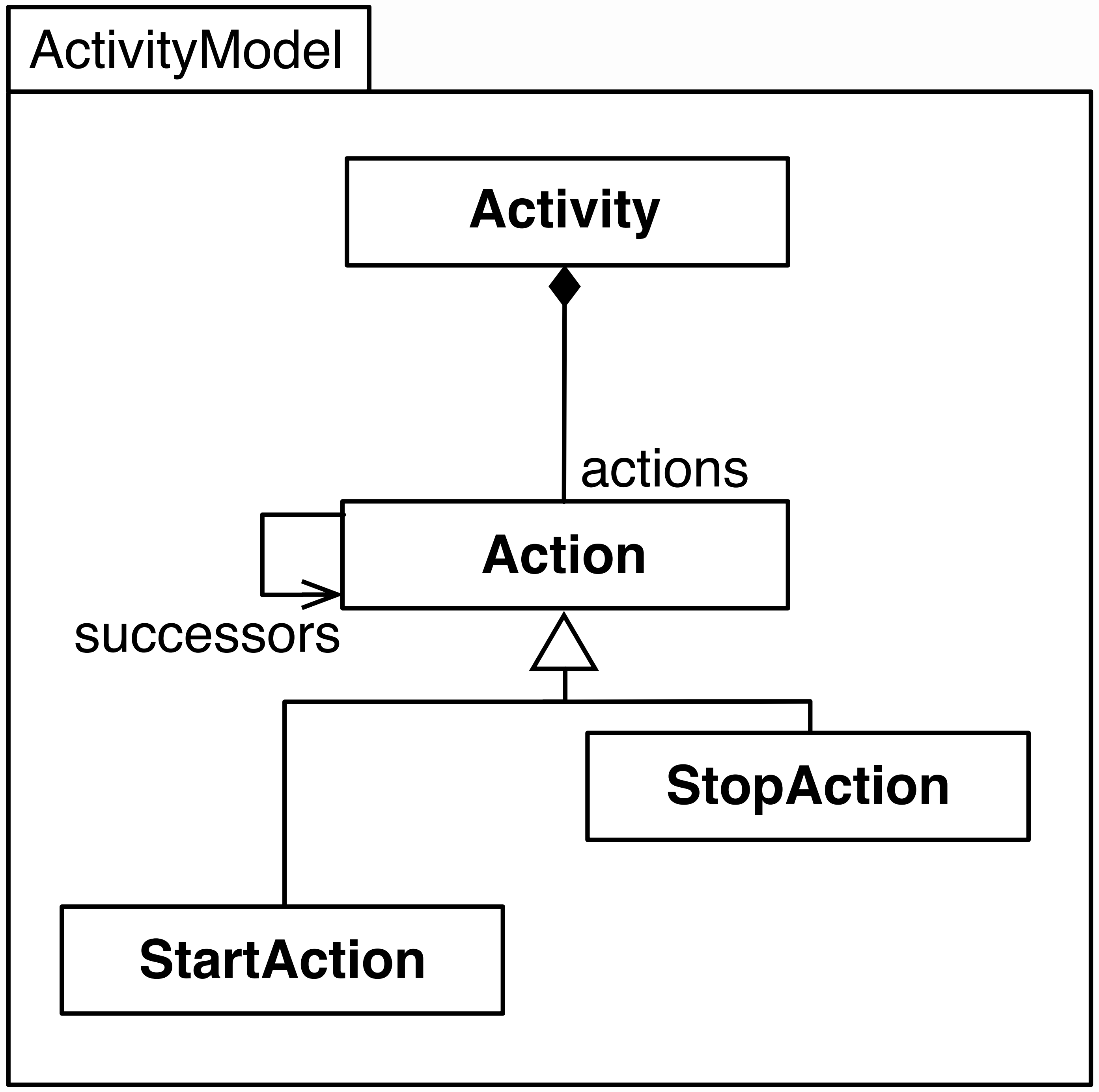
# A Minimalistic Example Transformation



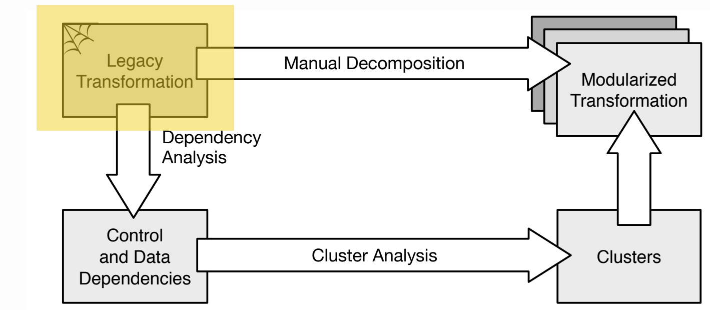
# A Minimalistic Example Transformation



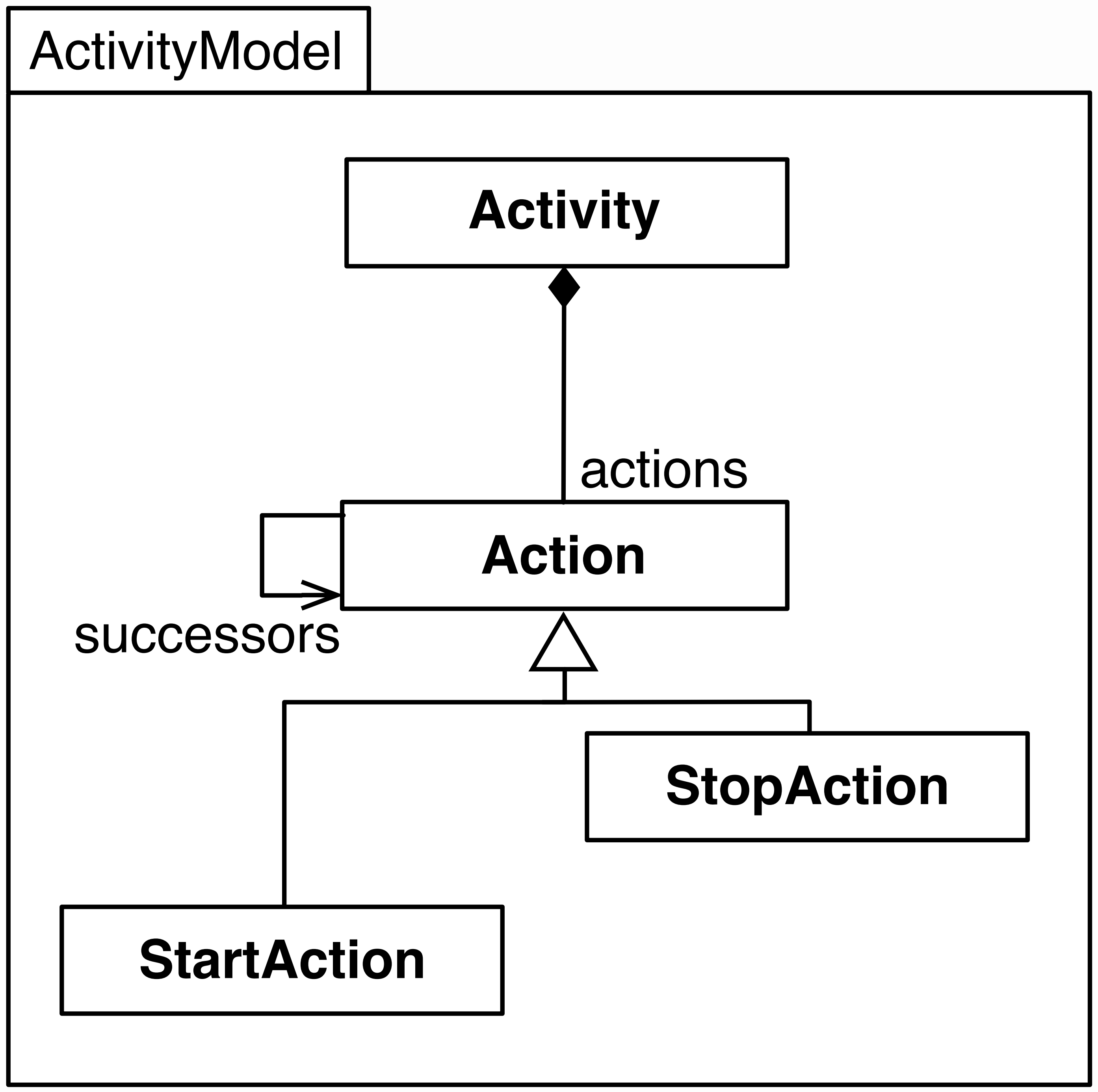

Source Model



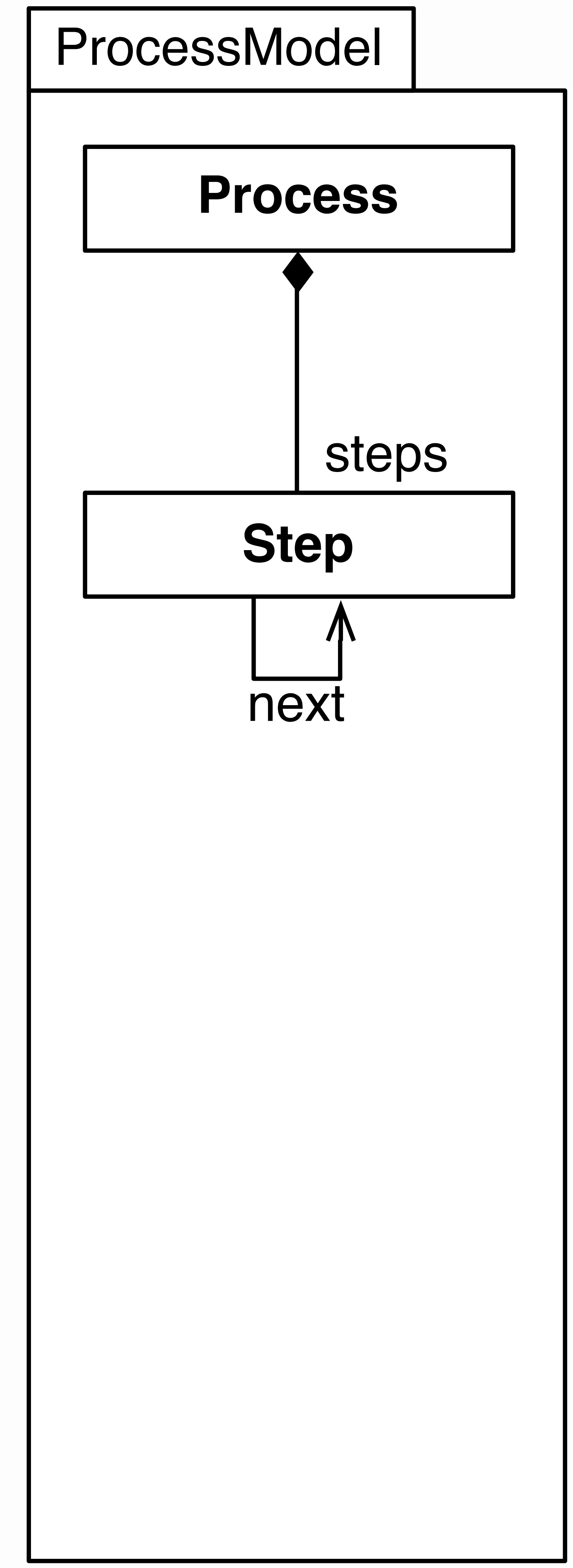
# A Minimalistic Example Transformation



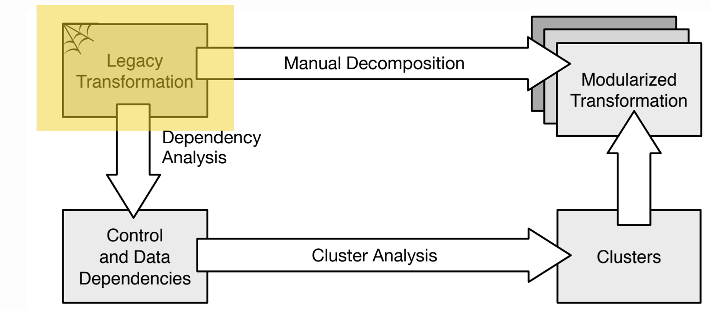

Source Model

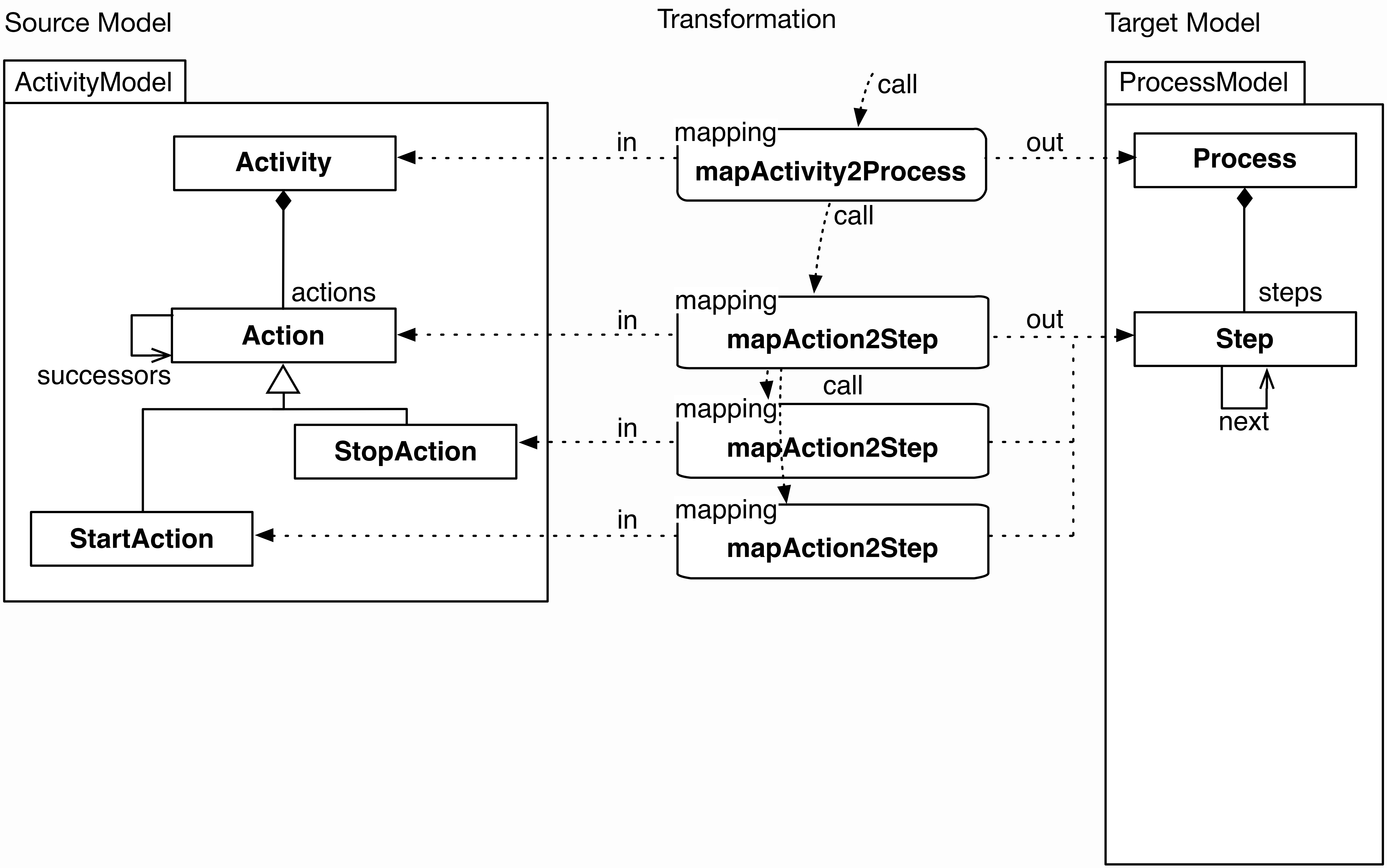



