Breaking Bad?

Semantic Versioning and Impact of Breaking Changes in Maven Central

BENEVOL 2020

Submitted to the Empirical Software Engineering (EMSE) Journal

CWI

LaBRI

TU/e

Lina Ochoa

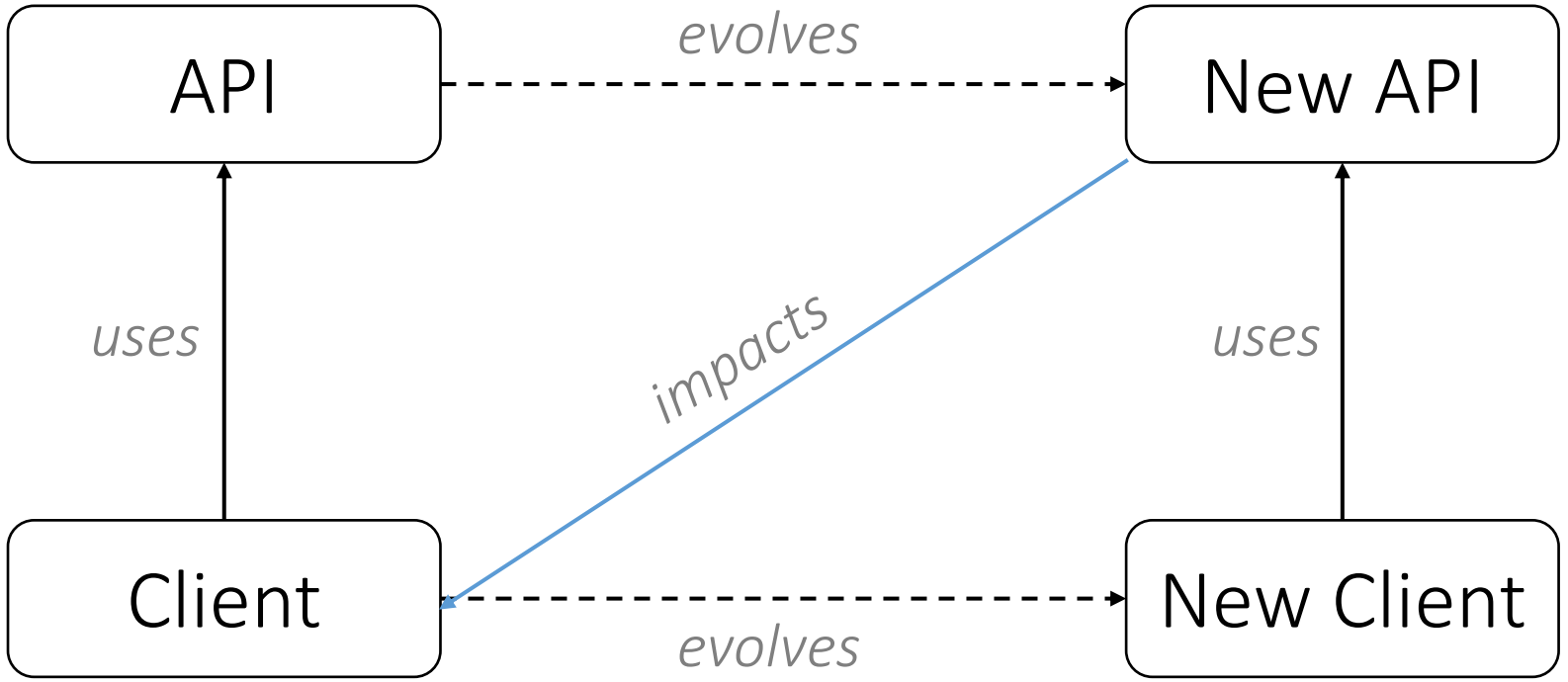
Thomas Degueule

Jean-Rémy Falleri

Jurgen Vinju

Software (API) **Evolution**

API: interface that exposes a set of services to be reused by client projects.



