
Dr. Antonela Tommasel
ISISTAN, CONICET-UNICEN, Argentina
antonela.tommasel@isistan.unicen.edu.ar
Who am I?

- Dr. Antonela Tommasel
  - PhD in Computer Sciences at UNICEN, December 2017

- Work at ISISTAN, CONICET-UNICEN.

- Teacher at UNICEN.

- Research Interests:
  - Recommender systems
  - Text Mining
  - Social Media
  - Machine Learning
  - …
# Table of Contents

1. Introduction & Motivation  
2. Predicting Dependencies  
3. Predicting Smells  
4. History-aware Smell Prediction  
5. Conclusions and Future Work
1. Introduction & Motivation

2. Predicting Dependencies

3. Predicting Smells

4. History-aware Smell Prediction

5. Conclusions and Future Work
As software systems evolve, the **amount and complexity of the interactions** amongst their components often increases.

- More coupling.
- “**Undesired**” dependencies amongst certain components (e.g., layer bridging, direct access to databases, cycles).
- Degradation of intended design.
Software Evolution & Dependencies

- As software systems evolve, the **amount and complexity of the interactions** amongst their components often increases.
  - More coupling.
  - "**Undesired**" dependencies amongst certain components (e.g., layer bridging, direct access to databases, cycles).
  - Degradation of intended design.
As software systems evolve, the amount and complexity of the interactions amongst their components often increases.

- More coupling.
- “Undesired” dependencies amongst certain components (e.g., layer bridging, direct access to databases, cycles).
- Degradation of intended design.
Software Evolution & Dependencies

**Apache Derby**
Introduction of a Cyclic dependency

**SubscriberDB**
Violation of architectural rules (based on a given architecture)

Version 10.8.3.0

org.apache.derby.catalog

org.apache.derby.impl.sql

org.apache.derby.impl.sql.catalog

org.apache.derby.impl.sql.execute

org.apache.derby.impl.sql.execute.rts

Version 10.9.1.0

**Keeping one-step ahead of Architectural Smells: A Machine Learning Application**

A. Tommasel

ISISTAN, CONICET-UNICEN
Software Evolution & Dependencies

Apache Derby
Introduction of a Hub-like dependency
Software Evolution & Dependencies

• Conscious efforts must be made to stop (or alleviate) degradation.
  • Plan for corrective actions (e.g., refactoring).
  • Monitor system health (e.g., via metrics).
  • Conformance checks.
Software Evolution & Dependencies

- Conscious efforts must be made to stop (or alleviate) degradation.
- Plan for corrective actions (e.g., refactoring).
- Monitor system health (e.g., via metrics).
- Conformance checks.

The early detection of such symptoms is important for developers, so that they can plan ahead for actions that preserve the quality of the system.
What can we do about it?

- Different tools available
  - LattixDSM, SonarQube, SonarGraph, JITTAC.

What can we do about it?

• Different tools available
  • LattixDSM, SonarQube, SonarGraph, JITTAC.

• Identification of problems **once they occurred** in the system!
  • Tools normally perform a dependency analysis of the source code.
  • Compute metrics/indicators, ranking of smells (e.g., by severity).
  • Show all these symptoms to developers.
However, developers may be reluctant to fix problems, when they were already introduced in the code.

Particularly, quality-related problems. Schedule pressures, “it still works”, loss of context.
However, developers may be **reluctant** to fix **problems**, when they were **already introduced** in the code.

Particularly, quality-related problems. Schedule pressures, “it still works”, loss of context.

**Predict when a dependency-related problem is likely to manifest!**
Social Network Analysis to the Rescue!

- Although there are approaches for computing coupling metrics, very few of them have dealt with the prediction of dependency relations amongst software components.
Social Network Analysis to the Rescue!

• Although there are approaches for computing coupling metrics, very few of them have dealt with the prediction of dependency relations amongst software components.

• A particular graph-based approach is social networks analysis (SNA), which has been used for modelling both nature and human phenomena.

  SNA techniques can predict links that yet do not exist between pairs of nodes in a network.
Social Network Analysis to the Rescue! …and Software Engineering?

• Evidence that the topological features of dependency graphs can reveal interesting properties of the software system under analysis.

• Nonetheless, SNA techniques has not yet greatly exploited in the Software Engineering community.
Social Network Analysis to the Rescue! …and Software Engineering?

- **Evidence** that the **topological** features of **dependency graphs** can reveal interesting properties of the software system under analysis.

- Nonetheless, **SNA** techniques has not yet greatly exploited in the Software Engineering community.

  We argue that **Social Network Analysis** techniques need to be revisited with respect to **software dependency prediction**!
1. Introduction & Motivation

2. Predicting Dependencies

3. Predicting Smells

4. Time-series Smell Prediction

5. Conclusions and Future Work
**First Try: Link Prediction Techniques**

- **Link Prediction** studies the **evolution of a network/graph** using models of network **features**.
  - Infer “missing” links between pairs of nodes.
  - Based on the observable links of the network and their attributes.

- **Homophily Principle (HP)**: interactions between similar nodes occur at a higher rate than interactions between dissimilar nodes.

- Most techniques rely on graph topological features that **assess similarity between pairs of nodes**.
First Try: Link Prediction Techniques

…but we need a graph

- Build a graph $DG(V, E)$ for system version $n$, where:
  - Each node $v$ in $V$ is Java package, and each edge $e$ in $E$ is a usage relationship between a pair of packages $v_1$ and $v_2$. 