Target Model

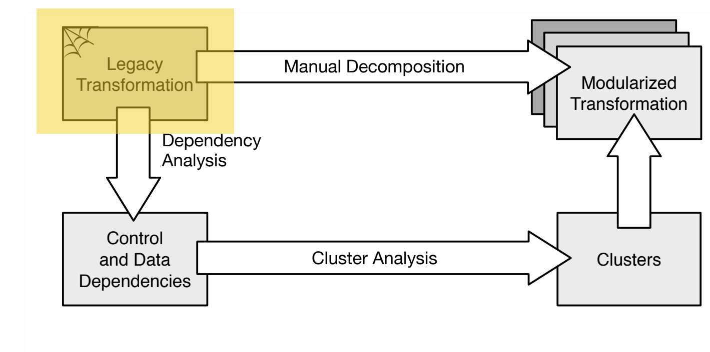


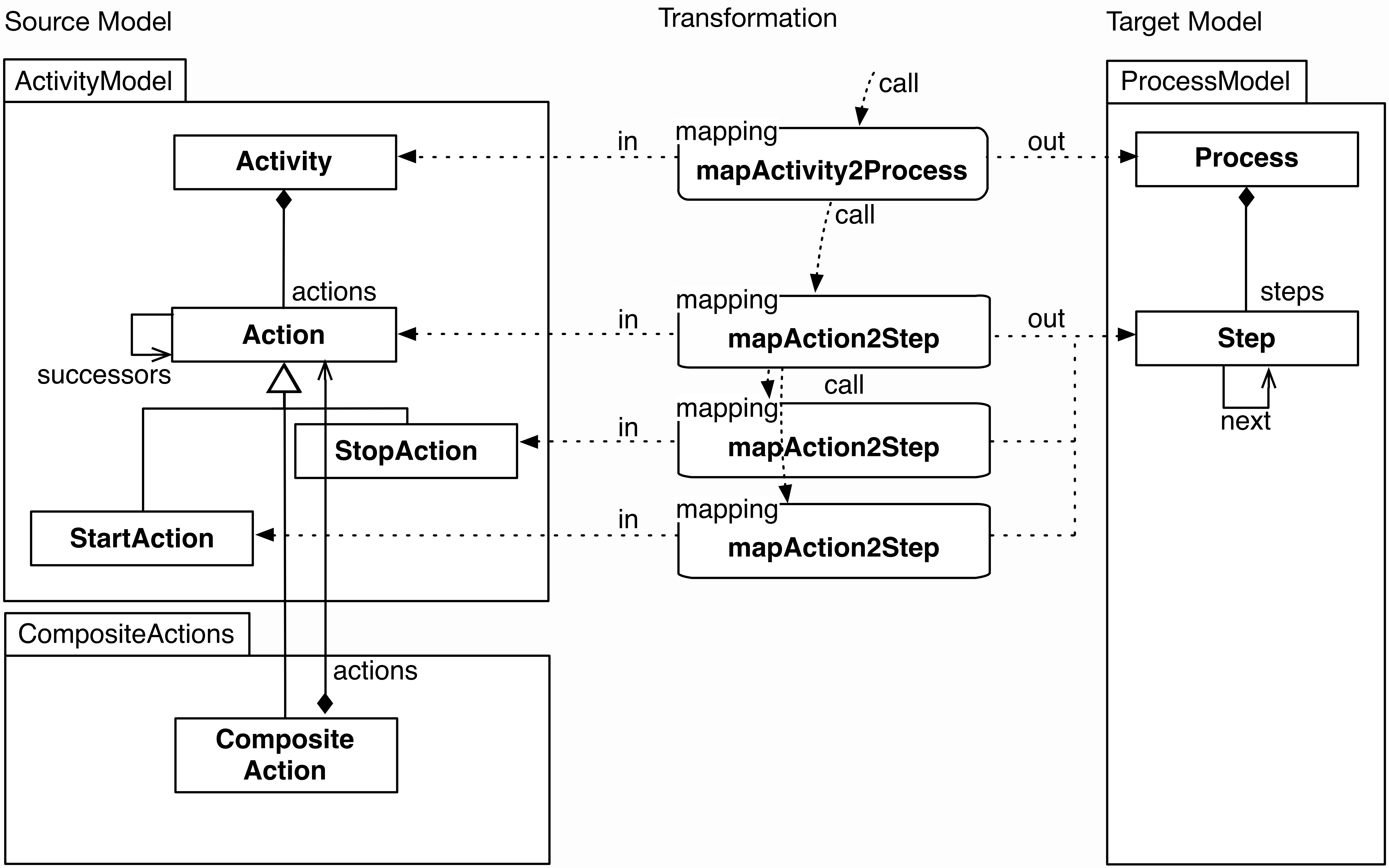
# A Minimalistic Example Transformation



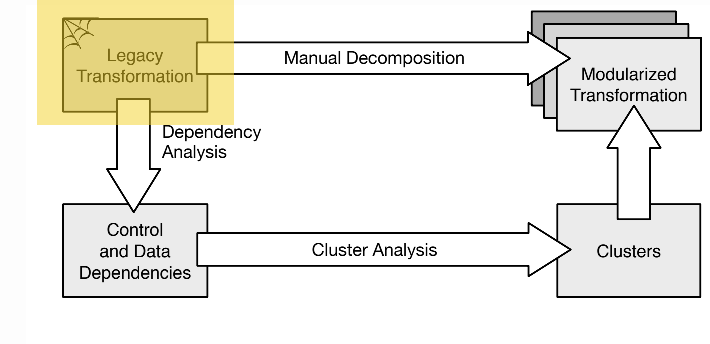



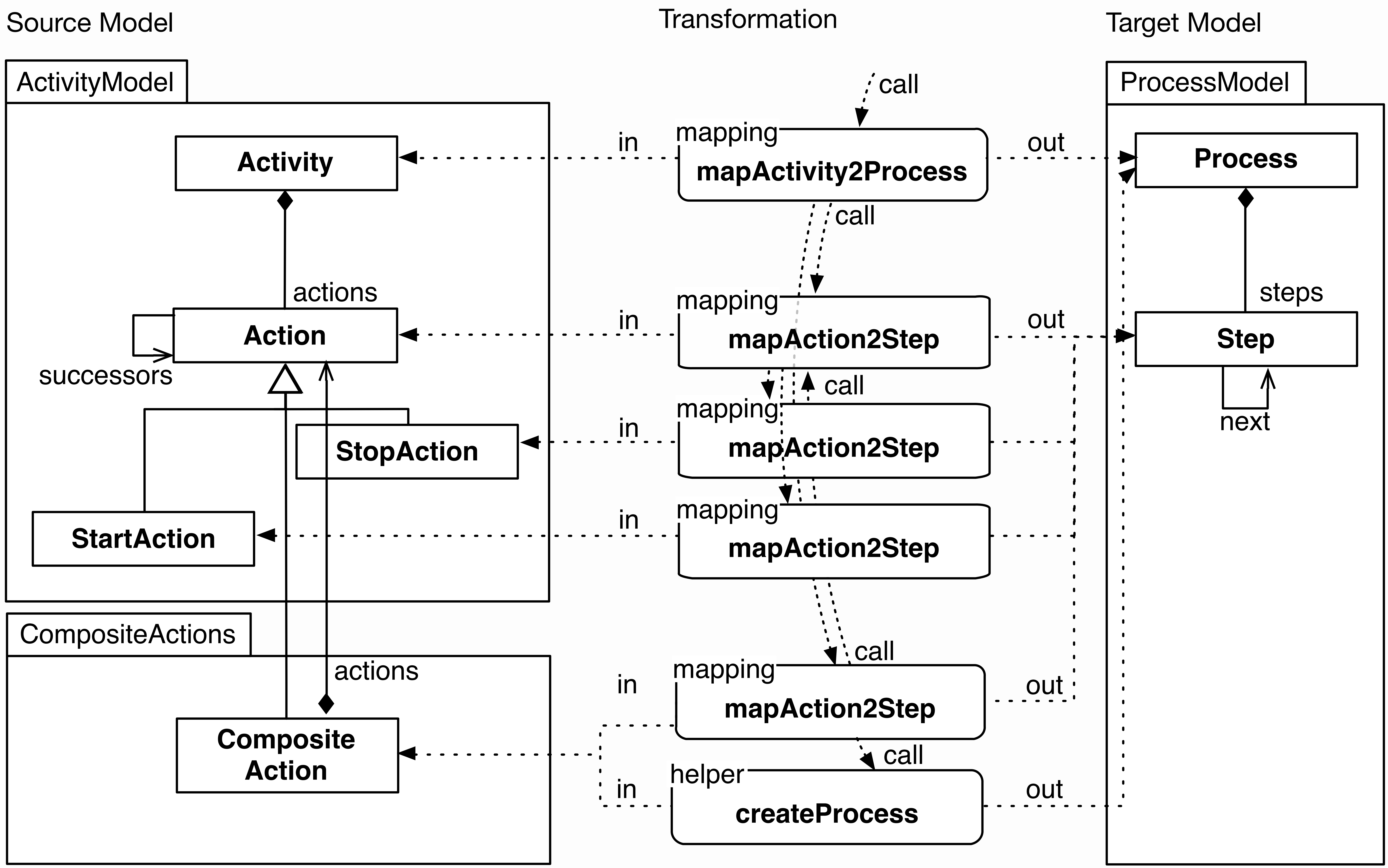
# A Minimalistic Example Transformation



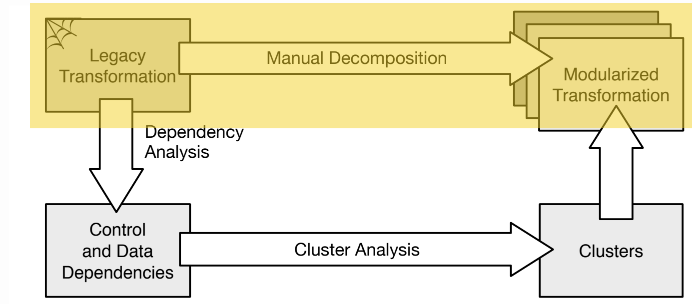



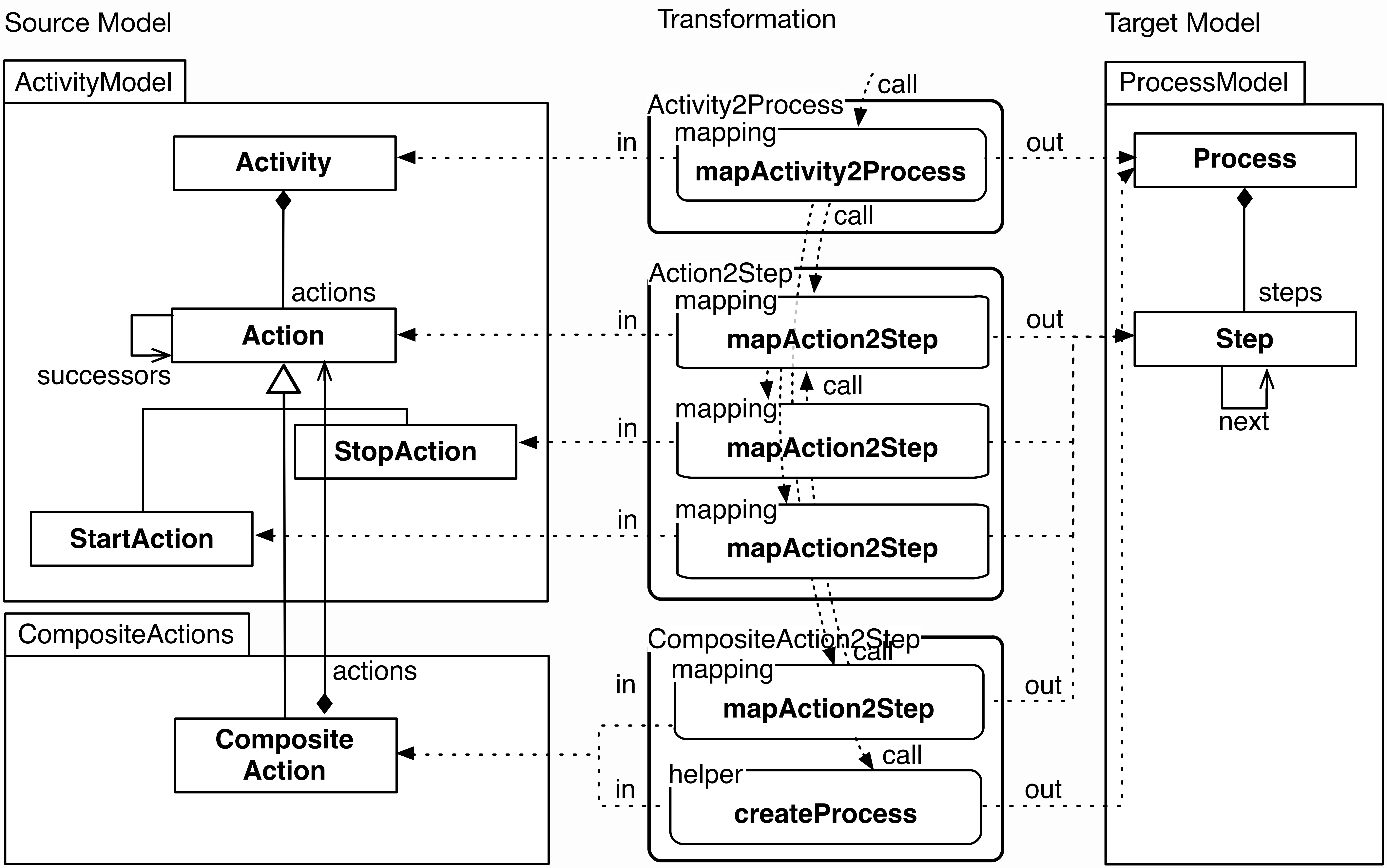