Adding and **Abstract Method**

JavaServlet 3.0.1

```
public interface HttpServletRequest
extends ServletRequest {

    public String getAuthType();
    public String getMethod();


    [...]
}
```

JavaServlet 3.1.0

```
public interface HttpServletRequest
extends ServletRequest {

    public String getAuthType();
    public String getMethod();
    public String changeSessionId();

    [...]
}
```

Impact on a Client Project

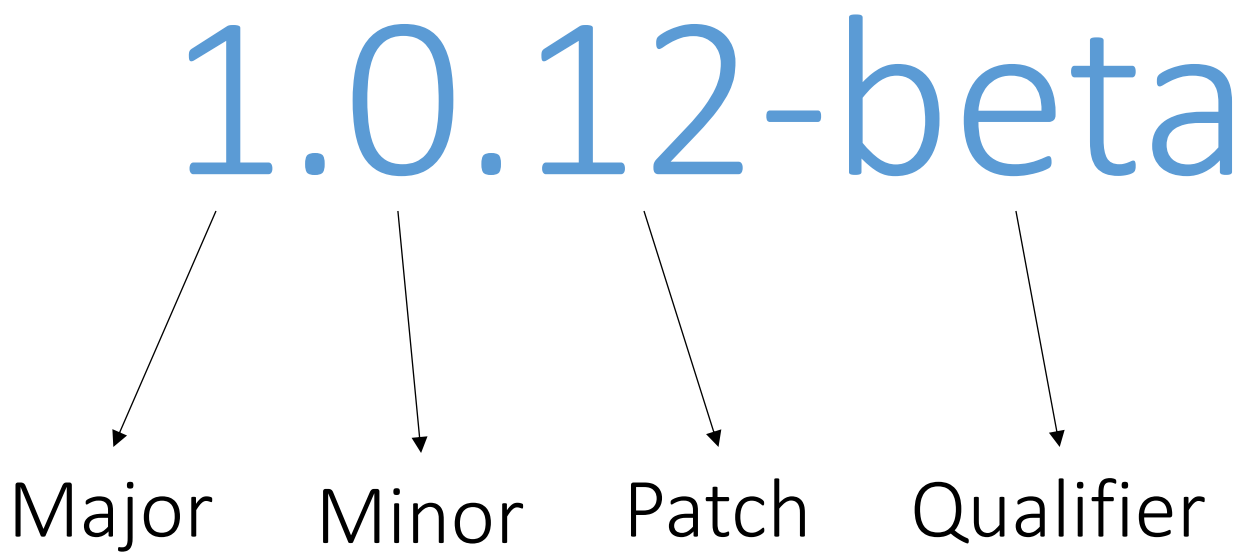
Spring TestContext 4.2.5-R

```
public class MockHttpServletRequest implements HttpServletRequest {
	@Override public String getAuthType() {
		return this.authType;
	}
	@Override public String getMethod() {
		return this.method;
	}
	[...]
}
```

MockHttpServletRequest must **implement** method HttpServletRequest.**changeSessionId**()

Semver (Semantic Versioning)

Can we trust version numbers in Maven?



Semantic Versioning and Impact of Breaking Changes in the Maven Repository

S. Raemaekers[a,b], A. van Deursen[b], J. Visser[c]

[a]*ING, Haarlemmerweg, Amsterdam, the Netherlands*
[b]*Technical University Delft, Delft, the Netherlands*
[c]*Software Improvement Group, Amsterdam, the Netherlands*