![Graph Diagram]

Version 10.8.3.0

- org.apache.derby.catalog
- org.apache.derby.impl.sql
- org.apache.derby.impl.sql.execute
- org.apache.derby.impl.sql.execute.rts

<<use>>
First Try: Link Prediction Techniques

...but we need a graph

- Build a graph $DG(V, E)$ for system version $n$, where:
  - Each node $v$ in $V$ is a Java package, and each edge $e$ in $E$ is a usage relationship between a pair of packages $v_1$ and $v_2$.
  - Assumption 1: The package structure remains stable over versions.
  - Assumption 2: Similar packages have a high chance to establish usage dependencies.
  - Compute $score(v_1, v_2)$ to assess similarity.
First Try: Link Prediction Techniques

…but we need a graph

- Build a graph $DG(V, E)$ for system version $n$, where:
  - Each node $v$ in $V$ is Java package, and each edge $e$ in $E$ is a usage relationship between a pair of packages $v_1$ and $v_2$.
  - Assumption 1: The package structure remains stable over versions.
  - Assumption 2: Similar packages have a high chance to establish usage dependencies.
  - Compute $score(v_1, v_2)$ to assess similarity.
First Try: Link Prediction Techniques

Measuring Similarity Between Nodes

Standard Topological Similarity Metrics

- Common Neighbours
- Adamic Adar
- Kats Score
- SimRank
- ... (more metrics)

Source-code Similarity Metrics

- Kunczynsky
- Relative Matching
- Russel Rao
- ... (more metrics)
First Try: Link Prediction Techniques

Measuring Similarity Between Nodes

Common Neighbors = |T(x) ∩ T(y)|

|T(org.apache.derby.impl.sql.execute.rts) ∩ T(org.apache.derby.impl.sql.catalog)|

{org.apache.derby.catalog}  {org.apache.derby.catalog}

Common Neighbors = 1
First Try: Link Prediction Techniques

What do we want?

To what extent LP can leverage on information from software versions to predict likely dependencies in the next version, for those pairs of modules that exist in the analysed versions.
First Try: Link Prediction Techniques

What do we want?

To what extent LP can leverage on information from software versions to predict likely dependencies in the next version, for those pairs of modules that exist in the analysed versions.

- For a package $p$, a ranking of packages is built, based on their chance of having a future dependency with $p$, according to a similarity metric.

- For pairs of consecutive versions, the quality of predictions was evaluated in terms of precision (i.e., the ratio of actual dependencies discovered to the total number of predictions) for the top-N dependencies of the ranking.
First Try: Link Prediction Techniques

What do we want?

dependency graph for $v_n$

impl.sql.execute.rts

impl.sql.execute

impl.sql.catalog

impl.sql. impl.sql.
First Try: Link Prediction Techniques

What do we want?

Will this dependency appear on $v_{n+1}$?
First Try: Link Prediction Techniques

What do we want?

Output for $v_{n+1}$

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Common Neighbours</th>
<th>Adamic-Adar</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>impl.sql</td>
<td>impl.sql.execute</td>
</tr>
<tr>
<td>2</td>
<td>impl.sql.execute</td>
<td>impl.sql</td>
</tr>
<tr>
<td>3</td>
<td>impl.sql.conn</td>
<td>impl.sql.con</td>
</tr>
<tr>
<td>4</td>
<td>impl.db</td>
<td>impl.db</td>
</tr>
<tr>
<td>5</td>
<td>impl.store.raw.data</td>
<td>impl.jdbc</td>
</tr>
</tbody>
</table>

Will this dependency appear on $v_{n+1}$?
First Try: Link Prediction Techniques

Study Settings

• We analysed **package dependencies** in general.
  • Unrelated to specific design problems.

• Dependencies between **classes** were **ignored**.
First Try: Link Prediction Techniques

Study Settings

- We analysed **package dependencies** in general.
  - Unrelated to specific design problems.

- Dependencies between **classes** were **ignored**.

We are going to tackle this later!
First Try: Link Prediction Techniques

Study Settings

• We analysed **package dependencies** in general.
  • Unrelated to specific design problems.

• Dependencies between **classes** were **ignored**.

• For Link Prediction to produce reasonable outputs, a pair of consecutive versions:
  • $v_n$ and $v_{n+1}$ have approximately the same number of packages.
  • $v_{n+1}$ adds new dependencies between known packages.
  • New dependencies in $v_{n+1}$ between new packages are disregarded.
### Study Settings

Two small Java systems.