# A Minimalistic Example Transformation



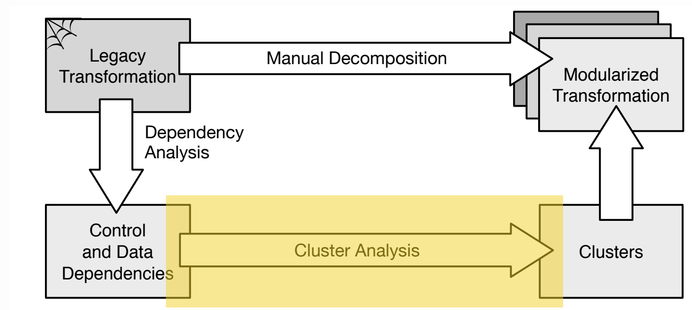



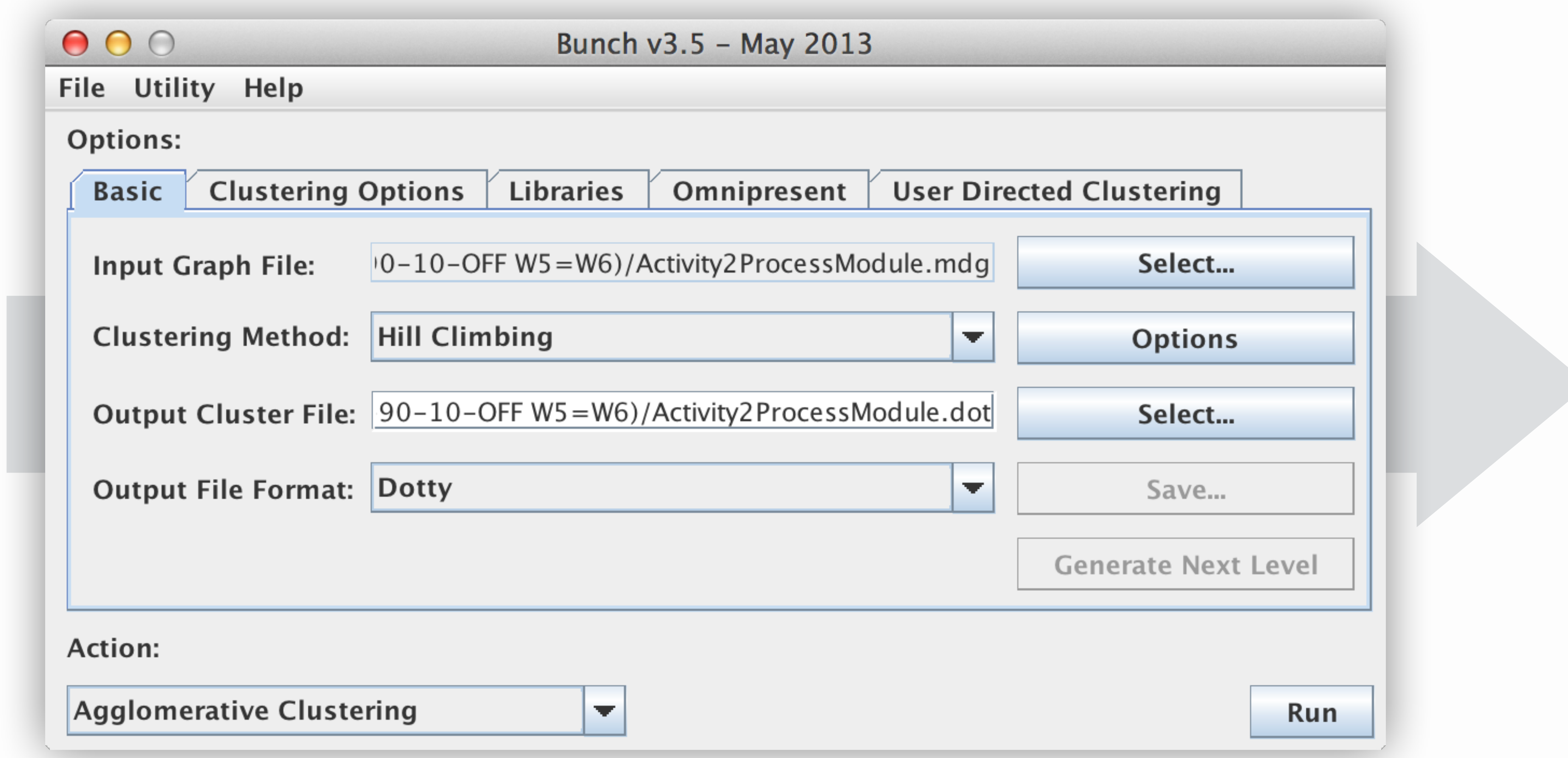
# A Minimalistic Example Transformation



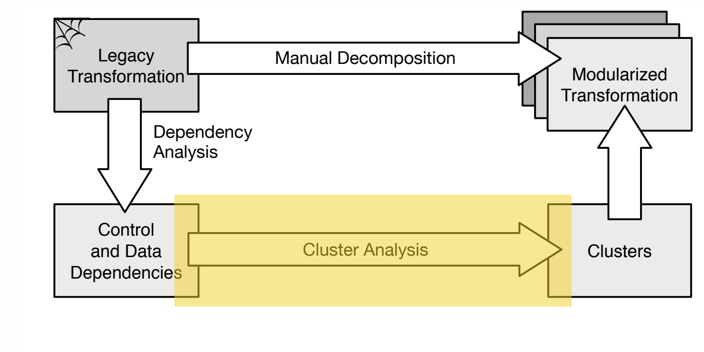
# The *Bunch* Clustering Approach*

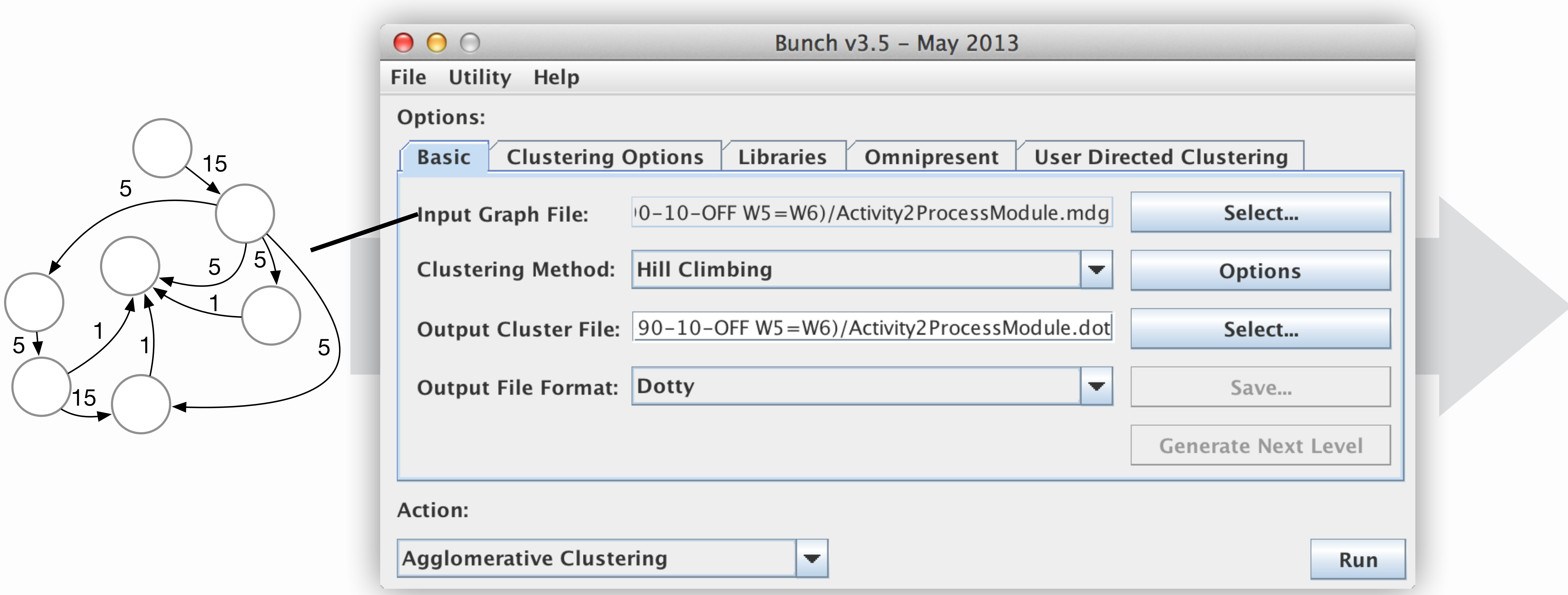


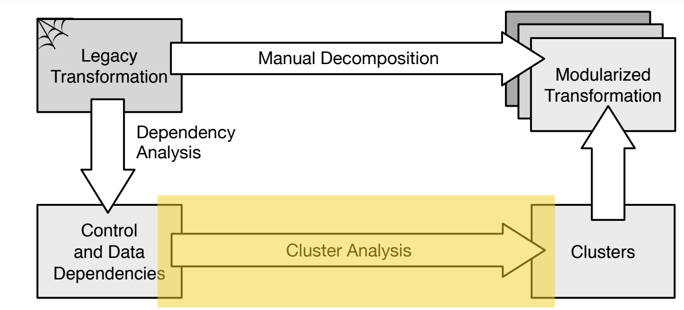

* by [Mitchell06]; alternative approaches are ARCH, ACDC, LIMBO, … [Shtern12]