Abstract

Systems that depend on third-party libraries may have to be updated when updates to these libraries become available in order to benefit from new functionality, security patches, bug fixes, or API improvements. However, often such changes come with changes to the existing interfaces of these libraries, possibly causing rework on the client system. In this paper, we investigate versioning practices in a set of more than 100,000 jar files from Maven Central, spanning over 7 years of history of more than 22,000 different libraries. We investigate to what degree versioning conventions are followed in this repository. Semantic versioning provides strict rules regarding major (breaking changes allowed), minor (no breaking changes allowed), and patch releases (only backward-compatible bug fixes allowed). We find that around one third of all releases introduce at least one breaking change. We perform an empirical study on potential rework caused by breaking changes in library releases and find that breaking changes have a significant impact on client libraries using the changed functionality. We find out that minor releases generally have larger release intervals than major releases. We also investigate the use of deprecation tags and find out that these tags are applied improperly in our dataset.

Keywords: Semantic versioning, Breaking changes, Software libraries

1. Introduction

For users of software libraries or application programming interfaces (APIs), backward compatibility is a desirable trait. Without backward compatibility, library users will face increased risk and cost when upgrading their dependencies. In spite of these costs and risks, library upgrades may be desirable or even necessary, for example if the newer version contains required additional functionality or critical security fixes. To conduct the upgrade, the library user will need to know whether there are incompatibilities, and, if so, which ones.

Email addresses: stevenraemaekers@gmail.com (S. Raemaekers), arie.vandeursen@tudelft.nl (A. van Deursen), j.visser@sig.eu (J. Visser)

Raemaekers' Study

- Use of **semver** to signal API instability
- API evolution **impact on clients**
- Breaking changes defined in the **JLS**
- Snapshot of **Maven Central Repository** 2011

Semver **Questions** and **Findings**

Q1: How are **semver** principles applied in the MCR (in terms of **BCs**)?

- "**BCs** are **widespread** without regard for versioning principles."

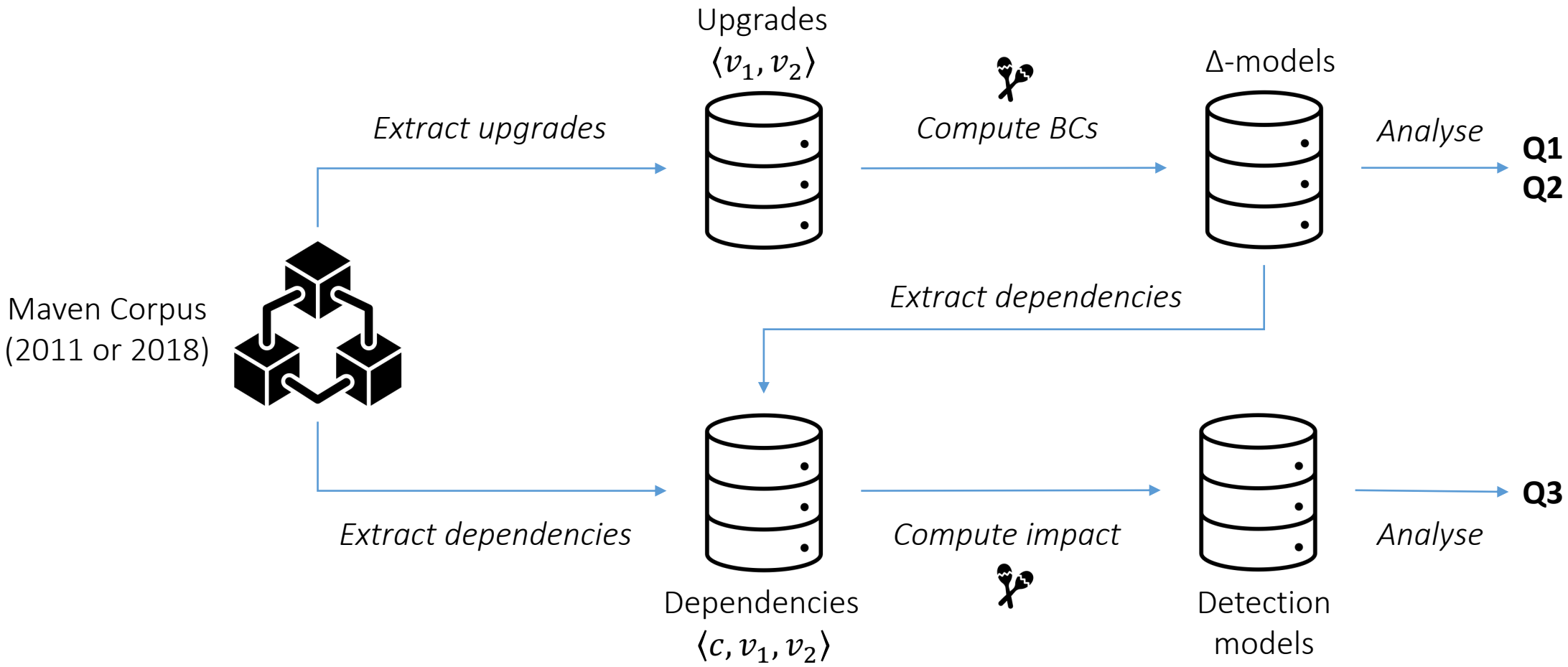
Q2: Has the adherence to semver **increased over time**?