- HealthWatcher (HW)
  - 49 KLOC

- SubscriberDB (SDB)
  - 10 KLOC

<table>
<thead>
<tr>
<th></th>
<th>HWv1</th>
<th>HWv2</th>
<th>HWv3</th>
<th>HWv4</th>
<th>HWv5</th>
<th>HWv6</th>
<th>HWv7</th>
<th>HWv8</th>
<th>HWv9</th>
<th>HWv10</th>
</tr>
</thead>
<tbody>
<tr>
<td>c</td>
<td>88</td>
<td>92</td>
<td>104</td>
<td>106</td>
<td>108</td>
<td>112</td>
<td>116</td>
<td>120</td>
<td>132</td>
<td>135</td>
</tr>
<tr>
<td>p</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>22</td>
<td>23</td>
<td>23</td>
<td>24</td>
<td>24</td>
<td>25</td>
</tr>
<tr>
<td>deps</td>
<td>67</td>
<td>70 (+8,-5)</td>
<td>75 (+5)</td>
<td>85 (+10)</td>
<td>86 (+7,-2)</td>
<td>91</td>
<td>91</td>
<td>96 (+5)</td>
<td>97 (+1)</td>
<td>101 (+4)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>SDBv1</th>
<th>SDBv2</th>
<th>SDBv3</th>
<th>SDBv4</th>
<th>SDBv5</th>
<th>SDBv6</th>
<th>SDBv7</th>
<th>SDBv8</th>
<th>SDBv9</th>
<th>SDBv10</th>
</tr>
</thead>
<tbody>
<tr>
<td>c</td>
<td>98</td>
<td>167</td>
<td>192</td>
<td>193</td>
<td>193</td>
<td>193</td>
<td>195</td>
<td>195</td>
<td>195</td>
<td>195</td>
</tr>
<tr>
<td>p</td>
<td>14</td>
<td>16</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>deps</td>
<td>30</td>
<td>47 (+17)</td>
<td>50 (+4,-1)</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>51 (+1)</td>
<td>51</td>
<td>51</td>
</tr>
</tbody>
</table>

---

A. Tommasel

*Keeping one-step ahead of Architectural Smells: A Machine Learning Application*
First Try: Link Prediction Techniques

So far not so good

• Unfortunately, ranking-based LP is not enough for software dependencies.
  • Precision of predicted links is rather low (0.14-0.25 at most).

• The Homophily Principle does not always hold for Java packages.
  • e.g., dependencies might still appear between dissimilar packages.

• Two similar packages can intentionally be designed to not become dependent on each other.
  • e.g., based on business logic or modularity considerations.
To what extent Link Prediction can leverage on information from the current version to predict dependencies in the next version?

Use statistical techniques to give computer systems the ability to "learn" (on a specific task) with data, without being explicitly programmed.
To what extent Link Prediction can leverage on information from the current version to predict dependencies in the next version?

Use statistical techniques to give computer systems the ability to "learn" (on a specific task) with data, without being explicitly programmed.
Second Try: Apply Machine Learning Models

We need a “dataset”

A binary classifier is trained using the topological information provided by a given graph version.
Second Try: Apply Machine Learning Models

We need a “dataset”

A binary classifier is trained using the topological information provided by a given graph version.

- An instance for the classifier consists of:
  - A pair of nodes.
  - A list of features (e.g., structural metrics) for the pair.
  - A label indicating if the nodes are linked (positive class) or not (negative class).
Second Try: Apply Machine Learning Models

We need a “dataset”

• **Existing dependencies** are used to compute features for instances of the positive class.

• **Missing dependencies** are used to compute features for instances of the negative class.

• Both training and test sets need to be defined.
  - The training set considers the known structure of \( v_n \).
  - The test set considers the full graph of \( v_{n+1} \).
Second Try: Apply Machine Learning Models

We need a “dataset”

A. Tommasel

• The predictions were considered over selected versions.
  • The first item is the version for the training set.
  • The second one is the version for the test set
How did it go?

The classifier finds all new dependencies (high recall) but it also mistakenly reports non-existing dependencies (low precision).
Second Try: Apply Machine Learning Models

How did it go?

- Better values for the **weighted class** (both positive and negative instances).
  - Average precision values of 0.85 (SDB) and 0.96 (HW)
- However, **precision for the positive class was far from ideal!**
  - Average values of 0.74 (SDB) and 0.23 (HW)
• Variations imply it might be difficult to differentiate between dependencies and non-dependencies due to similar structural characteristics.
  • Need to consider additional information for having good predictions.
To what extent Link Prediction can leverage on information from past versions to predict dependencies in the next version?
To what extent Link Prediction can leverage on information from past versions to predict dependencies in the next version?

Dynamic SNA (i.e., observations of the graph at different time periods) + topological features = Learn a robust ML model able to predict new links.
Third Try: Time Series Forecasting

We are not yet predicting new dependencies, but estimating the features’ scores based on previous versions.
Third Try: Time Series Forecasting

- Prediction is based on a classifier trained with the last known version of the system, \( v_n \).
- The test set considers the estimated feature scores for \( v_{n+1} \).

<table>
<thead>
<tr>
<th>source</th>
<th>target</th>
<th>uses target?</th>
<th>Common Neighbours</th>
</tr>
</thead>
<tbody>
<tr>
<td>A - B</td>
<td>true</td>
<td>0.353</td>
<td></td>
</tr>
<tr>
<td>A - D</td>
<td>false</td>
<td>0.618</td>
<td></td>
</tr>
<tr>
<td>C - B</td>
<td>true</td>
<td>0.389</td>
<td></td>
</tr>
<tr>
<td>A - E</td>
<td>true</td>
<td>0.385</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>B - D</td>
<td>false</td>
<td>0.605</td>
<td></td>
</tr>
<tr>
<td>E - F</td>
<td>true</td>
<td>0.171</td>
<td></td>
</tr>
<tr>
<td>C - F</td>
<td>true</td>
<td>0.1</td>
<td></td>
</tr>
</tbody>
</table>

estimation for \( v_{n+1} \)

<table>
<thead>
<tr>
<th>source</th>
<th>target</th>
<th>uses target?</th>
<th>Common Neighbours</th>
</tr>
</thead>
<tbody>
<tr>
<td>A - B</td>
<td>?</td>
<td>0.453</td>
<td></td>
</tr>
<tr>
<td>A - D</td>
<td>?</td>
<td>0.718</td>
<td></td>
</tr>
<tr>
<td>C - B</td>
<td>?</td>
<td>0.289</td>
<td></td>
</tr>
<tr>
<td>A - E</td>
<td>?</td>
<td>0.685</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>B - D</td>
<td>?</td>
<td>0.805</td>
<td></td>
</tr>
<tr>
<td>E - F</td>
<td>?</td>
<td>0.171</td>
<td></td>
</tr>
<tr>
<td>C - F</td>
<td>?</td>
<td>0.11</td>
<td></td>
</tr>
</tbody>
</table>

source and target dependency graphs for \( v_{n-1} \) and \( v_n \)
Third Try: Time Series Forecasting

• The versions represent the span for the estimations.
  • v1-v3 means that v1, v2 and v3 served to estimate the features for v4 (test set).