# The *Bunch* Clustering Approach*

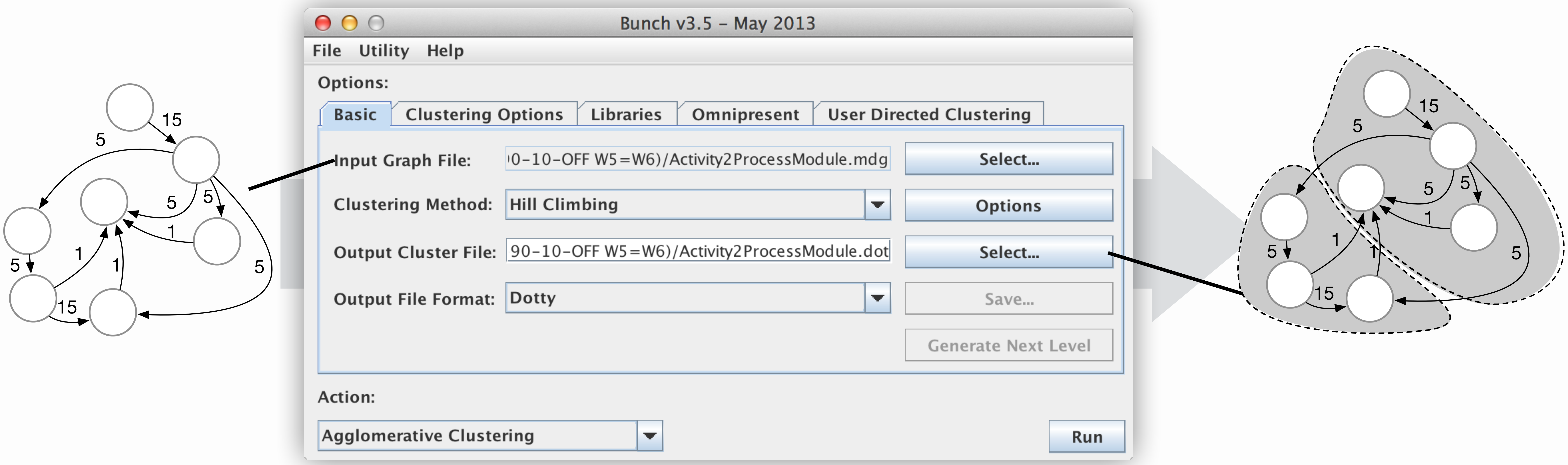



* by [Mitchell06]; alternative approaches are ARCH, ACDC, LIMBO, … [Shtern12]

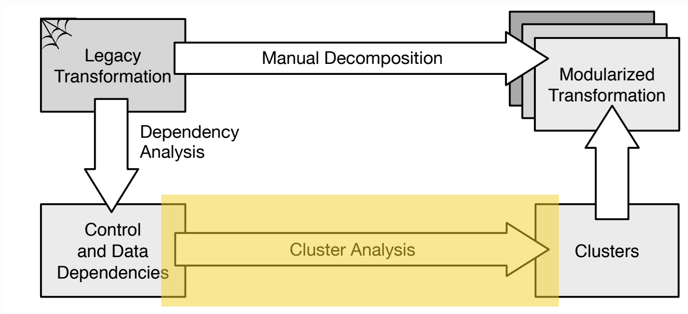



# The *Bunch* Clustering Approach*

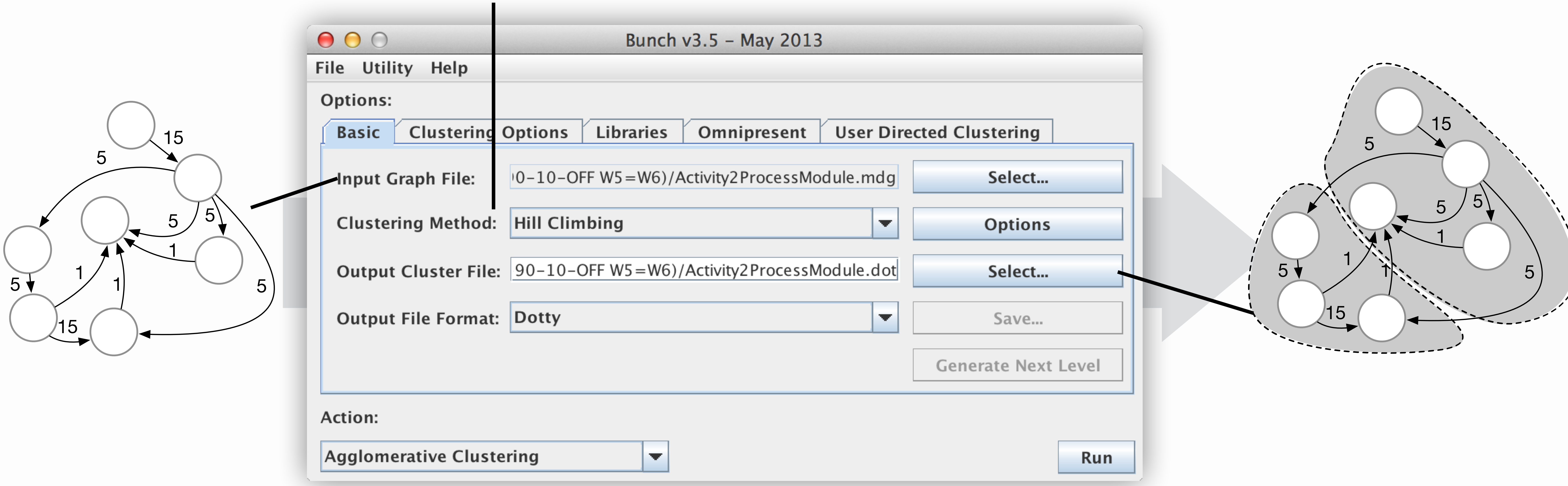



* by [Mitchell06]; alternative approaches are ARCH, ACDC, LIMBO, … [Shtern12]

# The *Bunch* Clustering Approach*

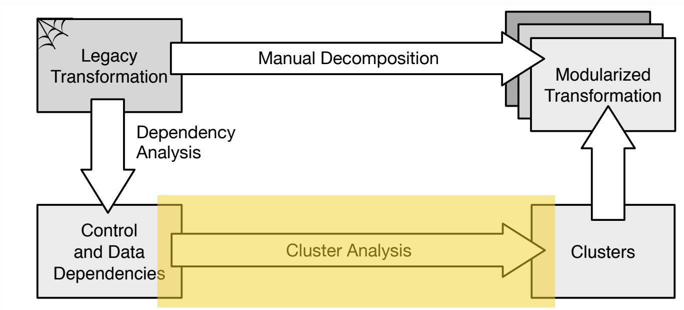



Clustering methods: hill climbing, genetic, exhaustive

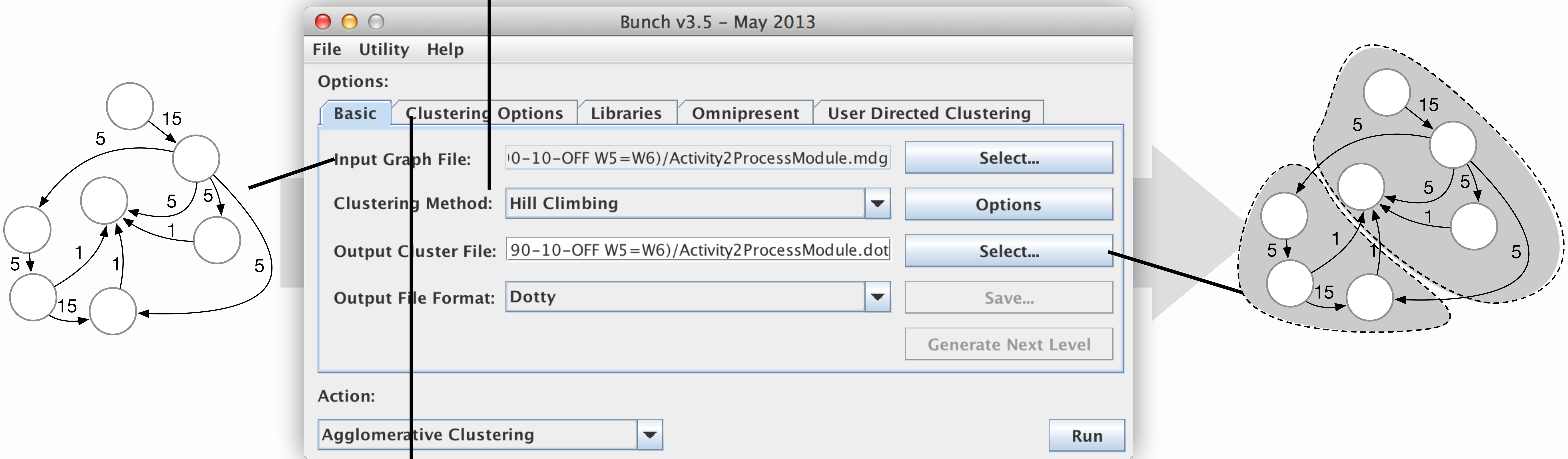



* by [Mitchell06]; alternative approaches are ARCH, ACDC, LIMBO, … [Shtern12]

# The *Bunch* Clustering Approach*




Clustering methods: hill climbing, genetic, exhaustive

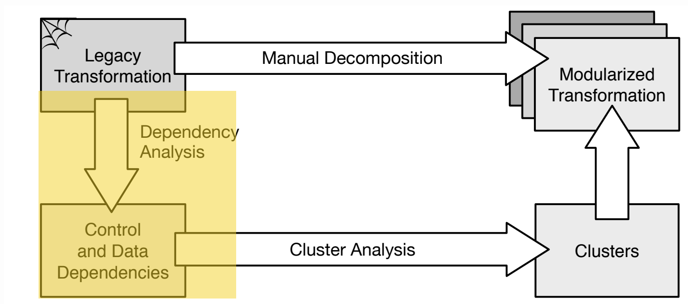


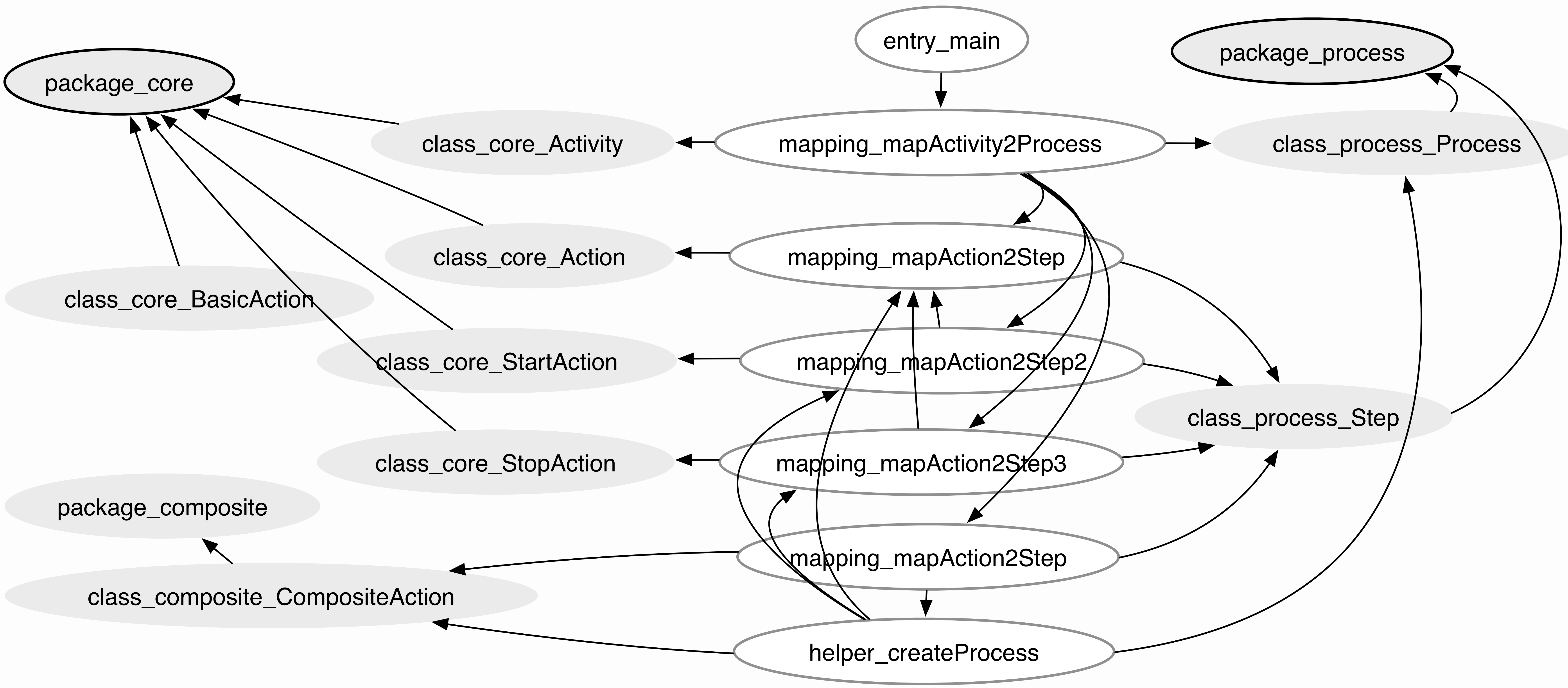

Modularity quality index: favor low coupling & high cohesion

$$MQ = \sum_{i=1}^{k} CF_i, \quad CF_i = \begin{cases} 0, & \mu_i = 0 \\ \frac{\mu_i}{\mu_i + \frac{1}{2}\sum_{\substack{j=1 \\ j \neq i}}^{k} (\epsilon_{i,j} + \epsilon_{j,i})}, & \text{otherwise} \end{cases}$$