- "The adherence to **semver** principles has **increased** over **time**."

Q3: What is the **impact** of BCs on **clients**?

- "BCs have a **significant impact** on clients."

Design of the Replication Study

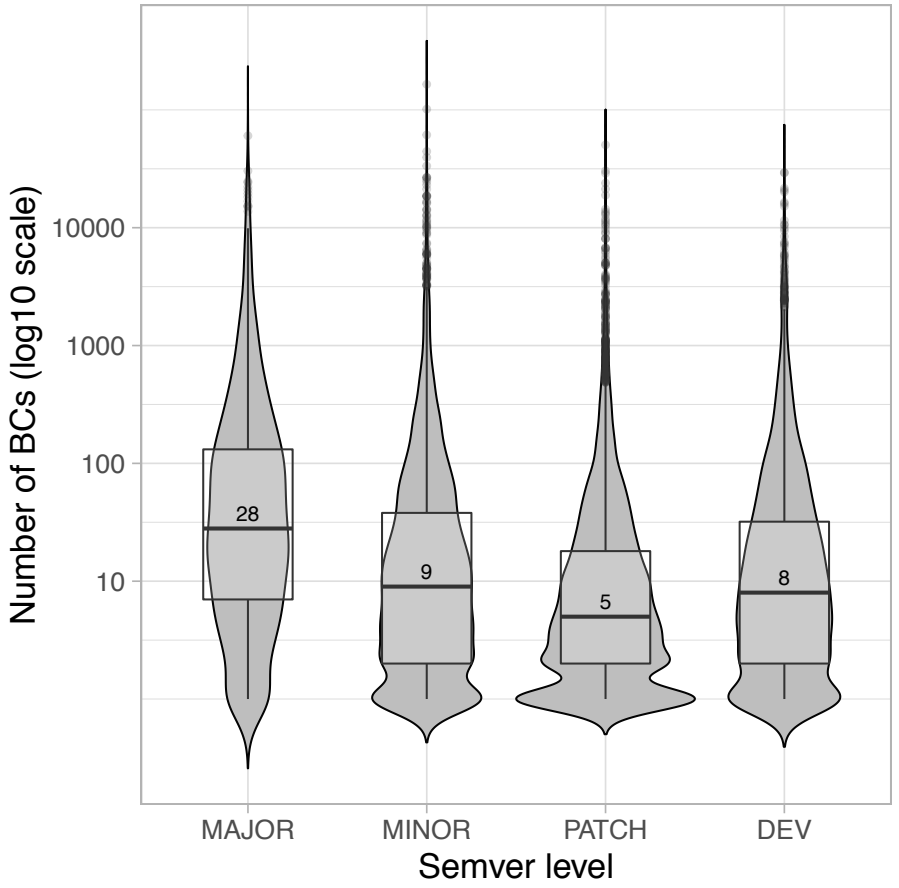


2011 Maven Dependency Dataset (MDD), 2018 Maven Dependency Graph (MDG)

Q1: How are **semver** principles applied in the MCR (in terms of **BCs**)?