• Each pair represents the span of estimations, with real versus estimated features.

A. Tommasel

• Better values for the **positive class**!
  
  • Average values of 0.84.
  • Estimated features are “better predictors” than real features.
The choice of versions for forecasting was relevant!
  - More versions sometimes decreases the quality of the predictions.

This effect could be related to the structural changes in each version.
Lessons Learned
What did we learn?

• We wanted to
  • Assess the LP performance in dependency graphs
  • Assess the kind of information required for having reasonable predictions.

• Naïve LP techniques are not adequate for the task.

• Leveraging on information from previous versions gives reasonable predictions, although not all versions seem useful.
Lessons Learned

What did we learn?

• We wanted to
  • Assess the LP performance in dependency graphs
  • Assess the kind of information required for having reasonable predictions.

• Naïve LP techniques are not adequate for the task.

• Leveraging on information from previous versions gives reasonable predictions, although not all versions seem useful.

Machine Learning techniques have the potential for Link Prediction applied to software dependencies
Lessons Learned

What do we do now?

• Despite the potential of LP techniques, further investigation is needed.

• A systematic study with more systems is required to corroborate our initial findings.

• The features currently used can be extended.
Lessons Learned

What do we do now?

- Despite the potential of LP techniques, further investigation is needed.
- A systematic study with more systems is required to corroborate our initial findings.
- The features currently used can be extended.

Develop **customized LP algorithms for dependency-related problems**

(e.g., layering violations, cycles, hub-like configurations)
Table of Contents

1. Introduction & Motivation
2. Predicting Dependencies
3. Predicting Smells
4. History-aware Smell Prediction
5. Conclusions and Future Work
Dependency-based Smells

• An architectural bad smell is a commonly used set of architectural design decisions that negatively impacts system lifecycle properties.
  • E.g. understandability, testability, extensibility, and reusability.
Dependency-based Smells

- An architectural **bad smell** is a commonly used set of **architectural design decisions** that **negatively** impacts system lifecycle properties.
  - E.g. understandability, testability, extensibility, and reusability.

- **Dependency-based smells** involve **interactions** amongst system components.
  - Occur when one or more components **violate design principles** or rules.
  - Often manifest themselves as **undesired dependencies** in the source code.
**Dependency-based Smells**

- An architectural **bad smell** is a commonly used set of **architectural design decisions** that **negatively** impacts system lifecycle properties.
  - E.g. understandability, testability, extensibility, and reusability.

- **Dependency-based smells** involve **interactions** amongst system components.
  - Occur when one or more components **violate design principles** or rules.
  - Often manifest themselves as **undesired dependencies** in the source code.
Dependency-based Smells

Cyclic Dependencies

- Various components directly or indirectly depend on each other to function properly.

- A case of an undesired dependency.
  - Breaks the desirable acyclic nature of a subsystem’s dependency structure.

- Components involved in a cycle can be hard to maintain, test or reuse in isolation.

- Cycles might have different shapes.
  - Different harmful levels for the system health than others.
Dependency-based Smells

Hub-like Dependencies

• A component has outgoing and ingoing dependencies with a large number of other components.

• Detecting hubs:
  1. Computes the median of the number of incoming and outgoing dependencies of all packages.
  2. For each package: Are both its incoming and outgoing dependencies greater than the incoming and outgoing medians?
  3. incoming - outgoing dependencies < than a fraction of the total dependencies of that package.
Dependency-based Smells

Once again, we resort to Machine Learning!
Dependency-based Smells
Once again, we resort to Machine Learning!

predict ➔ filter
Dependency-based Smells

Once again, we resort to Machine Learning!

predict

filter

predict the appearance of new dependencies in the next system version

filter the predicted dependencies according to the characteristics of specific types of smells.
**Dependency-based Smells**

1. **Prediction Phase**
   - Dependency graph for $v_{n-1}$
   - Dependency Predictor
   - Dependency graph + predicted dependencies

2. **Filtering Phase**
   - Topological features
   - Content features
   - Train Model
   - Evaluate Model
   - Predicted Smells

Dataset

- Topological features
- Content features
- Evaluate Model
- Predicted Smells

- Train Model
- Evaluate Model
- Predicted Smells

A. Tommasel

*Keeping one step ahead of Architectural Smells: A Machine Learning Application*
Dependency-based Smells

Prediction Phase

- Individual dependencies are inferred based on training a binary classification model.
- Dependency graphs of the current ($v_n$) and previous versions ($v_{n-1}$) are used as inputs.
- The output is the set of dependencies that are likely to appear in the next system version $v_{n+1}$.
- This phase is **smell independent**.
  - Only identifies dependencies that might prefigure different smells in the second phase.
The prediction phase internally involves 3 steps.