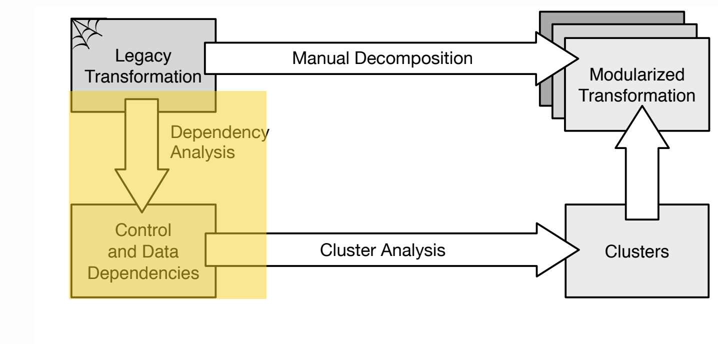
* by [Mitchell06]; alternative approaches are ARCH, ACDC, LIMBO, … [Shtern12]

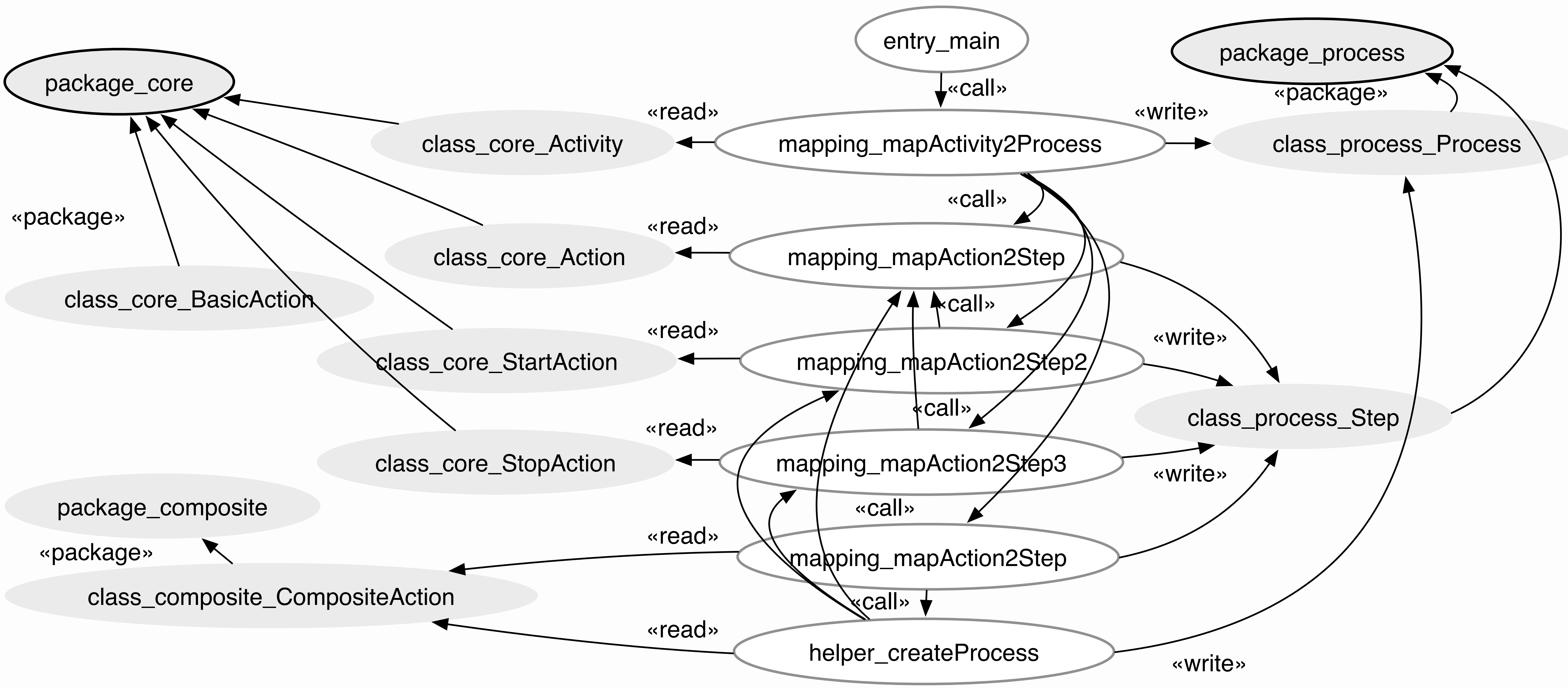
# Dependence Analysis

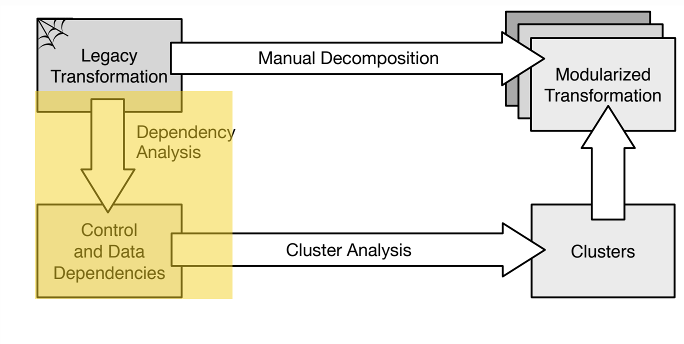


# Dependence Analysis



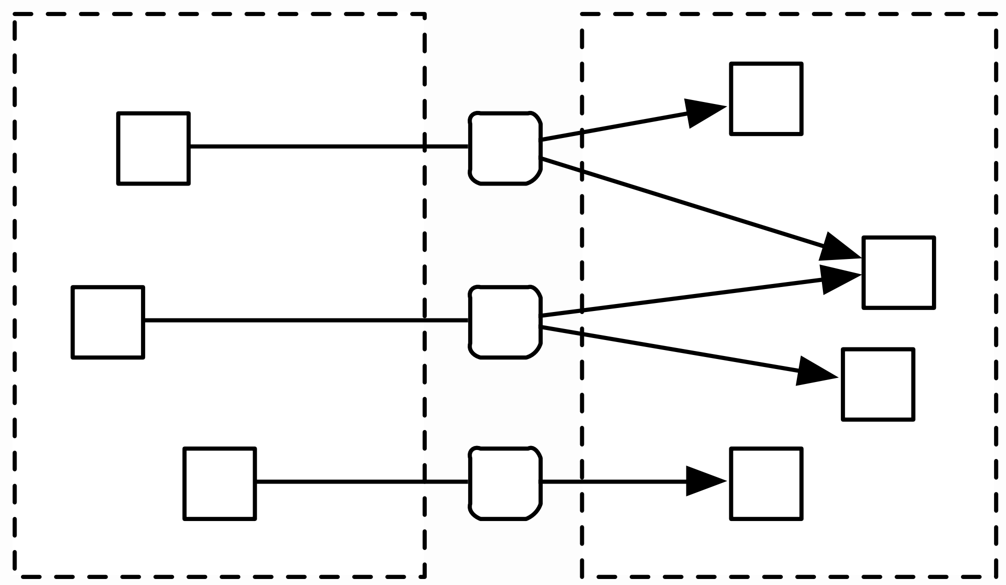





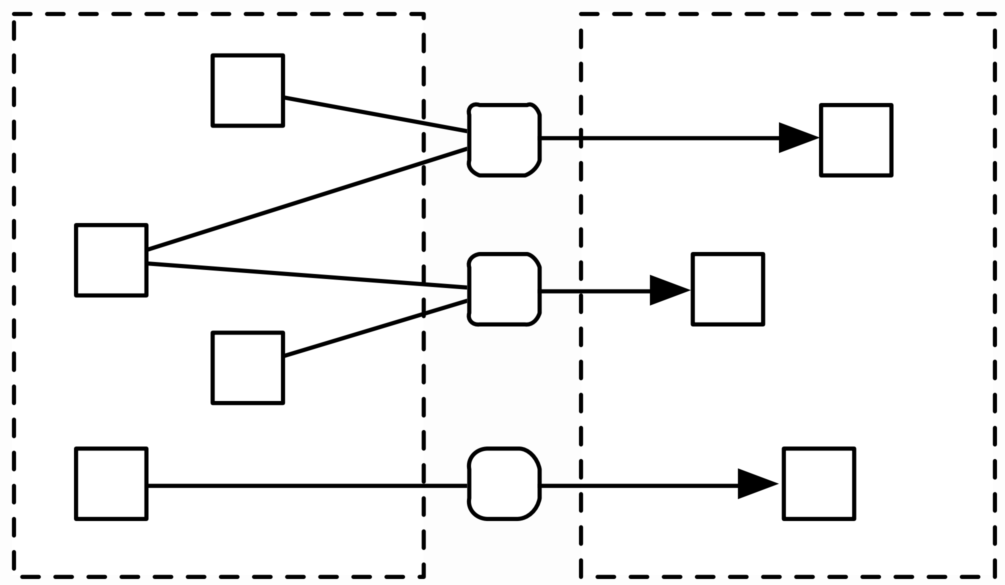

# Weight Configuration

Depending on the style…

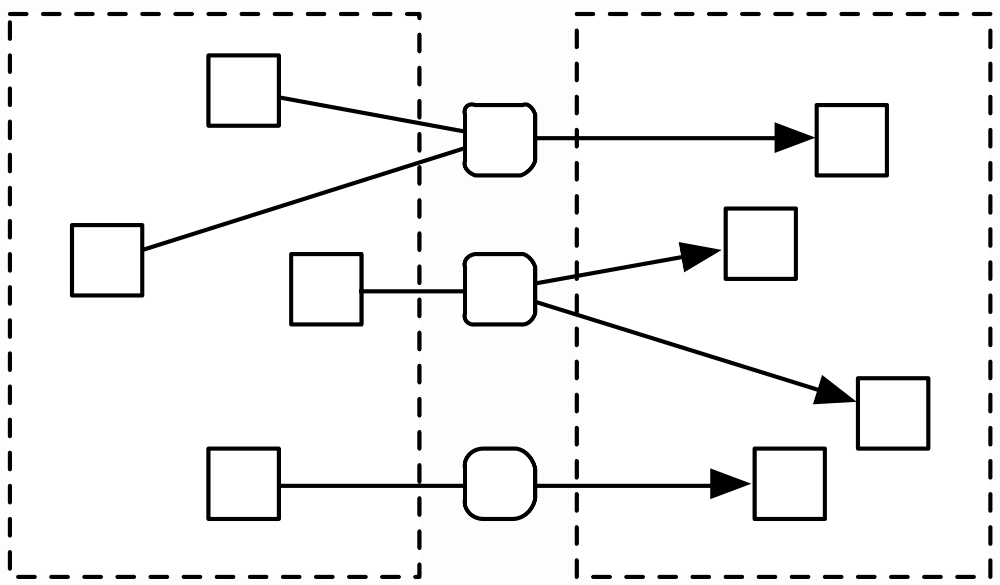
Source-driven



Target-driven



Aspect-driven



…initialize a configuration vector for weighting the various types of edges:

$$WC := \langle W_{write}, W_{read}, W_{call}, W_{package} \rangle \in \mathbb{N}_0^4$$