Original study: "**BCs** are **widespread** without regard for versioning principles."

Non-major artefacts	Breaking
Original (2011)	29.0%
Replication (2011)	30.5%
Replication (2018)	20.1%

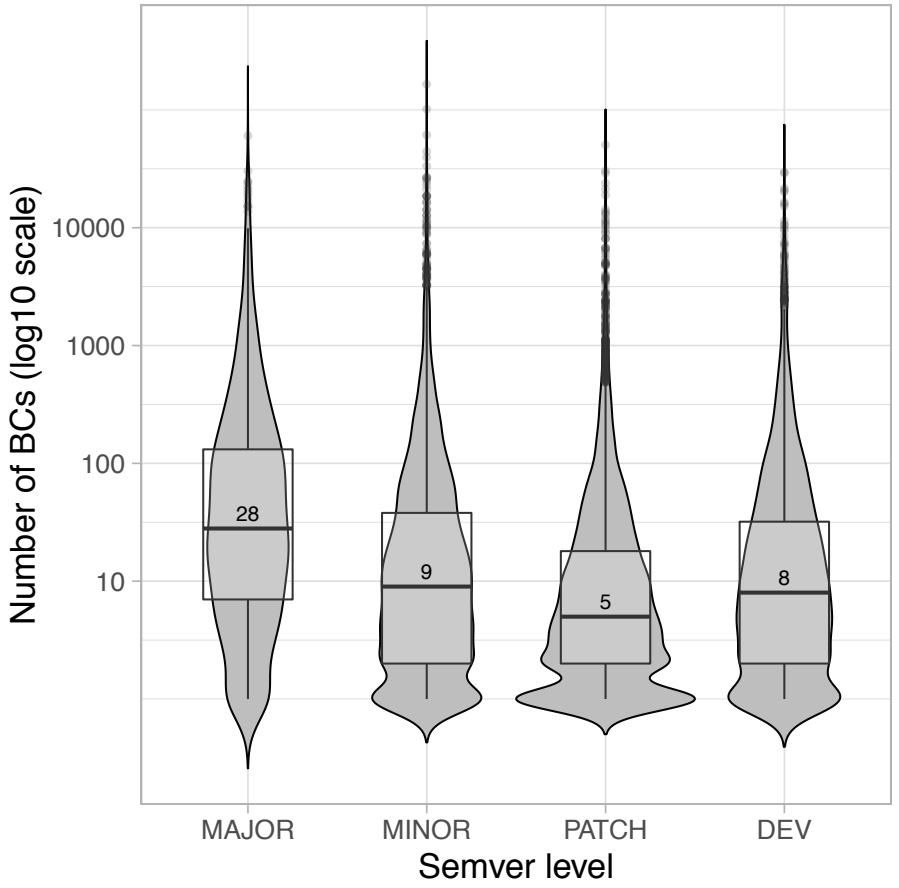


Q1: How are **semver** principles applied in the MCR (in terms of **BCs**)?

Conclusion: Semver principles are **not strictly applied** in practice, however they are **largely followed** (83.43% of all upgrades comply with semver).

≠

Non-major artefacts	Breaking
Original (2011)	29.0%
Replication (2011)	30.5%
Replication (2018)	20.1%



Q2: Has the adherence to semver **increased over time**?

Original study: "The adherence to **semver** principles has **increased** over **time**."