**Step 1**
- The instance-based representation are constructed, based on both topological and content-based features.
  - Existing dependencies → positive class.
  - Missing dependencies → negative class.
- The training set includes:
  - Existing dependencies in $v_{n-1}$.
  - Missing dependencies in $v_{n-1}$.
  - Existing dependencies in $v_n$.
Dependency-based Smells

Prediction Phase

- **Content-based features** are an alternative (and complementary) similarity criterion to topological features.

- Natural language processing routines are used to transform texts into their **bag-of-words representations** by considering different aspects of the original texts.
  - Restricted to only the appearing nouns, adjective or verbs…
  - Remove punctuation…
**Dependency-based Smells**

**Prediction Phase**

- **Content-based features** are an alternative (and complementary) similarity criterion to topological features.

- Natural language processing routines are used to transform texts into their **bag-of-words representations** by considering different aspects of the original texts.
  - Restricted to only the appearing nouns, adjective or verbs...
  - Remove punctuation...

- The bag-of-words class representations can be used to assess the similarity amongst the classes.
  - Cosine similarity is commonly used.

- Each Java class $c$ as a bag-of-words containing the most representative tokens that characterize its source code.
  - Either considering the name of the field attributes of the classes, the name of the declared methods or the class comments and documentation.
Dependency-based Smells

Prediction Phase

org.apache.derby.iapi.sql.dictionary.
ColumnDescriptor.java

Fields
Method Names
Comments
The prediction phase internally involves 3 steps.

**Step 2**

- The classification model is built.
  - The classifier is trained for properly learning instances of both the positive and negative classes.
    - Includes information of dependencies in $v_{n-1}$ being guaranteed that are not going to appear in $v_n$. 

The prediction phase internally involves 3 steps.

**Step 3**

- Dependencies are predicted.
- Only potential dependencies considering the packages already existing in $v_n$ are considered.
The prediction of a dependency is not enough to predict the appearance of an architectural smell.

- Not every predicted dependency might cause an smell to emerge.

Predicted dependencies undergo a filtering process.
- Filters are smell-dependent.
Keeping one step ahead of Architectural Smells: A Machine Learning Application

**Dependency-based Smells**

**Filtering Phase - Cycles**

- Considers only predicted dependencies that would lead to the closure of new cycles in $v_{n+1}$.

Two variants:

- *All predicted dependencies are simultaneously considered.*
  - Allows to detect cycles needed more than one dependency to be closed.

- *Dependencies are individually analysed.*
  - Allows to detect cycles needed only one dependency to be closed.
• Considers only predicted dependencies that would lead to the closure of new cycles in $v_{n+1}$.

Two variants:

• All predicted dependencies are simultaneously considered.
  • Allows to detect cycles needing more than one dependency to be closed.

• Dependencies are individually analysed.
  • Allows to detect cycles needing only one dependency to be closed.
• Considers only predicted dependencies that would lead to the closure of new cycles in $v_{n+1}$.

Two variants:

• All predicted dependencies are simultaneously considered.
  • Allows to detect cycles needing more than one dependency to be closed.

• Dependencies are individually analysed.
  • Allows to detect cycles needing only one dependency to be closed.
Dependency-based Smells

Filtering Phase - Cycles

• Considers only predicted dependencies that would lead to the closure of new cycles in $v_{n+1}$.

Two variants:

• All predicted dependencies are simultaneously considered.
  • Allows to detect cycles needing more than one dependency to be closed.

• Dependencies are individually analysed.
  • Allows to detect cycles needing only one dependency to be closed.

This cycle requires one new dependency to be closed.

The cycle requiring two dependencies is not going to be found.

A. Tommasel

ISISTAN, CONICET-UNICEN
Dependency-based Smells

Filtering Phase - Hubs

- Only the nodes incidental to the predicted edges that fit with the hub definition are actually predicted.
  - Allow the detection of those nodes becoming hubs due to the addition of new dependencies.
  - Disregard nodes that might become hubs due to changes in the overall structure of the dependency graph.

Three variants:
- Dependencies are individually analysed.
- Dependencies are grouped per node.
- All predicted dependencies are simultaneously considered.
Dependency-based Smells
Filtering Phase - Hubs

- Only the nodes incidental to the predicted edges that fit with the hub definition are actually predicted.
  - Allow the detection of those nodes becoming hubs due to the addition of new dependencies.
  - Disregard nodes that might become hubs due to changes in the overall structure of the dependency graph.

Three variants:
- Dependencies are individually analysed.
- Dependencies are grouped per node.
- All predicted dependencies are simultaneously considered.

- Only the nodes incidental to the predicted edges that fit with the hub definition are actually predicted.
  - Allow the detection of those nodes becoming hubs due to the addition of new dependencies.
  - Disregard nodes that might become hubs due to changes in the overall structure of the dependency graph.

Three variants:
- Dependencies are individually analysed.
- Dependencies are grouped per node.
- All predicted dependencies are simultaneously considered.
Dependency-based Smells

Filtering Phase - Hubs

- Only the nodes incidental to the predicted edges that fit with the hub definition are actually predicted.
  - Allow the detection of those nodes becoming hubs due to the addition of new dependencies.
  - Disregard nodes that might become hubs due to changes in the overall structure of the dependency graph.