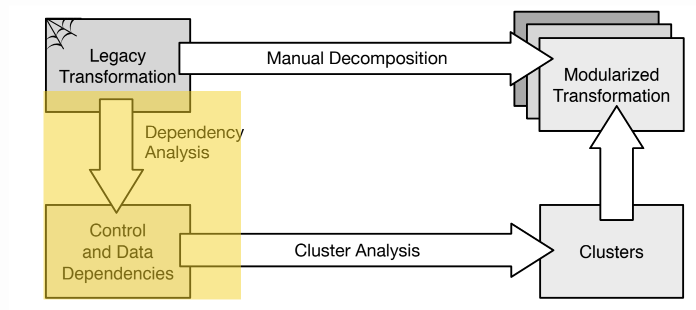
Model element is contained in package

Method invokes another method

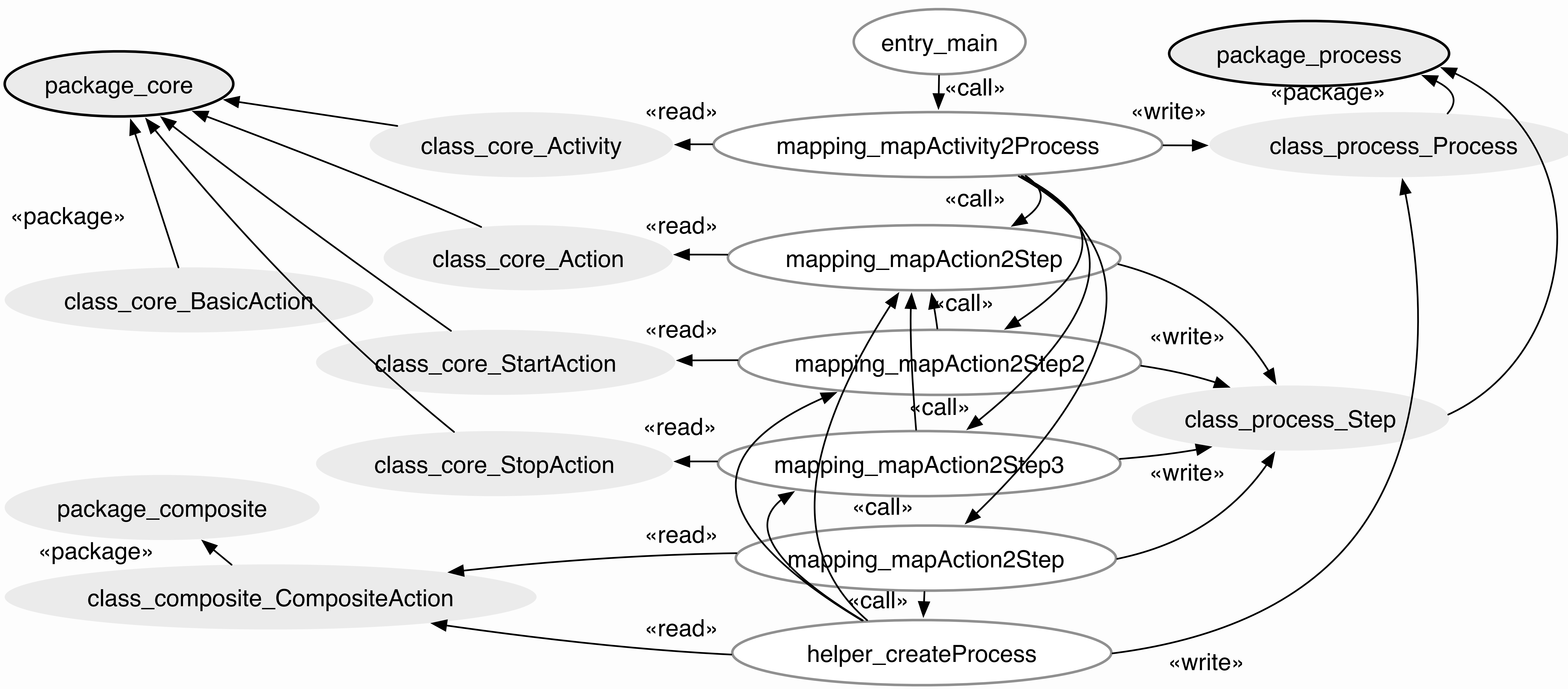
Method reads model element

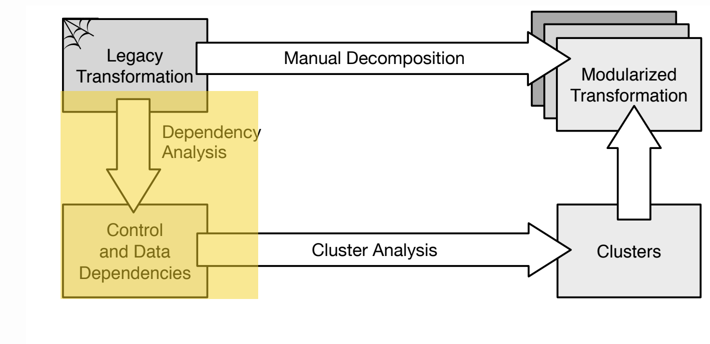
Method creates/updates model element

# Assigning Weights



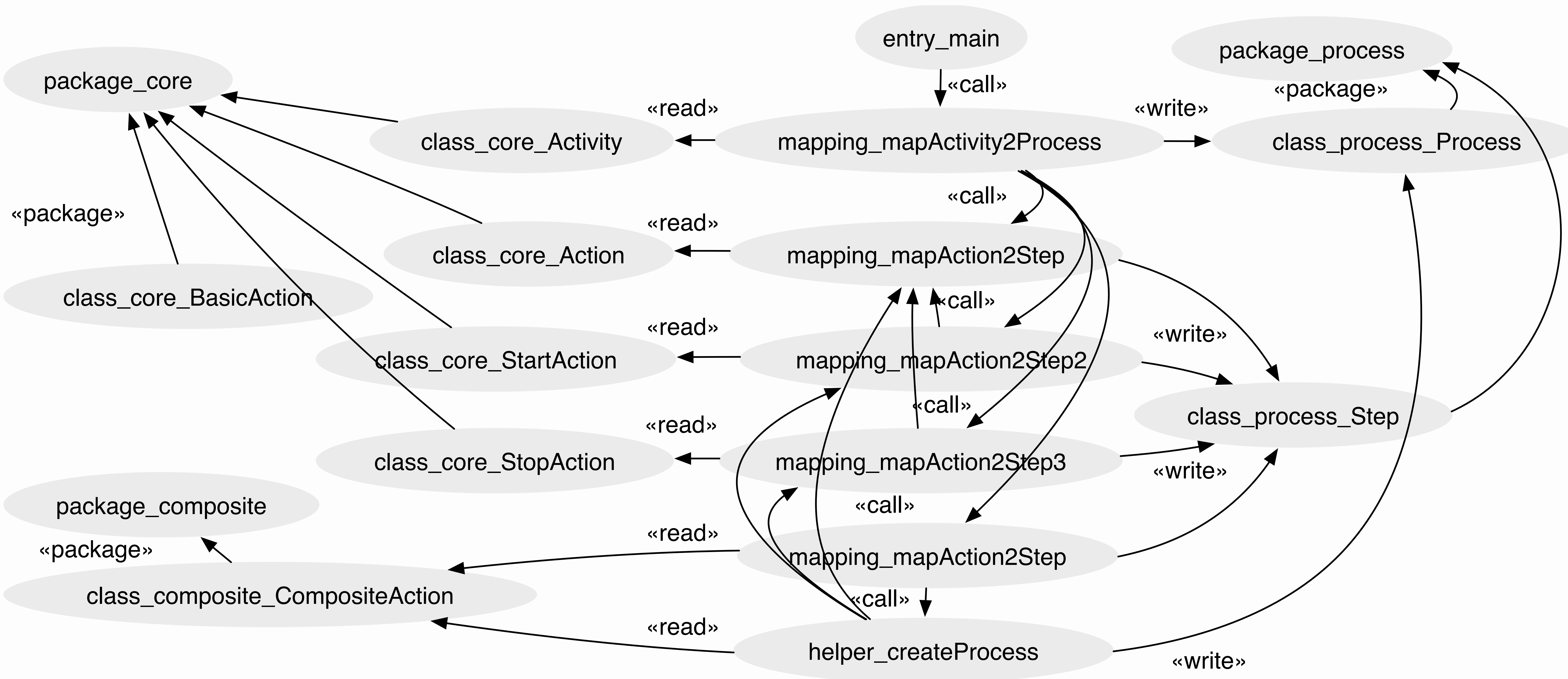

$$
\begin{aligned}
WC &:= \langle W_{write}, W_{read}, W_{call}, W_{package} \rangle \\
&= \langle 1, 15, 5, 15 \rangle
\end{aligned}
$$



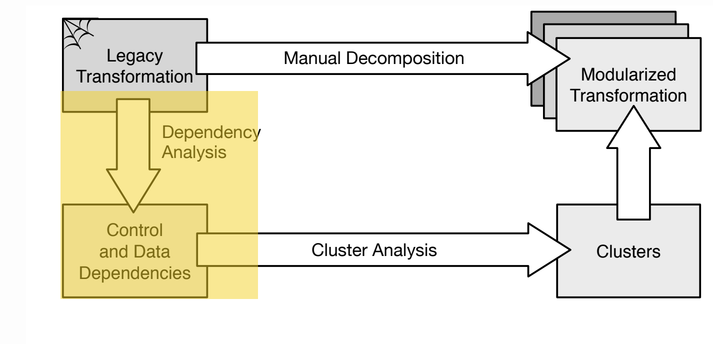



# Assigning Weights

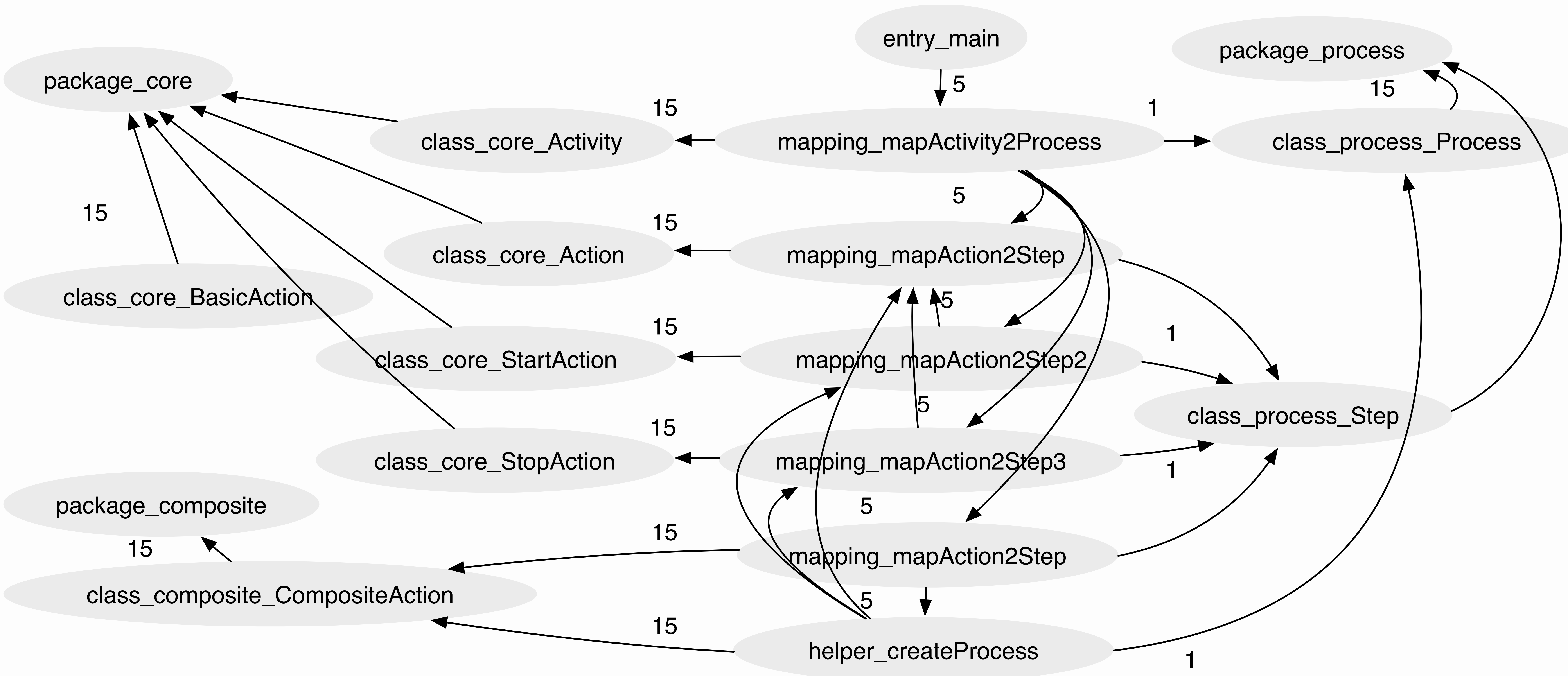
$$WC := \langle W_{write}, W_{read}, W_{call}, W_{package} \rangle$$
$$= \langle 1, 15, 5, 15 \rangle$$