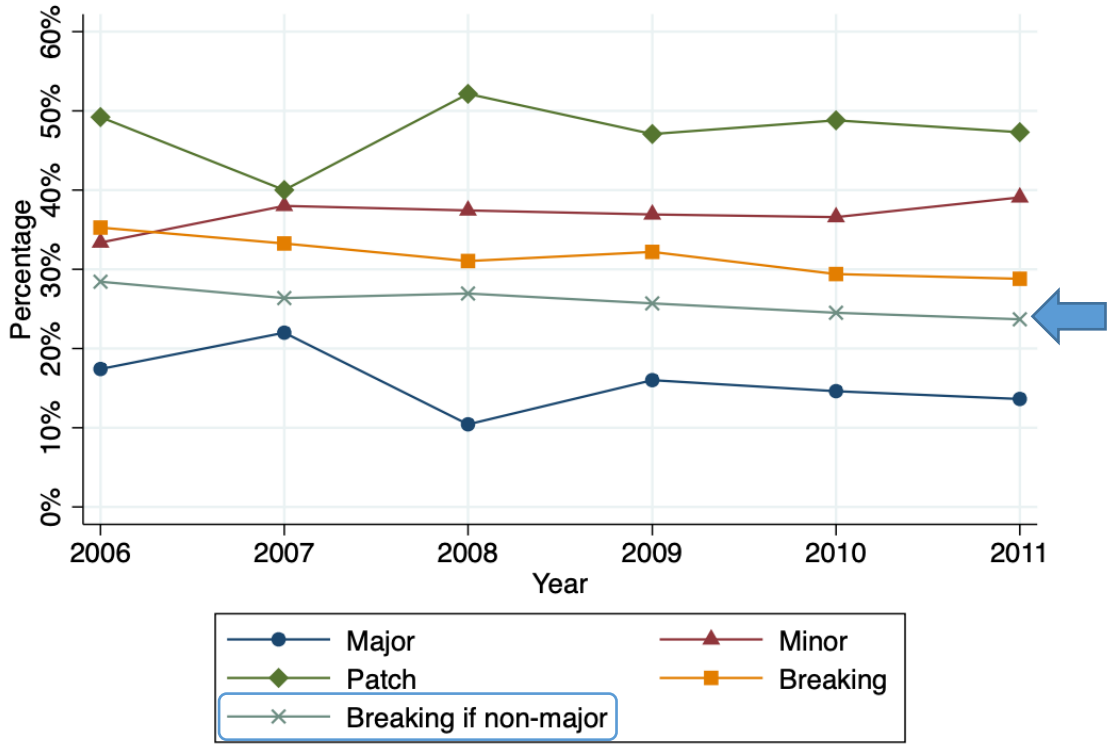
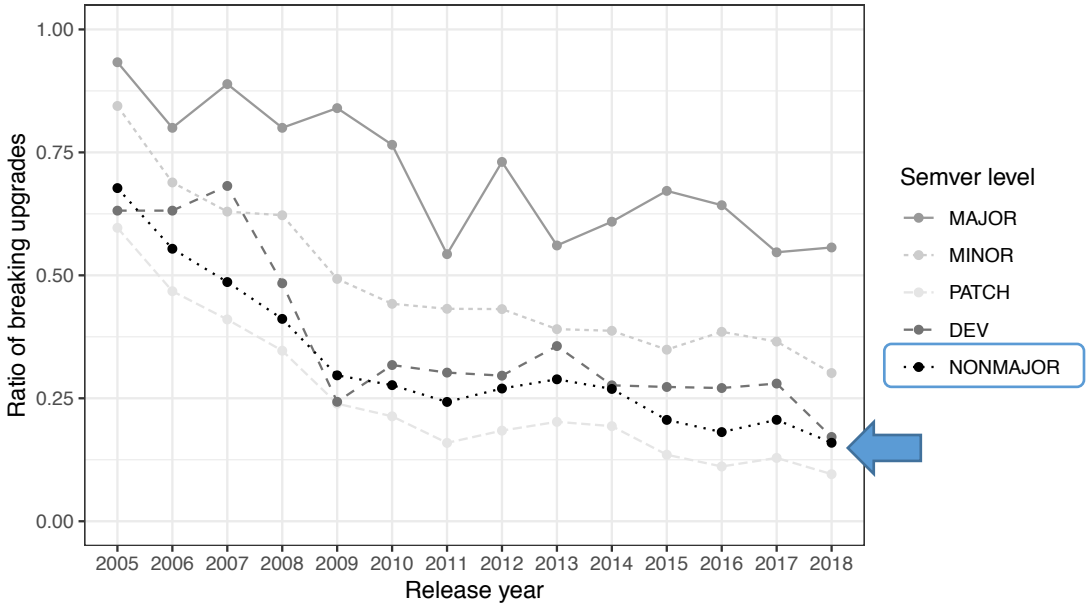