Three variants:

- Dependencies are individually analysed.
- Dependencies are grouped per node.
- All predicted dependencies are simultaneously considered.
Only the nodes incidental to the predicted edges that fit with the hub definition are actually predicted.
- Allow the detection of those nodes becoming hubs due to the addition of new dependencies.
- Disregard nodes that might become hubs due to changes in the overall structure of the dependency graph.

Three variants:
- Dependencies are individually analysed.
- Dependencies are grouped per node.
- All predicted dependencies are simultaneously considered.
## Dependency-based Smells

### Study Settings

<table>
<thead>
<tr>
<th>System</th>
<th>Version</th>
<th>#c</th>
<th>#p</th>
<th>#deps</th>
<th>#cycles</th>
<th>cycle length</th>
<th>#hubs</th>
<th>hub degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>derby</td>
<td>10.5.1.1</td>
<td>1344</td>
<td>96</td>
<td>767</td>
<td>234</td>
<td>11.59</td>
<td>28</td>
<td>+2</td>
</tr>
<tr>
<td>derby</td>
<td>10.5.3.0</td>
<td>1344</td>
<td>96</td>
<td>768</td>
<td>234</td>
<td>11.59</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>derby</td>
<td>10.6.1.0</td>
<td>1387</td>
<td>98</td>
<td>804</td>
<td>254</td>
<td>12.99</td>
<td>29</td>
<td>+3</td>
</tr>
<tr>
<td>derby</td>
<td>10.6.2.1</td>
<td>1387</td>
<td>98</td>
<td>805</td>
<td>255</td>
<td>13.02</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>derby</td>
<td>10.7.1.1</td>
<td>1389</td>
<td>98</td>
<td>807</td>
<td>257</td>
<td>12.98</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>derby</td>
<td>10.8.1.2</td>
<td>1395</td>
<td>97</td>
<td>837</td>
<td>305</td>
<td>15.17</td>
<td>30</td>
<td>+1</td>
</tr>
<tr>
<td>derby</td>
<td>10.8.3.0</td>
<td>1395</td>
<td>96</td>
<td>841</td>
<td>306</td>
<td>15.13</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>derby</td>
<td>10.9.1.0</td>
<td>1406</td>
<td>96</td>
<td>851</td>
<td>280</td>
<td>13.43</td>
<td>29</td>
<td>+1</td>
</tr>
<tr>
<td>derby</td>
<td>10.10.1.1</td>
<td>1453</td>
<td>100</td>
<td>938</td>
<td>291</td>
<td>13.32</td>
<td>29</td>
<td>-1</td>
</tr>
</tbody>
</table>

Two medium Java systems.

- **Apache Derby**
  - 14 versions.
  - 40 KLOC.

- **Apache Ant**
  - 18 versions.
  - 60 KLOC.
## Dependency-based Smells

### Study Settings

Two medium Java systems.

- **Apache Derby**
  - 14 versions.
  - 40 KLOC.

- **Apache Ant**
  - 18 versions.
  - 60 KLOC.

<table>
<thead>
<tr>
<th>System</th>
<th>#c</th>
<th>#p</th>
<th>#deps</th>
<th>#cycles</th>
<th>cycle length</th>
<th>#hubs</th>
<th>hub degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apache Derby 1.6.0</td>
<td>352</td>
<td>24</td>
<td>90 +20,-1</td>
<td>30 +1</td>
<td>3.73</td>
<td>9 +2</td>
<td>14.22</td>
</tr>
<tr>
<td>Apache Derby 1.6.2</td>
<td>369</td>
<td>24</td>
<td>92 +2</td>
<td>43 +2</td>
<td>4.12</td>
<td>9</td>
<td>14.67</td>
</tr>
<tr>
<td>Apache Derby 1.6.3</td>
<td>380</td>
<td>25</td>
<td>97 +5</td>
<td>43 +1</td>
<td>4.7</td>
<td>9</td>
<td>15.33</td>
</tr>
<tr>
<td>Apache Derby 1.7.1</td>
<td>502</td>
<td>29</td>
<td>137 +46,-6</td>
<td>63 +5</td>
<td>5.1</td>
<td>12 +1</td>
<td>17.42</td>
</tr>
</tbody>
</table>

Dependency-based Smells

How did it go? – Prediction Phase

- Compares results of considering either topological or topological + content features.
- Results are presented for those sets of versions in which new dependencies between already existing packages were added.


Dependency-based Smells

How did it go? – Prediction Phase

- Adding content-based features increased the quality of the predicted dependencies.
- Average improvements of 27%.
Dependency-based Smells

How did it go? – Prediction Phase

- High F-Measure values are due to a **high recall** and a **moderate precision**.
- The trained model is capable of finding most future dependencies, but it also predicts false dependencies.
## Dependency-based Smells