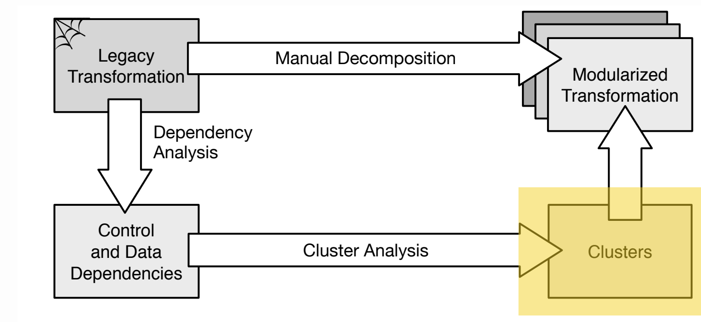
# Assigning Weights

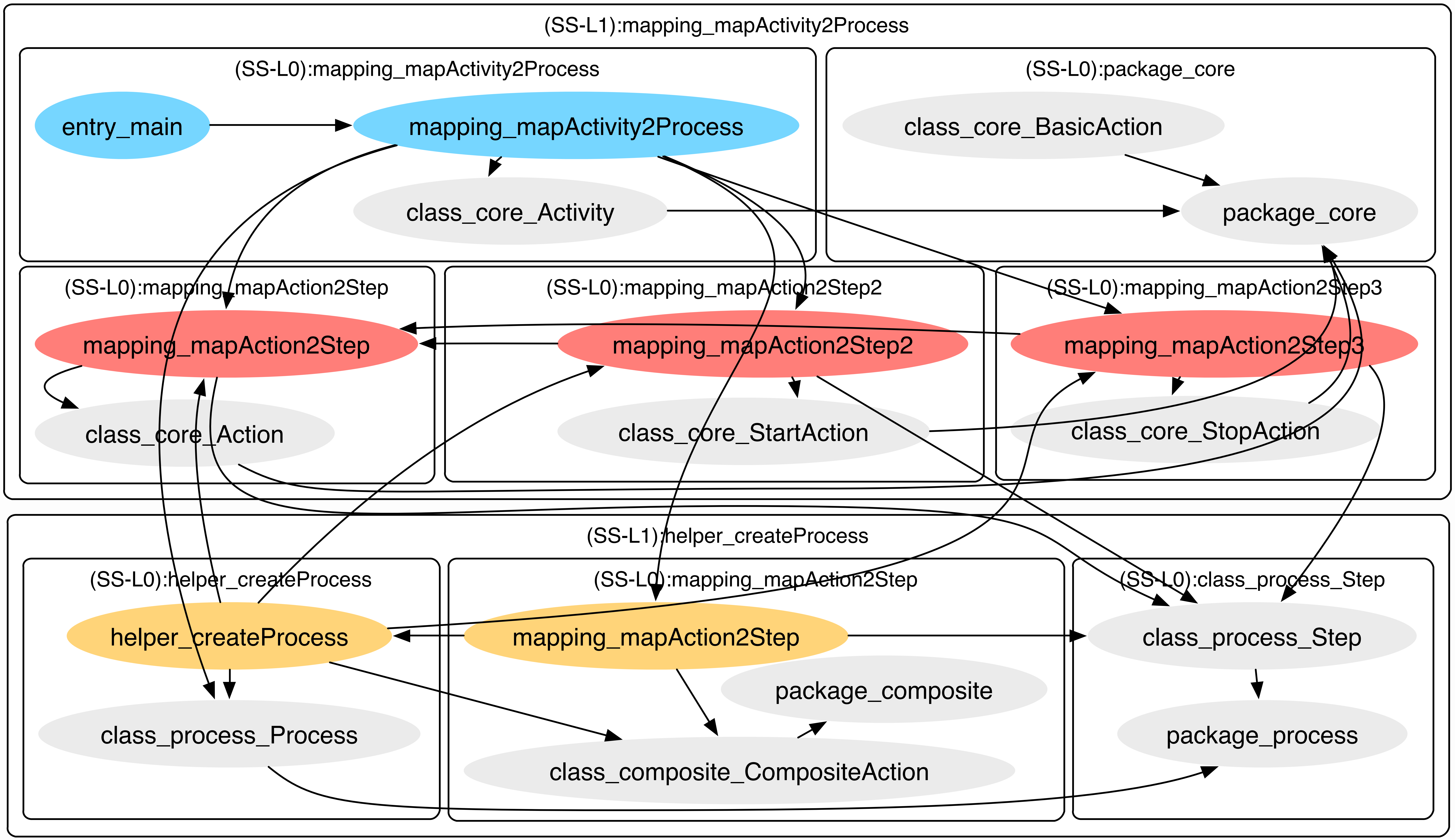



$$
\begin{aligned}
WC &:= \langle W_{write}, W_{read}, W_{call}, W_{package} \rangle \\
&= \langle 1, 15, 5, 15 \rangle
\end{aligned}
$$

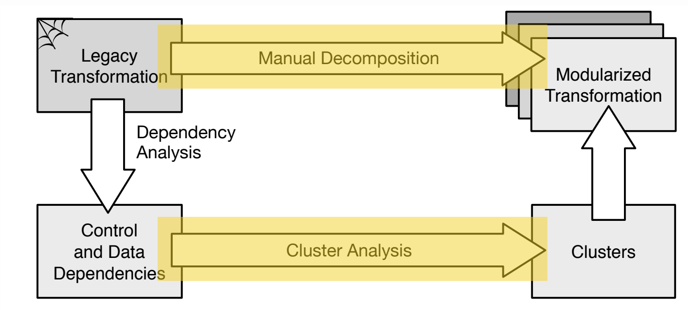


# Example Clustering



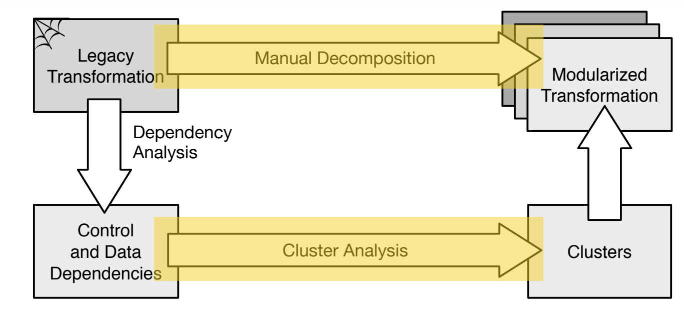



# Assessing Results




- Compare derived clustering with a manual expert clustering
- Using three similarity/distance metrics

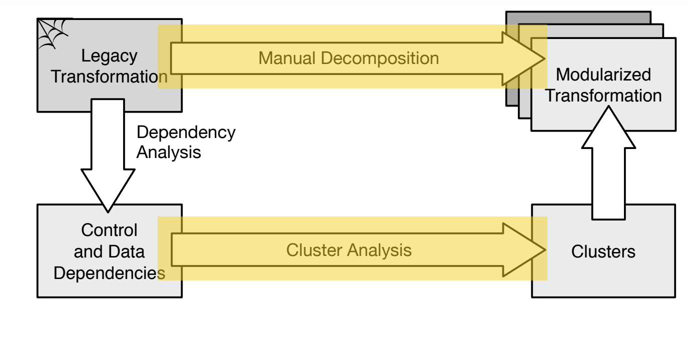
# Assessing Results




- Compare derived clustering with a manual expert clustering
- Using three similarity/distance metrics

| Configuration | Statistics | | Similarity to expert clustering | | | |
|---|---|---|---|---|---|---|
| | # Clusters | MQ index | Precision | Recall | EdgeSim | MeCl |
| **Expert clustering**<br>Derived manually | 3 | 1.067 | 100% | 100% | 100 | 100% |

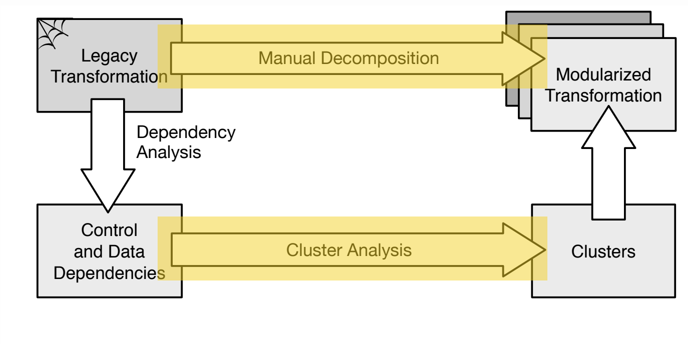
# Assessing Results




- Compare derived clustering with a manual expert clustering
- Using three similarity/distance metrics

| Configuration | Statistics | | Similarity to expert clustering | | | |
|---|---|---|---|---|---|---|
| | # Clusters | MQ index | Precision | Recall | EdgeSim | MeCl |
| **Expert clustering**<br>Derived manually | 3 | 1.067 | 100% | 100% | 100 | 100% |
| **Method-call dependencies only**<br>Hill Climbing, $WC = \langle 0, 0, 1, 0 \rangle$ | 2 | 1.214 | 20.00% | 100% | 54.54 | 60% |

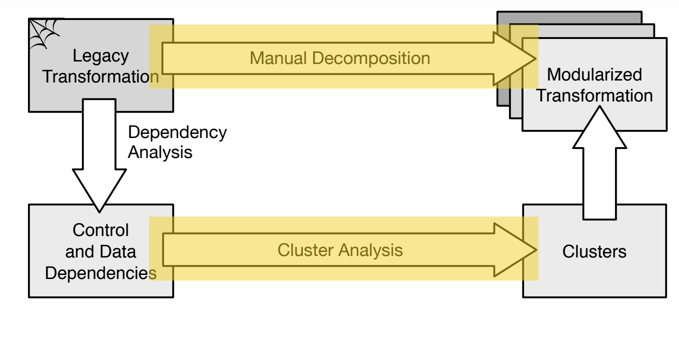
# Assessing Results




- Compare derived clustering with a manual expert clustering
- Using three similarity/distance metrics

| Configuration | Statistics | | Similarity to expert clustering | | | |
|---|---|---|---|---|---|---|
| | # Clusters | MQ index | Precision | Recall | EdgeSim | MeCl |
| **Expert clustering**<br>Derived manually | 3 | 1.067 | 100% | 100% | 100 | 100% |
| **Method-call dependencies only**<br>Hill Climbing, $WC = \langle 0, 0, 1, 0 \rangle$ | 2 | 1.214 | 20.00% | 100% | 54.54 | 60% |
| **Class-level dependencies**<br>Hill Climbing, $WC = \langle 1, 15, 5, 15 \rangle$ | 2 | 1.083 | 33.33% | 100% | 72.72 | 85% |

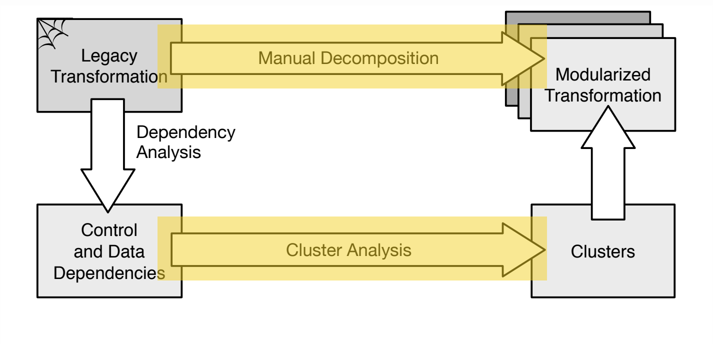
# Assessing Results




- Compare derived clustering with a manual expert clustering
- Using three similarity/distance metrics

| Configuration | Statistics | | Similarity to expert clustering | | | |
|---|---|---|---|---|---|---|
| | # Clusters | MQ index | Precision | Recall | EdgeSim | MeCl |
| **Expert clustering**<br>Derived manually | 3 | 1.067 | 100% | 100% | 100 | 100% |
| **Method-call dependencies only**<br>Hill Climbing, $WC = \langle 0, 0, 1, 0 \rangle$ | 2 | 1.214 | 20.00% | 100% | 54.54 | 60% |
| **Class-level dependencies**<br>Hill Climbing, $WC = \langle 1, 15, 5, 15 \rangle$ | 2 | 1.083 | 33.33% | 100% | 72.72 | 85% |

# Assessing Results




- Compare derived clustering with a manual expert clustering
- Using three similarity/distance metrics

| Configuration | Statistics | | Similarity to expert clustering | | | |
|---|---|---|---|---|---|---|
| | # Clusters | MQ index | Precision | Recall | EdgeSim | MeCl |
| **Expert clustering**<br>Derived manually | 3 | 1.067 | 100% | 100% | 100 | 100% |
| **Method-call dependencies only**<br>Hill Climbing, $WC = \langle 0, 0, 1, 0 \rangle$ | 2 | 1.214 | 20.00% | 100% | 54.54 | 60% |
| **Class-level dependencies**<br>Hill Climbing, $WC = \langle 1, 15, 5, 15 \rangle$ | 2 | 1.083 | 33.33% | 100% | 72.72 | 85% |