Image taken from "Semantic versioning and impact of breaking changes in the Maven repository" (Raemaekers et al., 2017)

Q2: Has the adherence to Semver **increased over time**?

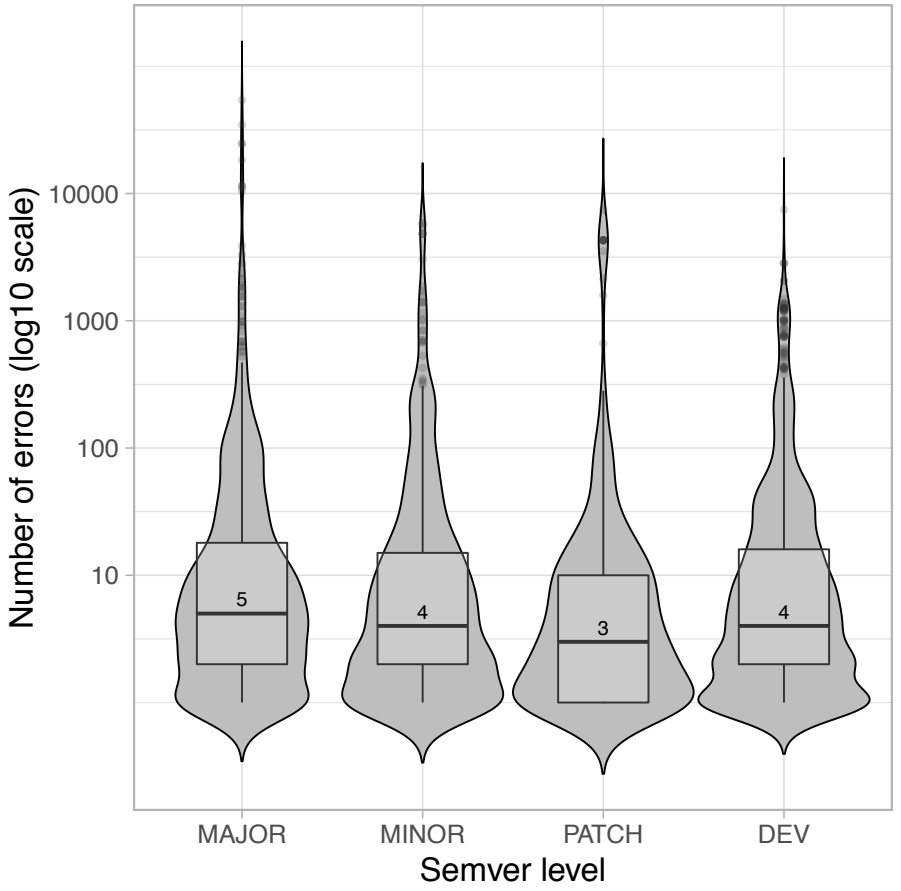
Conclusion: There is a **strong negative correlation** (r = -0.89) between the ratio of non-major breaking releases and time.



Q3: What is the **impact** of BCs on **clients**?

Original study: "BCs have a **significant impact** on clients."