### How did it go? – Cycle Prediction

<table>
<thead>
<tr>
<th>Version</th>
<th>Precision Individual Analysis</th>
<th>Recall Individual Analysis</th>
<th>Precision AllDependencies</th>
<th>Recall AllDependencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.4.1.0</td>
<td>0.98</td>
<td>0.97</td>
<td>0.96</td>
<td>0.95</td>
</tr>
<tr>
<td>10.4.2.0</td>
<td>0.97</td>
<td>0.96</td>
<td>0.95</td>
<td>0.94</td>
</tr>
<tr>
<td>10.5.1.1</td>
<td>0.96</td>
<td>0.95</td>
<td>0.94</td>
<td>0.93</td>
</tr>
<tr>
<td>10.5.3.0</td>
<td>0.95</td>
<td>0.94</td>
<td>0.93</td>
<td>0.92</td>
</tr>
<tr>
<td>10.6.1.0</td>
<td>0.94</td>
<td>0.93</td>
<td>0.92</td>
<td>0.91</td>
</tr>
<tr>
<td>10.6.2.1</td>
<td>0.93</td>
<td>0.92</td>
<td>0.91</td>
<td>0.90</td>
</tr>
<tr>
<td>10.7.1.1</td>
<td>0.92</td>
<td>0.91</td>
<td>0.90</td>
<td>0.89</td>
</tr>
<tr>
<td>10.8.1.2</td>
<td>0.91</td>
<td>0.90</td>
<td>0.89</td>
<td>0.88</td>
</tr>
<tr>
<td>10.8.2.2</td>
<td>0.90</td>
<td>0.89</td>
<td>0.88</td>
<td>0.87</td>
</tr>
<tr>
<td>10.8.3.0</td>
<td>0.89</td>
<td>0.88</td>
<td>0.87</td>
<td>0.86</td>
</tr>
<tr>
<td>10.9.1.0</td>
<td>0.88</td>
<td>0.87</td>
<td>0.86</td>
<td>0.85</td>
</tr>
<tr>
<td>10.10.1.1</td>
<td>0.87</td>
<td>0.86</td>
<td>0.85</td>
<td>0.84</td>
</tr>
</tbody>
</table>

- In most cases recall is almost perfect (almost every new dependency leading to the closure of a quasi-cycle was found).
- Precision indicates that some mistaken dependencies are also predicted.
  - At most 5 mistaken predictions (0.06% of total dependencies).
## Dependency-based Smells

### How did it go? – Cycle Prediction

<table>
<thead>
<tr>
<th>Number</th>
<th>10.4.1.0</th>
<th>10.4.2.0</th>
<th>10.5.1.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.4</td>
<td>0.6</td>
<td>0.8</td>
</tr>
<tr>
<td>0.2</td>
<td>0.6</td>
<td>0.8</td>
<td>1.0</td>
</tr>
<tr>
<td>0.4</td>
<td>0.8</td>
<td>1.0</td>
<td>1.2</td>
</tr>
<tr>
<td>0.6</td>
<td>1.0</td>
<td>1.2</td>
<td>1.4</td>
</tr>
<tr>
<td>0.8</td>
<td>1.2</td>
<td>1.4</td>
<td>1.6</td>
</tr>
<tr>
<td>1.0</td>
<td>1.4</td>
<td>1.6</td>
<td>1.8</td>
</tr>
</tbody>
</table>

- Similar performance for both variants.
  - Quasi-cycles are closed by adding only one dependency or by multiple dependencies that also individually close cycles.
How did it go? – Cycle Prediction

- Differences between the variants could be explained by the existence of quasi-cycles needing +1 dependency to be closed.
  - Precision of individual-analysis is not affected, but recall decreases.
Dependency-based Smells

How did it go? – Hub Prediction

- The performance of the variants differ.
  - individual-analysis. ↓recall (highest number of missed nodes) ↑precision (fewest mistaken predictions)
  - all-node. ↑recall → precision (mistaken predictions) → neighbourhood more important than overall structure
  - all-dependencies. ↓recall ↓precision
Dependency-based Smells

How did it go? – Hub Prediction

- At least one missed smell.
- Mistaken predictions in the first phase.
- Hubs might not only depend on the addition of new edges but on the overall graph structure.
- Hubs might also depend on the unknown structure of the graph (dependencies added between yet unknown packages).
• An initial evaluation with two types of smells showed a good performance!  
  • High recall, low precision.

• Including content-based features improves dependency prediction.

• The choice of the filter variant (for a given smell type) can affect both recall and precision.  
  • We preferred good recall over precision in the analysed cases.

• Smell predictions depended on both the current overall system structure and version history.
Lessons Learned

What do we do now?

• Perform a systematic study with more systems and other dependency-based smells.

• The prediction capabilities are sensitive to the prediction model.
  • Analyse and extend the set of features used.
  • Considering software specific-metrics?

• Smells might not be harmful.
  • How can we train a model to discard them?
Lessons Learned

What do we do now?

• Perform a systematic study with more systems and other dependency-based smells.

• The prediction capabilities are sensitive to the prediction model.
  • Analyse and extend the set of features used.
  • Considering software specific-metrics?

• Smells might not be harmful.
  • How can we train a model to discard them?

Increase precision!
# Table of Contents

1. Introduction & Motivation  
2. Predicting Dependencies  
3. Predicting Smells  
4. History-aware Smell Prediction  
5. Conclusions and Future Work
• Most link prediction approaches have been proposed based on **static** network representations.
  • A **snapshot** of the network is available and the goal is to predict the future links.
Time-series Smell Prediction

- Most link prediction approaches have been proposed based on static network representations.
  - A snapshot of the network is available and the goal is to predict the future links.

- Nonetheless, networks are dynamic and perhaps nondeterministic.
  - Changes in the underlying structure and parameters over time.

- In these cases, additional information could be extracted from the history of network evolution.
Most link prediction approaches have been proposed based on **static** network representations.
- A **snapshot** of the network is available and the goal is to predict the future links.

Nonetheless, networks are **dynamic** and perhaps nondeterministic.
- Changes in the underlying structure and parameters over time.

In these cases, additional information could be extracted from the **history of network evolution**.