➡ Significantly better results with model elements considered

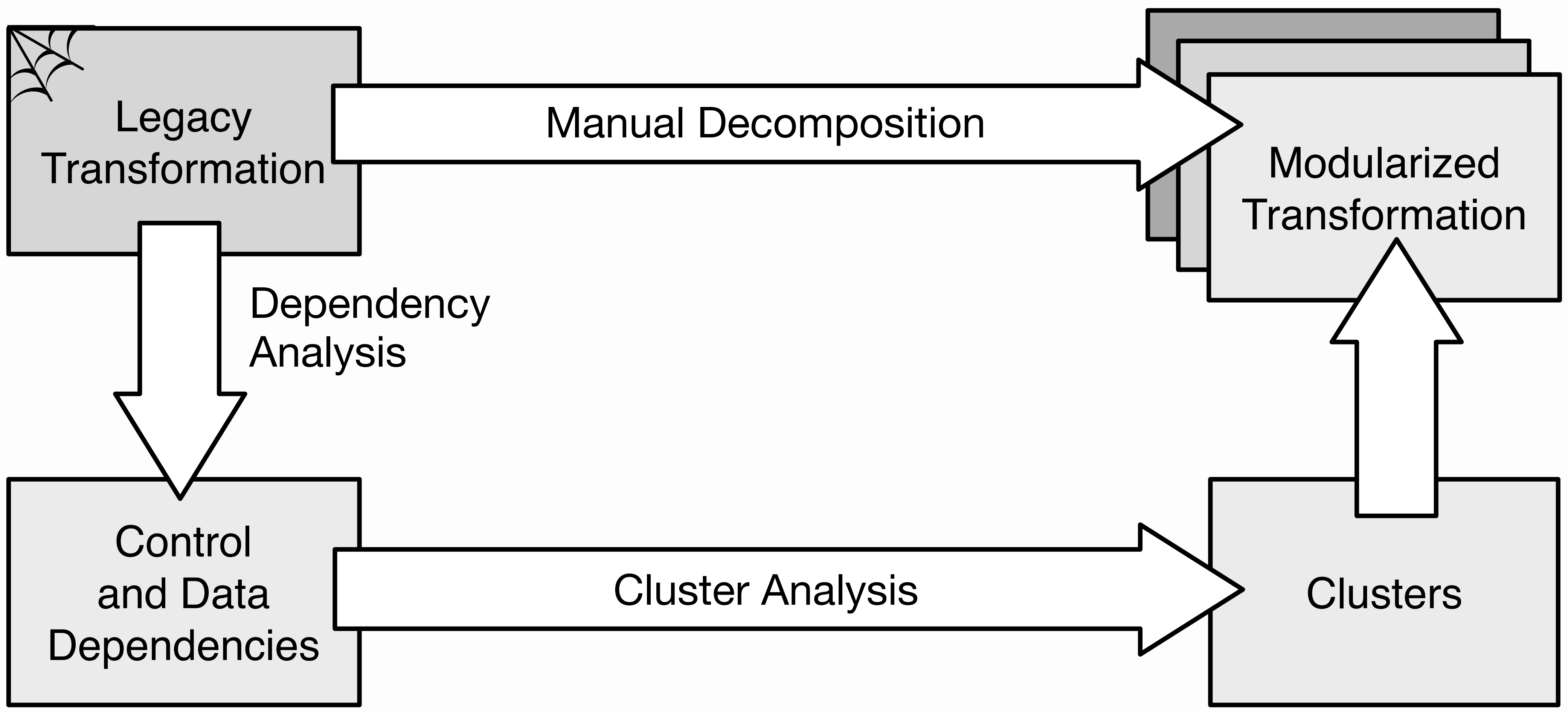
# Conclusion

**Problem**: Model transformation programs are often badly structured.

# Conclusion

**Problem**: Model transformation programs are often badly structured.

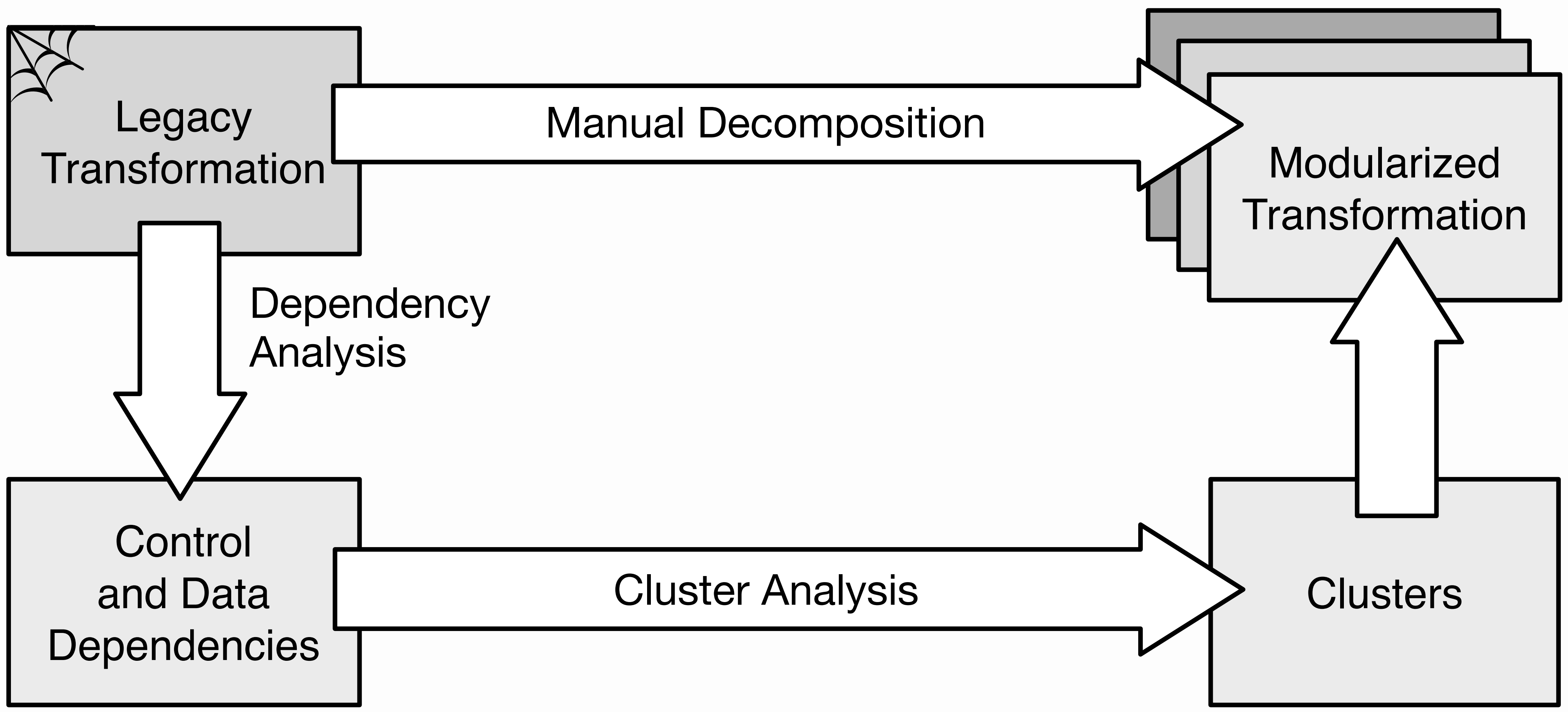
**Approach**: Apply automatic clustering algorithms to do the job.



# Conclusion

**Problem**: Model transformation programs are often badly structured.

**Approach**: Apply automatic clustering algorithms to do the job.



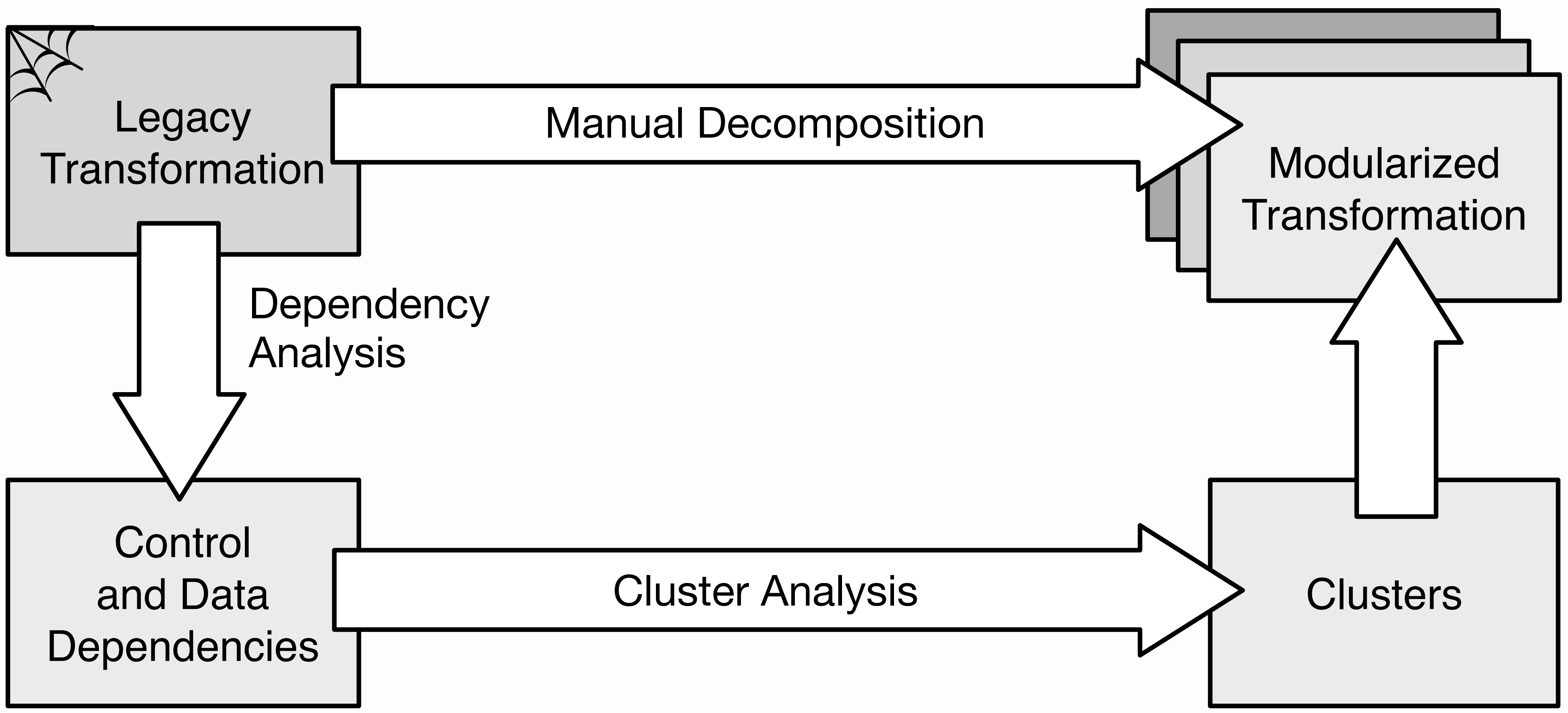

**Findings**:

- By including the model structure, we can derive clusterings that follow source and target-driven decompositional styles.
- Selecting 'good' weights requires expertise and/or experimenting.

# Conclusion

**Problem**: Model transformation programs are often badly structured.

**Approach**: Apply automatic clustering algorithms to do the job.



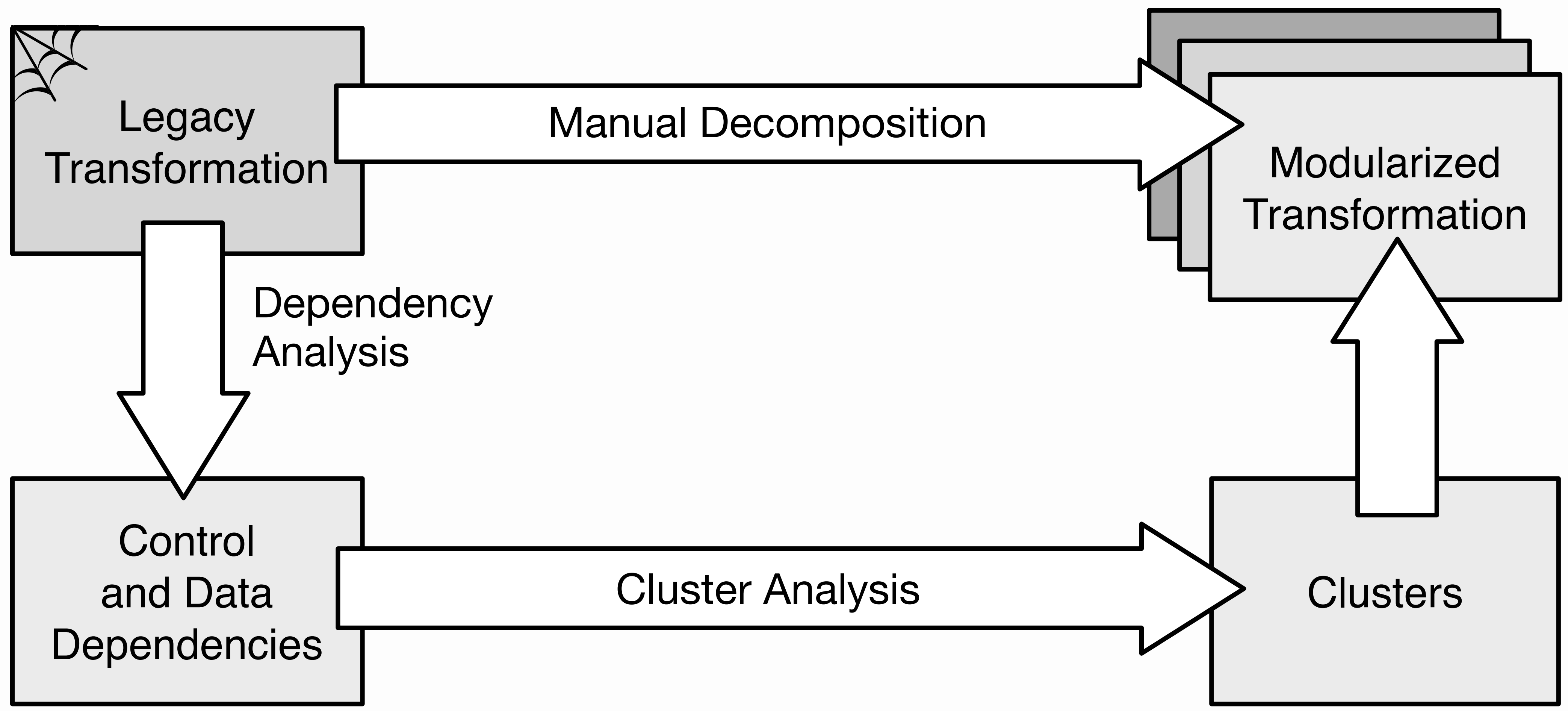

**Findings**:
- By including the model structure, we can derive clusterings that follow source and target-driven decompositional styles.
- Selecting ‘good’ weights requires expertise and/or experimenting.

**Open challenges**:
- Can we support aspectual decompositions?
- How do the results compare in maintenance scenarios?

# Conclusion

**Problem**: Model transformation programs are often badly structured.

**Approach**: Apply automatic clustering algorithms to do the job.




**Findings**:
- By including the model structure, we can derive clusterings that follow source and target-driven decompositional styles.
- Selecting ‘good’ weights requires expertise and/or experimenting.

**Open challenges**:
- Can we support aspectual decompositions?
- How do the results compare in maintenance scenarios?

Thanks!