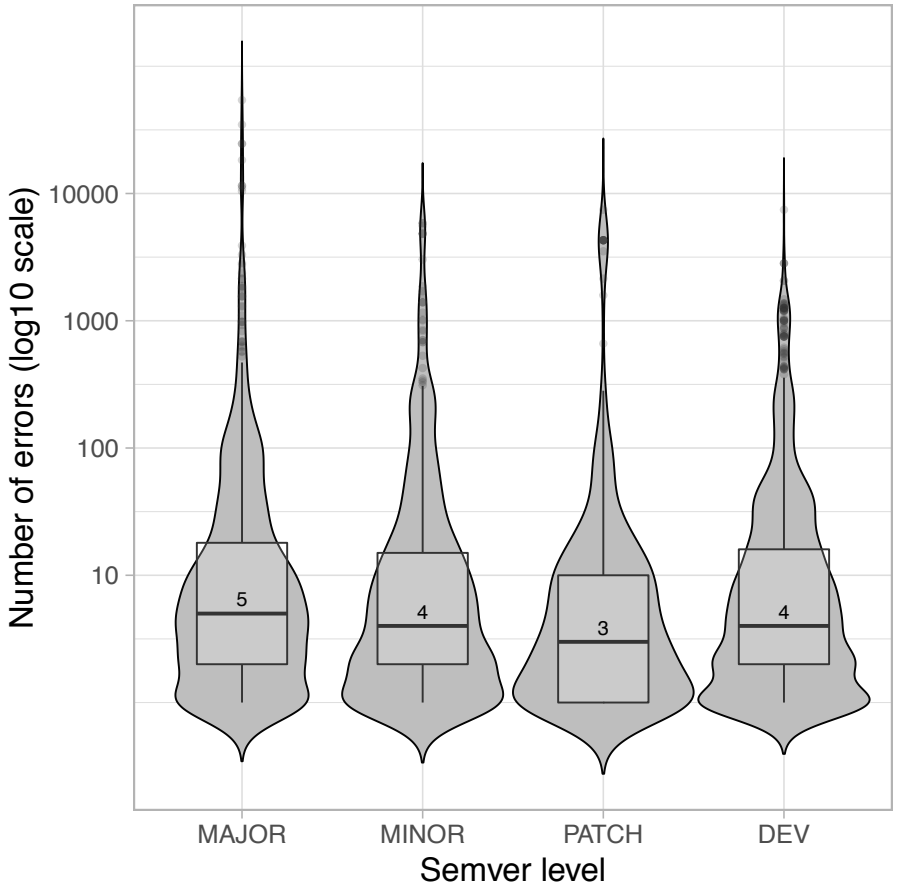
Level	Broken clients
Replication (2011)	6.12%
Replication (2018)	4.95%



Q3: What is the **impact** of BCs on **clients**?

Conclusion: In most cases, breaking declarations are **not used** by clients, which yields a **low number of broken clients**.

Level	Broken clients
Replication (2011)	6.12%
Replication (2018)	4.95%



Conclusions

Q1: How are **semver** principles applied in the MCR (in terms of **BCs**)?

- **83.43% of all upgrades** on MCR **comply** with semver principles.

Q2: Has the adherence to semver **increased over time**?

- The tendency to comply with semver practices has **significantly increased over time**.

Q3: What is the **impact** of BCs on **clients**?

- Only **4.97% of the clients** we analyse are **impacted** by BCs.

Breaking Bad? Semantic Versioning and Impact of Breaking Changes in Maven Central

Contact:

Lina Ochoa: l.m.ochoa.venegas@tue.nl

Thomas Degueule: thomas.degueule@labri.fr

Jean-Rémy Falleri: falleri@labri.fr

Jurgen Vinju: jurgen.vinju@cwi.nl

Data and code availability:

https://github.com/tdegueul/maven-methodo

Maracas:

https://github.com/crossminer/maracas