Link prediction techniques could be enriched by including **time series** information and **reinforcement learning** mechanisms.
Time-series Smell Prediction

predict ➔ filter
Time-series Smell Prediction

predict → filter → reinforce learning

adjust the confidence of predictions
Time-series Smell Prediction

1. Prediction Phase
   - Dependency graph for \( v_{n-1} \)
   - Dependency graph for \( v_n \)
   - Topological features
   - Content features

2. Filtering Phase
   - Dependency graph + predicted dependencies
   - Ranked dependencies:
     1. \( B \rightarrow D \)
     2. \( F \rightarrow A \)
     3. \( F \rightarrow F \)
     4. \( C \rightarrow E \)
   - Smell specific filter

3. Reinforcement Phase at time \( n + 1 \)
   - Potential dependencies:
     1. \( B \rightarrow D \)
     2. \( F \rightarrow D \)
     3. \( F \rightarrow A \)
     4. \( C \rightarrow E \)
   - penalise dependencies

A. Tommasel

ISISTAN, CONICET-UNICEN
Time-series Smell Prediction

Leverages on the history of software versions to estimate the confidence of predictions.

**Three** phases:
Time-series Smell Prediction

Leverages on the history of software versions to estimate the confidence of predictions.

**Three phases:**

1. Considering the information of two software versions, it predicts the appearance of new dependencies in the next system version.
Time-series Smell Prediction

Leverages on the history of software versions to estimate the confidence of predictions.

Three phases:

1. Considering the information of two software versions, it predicts the appearance of new dependencies in the next system version.

2. Smells are filtered and ranked according to:
   • The characteristics of the specific types of smells.
   • The confidence score of the predicted dependencies.
Time-series Smell Prediction

Leverages on the history of software versions to estimate the confidence of predictions.

Three phases:

1. Considering the information of two software versions, it predicts the appearance of new dependencies in the next system version.

2. Smells are filtered and ranked according to:
   - The characteristics of the specific types of smells.
   - The confidence score of the predicted dependencies.

3. When the next system version is known, the confidence of predicted dependencies is updated to reflect the actual changes in the actual dependency graph.
   - Applies an adaptation of reinforcement learning.
• Up to now, all predicted smells were presented to the developer, which resulted in the mistaken prediction of smells.
Time-series Smell Prediction

Filtering & Ranking

• Up to now, all predicted smells were presented to the developer, which resulted in the mistaken prediction of smells.

• Once smells are predicted and prioritised, we need to define which of them are going to be presented.

• Choosing the number of smells to recommend might not be easy!
Several alternatives:

- Set a **fixed threshold** and always recommend the same number of smells.
  - Threshold could be based on relevancy scores, a percentage of instances or the number of predicted items.

- This has several drawbacks.
  - Ignores the characteristics of the task at hand.
  - Might fail to acknowledge the possibility of rankings presenting different scores distributions.
Several alternatives:

- The number of smells to predict will be chosen according to the **history of discoverable smells in the previous versions**.

- The average number of predictable smells in the previous versions of the system plus its standard deviation.

- An “error” margin could also be included based on previous nDCG scores.
When the following software version is known, the reinforcement learning phase updates the relevance of dependencies based on the structure of the newest system version.

Includes additional information regarding the evolution of the network.
Time-series Smell Prediction

Reinforcement Learning

- A pool of predicted dependencies is maintained.
- In every iteration, new predicted dependencies are added to the pool and associated to a learning automaton that updates the confidence of the predicted dependency according to changes in the environment.
- The learning automaton starts with a confidence of 1.
For each predicted dependency there are two possibilities.
For each predicted dependency there are two possibilities.

1. The dependency appears on the new software version → It is removed from the pool.
Time-series Smell Prediction

For each predicted dependency there are two possibilities.

1. The dependency **appears** on the new software version → It is removed from the pool.

2. The dependency **does not appear** on the new software version → The associated the learning automaton decreases its confidence to penalise the incorrect prediction.

\[ C_{n+1} = 1 - b \times C_n \]
We still need to tailor the size of the ranking.
Time-series Smell Prediction

How did it go?

Apache Wicket

We still need to tailor the size of the ranking.
Table of Contents

1. Introduction & Motivation
2. Predicting Dependencies
3. Predicting Smells
4. History-aware Smell Prediction
5. Conclusions and Future Work
Summary

• As software systems evolve “undesired” dependencies appear.
• Degradation of intended design.
Summary

• As software systems evolve “undesired” dependencies appear.
• Degradation of intended design.

Machine Learning can help predict dependencies.
Summary

• As software systems evolve “undesired” dependencies appear.
  • Degradation of intended design.

Machine Learning can help predict dependencies.

Predicted dependencies can be used to predict smells.
Summary

• As software systems evolve “undesired” dependencies appear.
• Degradation of intended design.

Machine Learning can help predict dependencies.

Predicted dependencies can be used to predict smells.

Plan ahead for actions that preserve the quality of the system.
We are far from finished...

"Now this is not the end. It is not even the beginning of the end. But it is, perhaps, the end of the beginning."

• Can communities help boost predictions?

• More features.
  • Design metrics? OO metrics? Global characteristics of smells?

• Analyse other dependency-based problems!
  • Analyse other types of smells?

• Can we predict the appearance of new nodes (e.g. new packages, classes)?

• Can we predict the disappearance of dependencies?

• How about a tool?  ... and a lot more!
Questions?
Two and a Half Papers


Dr. Antonela Tommasel
ISISTAN, CONICET-UNICEN, Argentina
antonela.tommasel@isistan.unicen.edu.ar