Beyond Duplicates:
Towards Understanding and Predicting Link Types in Issue Tracking Systems

Clara Marie Lüders, Abir Bouraffa, Walid Maalej

MOTIVATION

- Issue networks are **large** and **complex**
- Linking issues is a **pain point** for stakeholders [2], manually linking is harder with many issues
- Duplicates are a popular and well studied link type, but there are many more link types

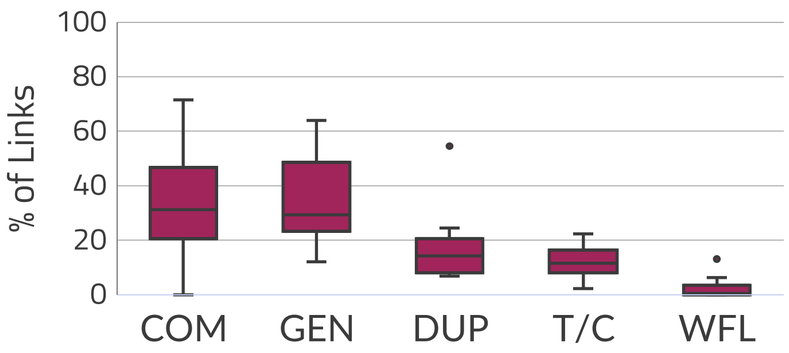
DATASET [1]
15 JIRA Repositories
171 Link Types
2M Issues
600k Links

RQ1 - USAGE

How are various link types used in practice, particularly in term of prevalence and structural properties?

CATEGORIZATION

Duplication
Composition
Temporal / **C**ausal
Work**fl**ow
General Relation



STRUCTURE

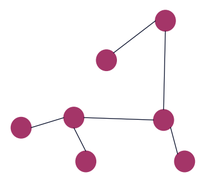
Complexity

Isolated: issues without links
2Comp: components with 2 issues
3Comp+: components with 3 or more issues
Avg. Density:

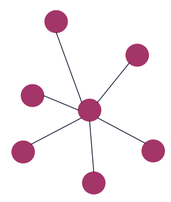
$$\text{Density} = \frac{number\ of\ Links}{maximum\ number\ of\ links}$$

Shape

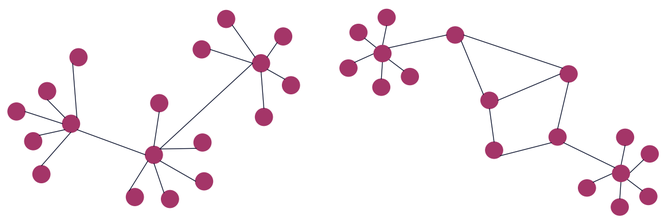
Tree



Stars



Assortativity
Pearson Correlation of degree of nodes at both ends of a link

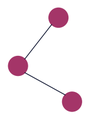


Assortative

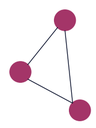
Disassortative

Transitivity

$$= \frac{3 * number\ of\ triangles}{number\ of\ triads}$$



Intransitive



Transitive

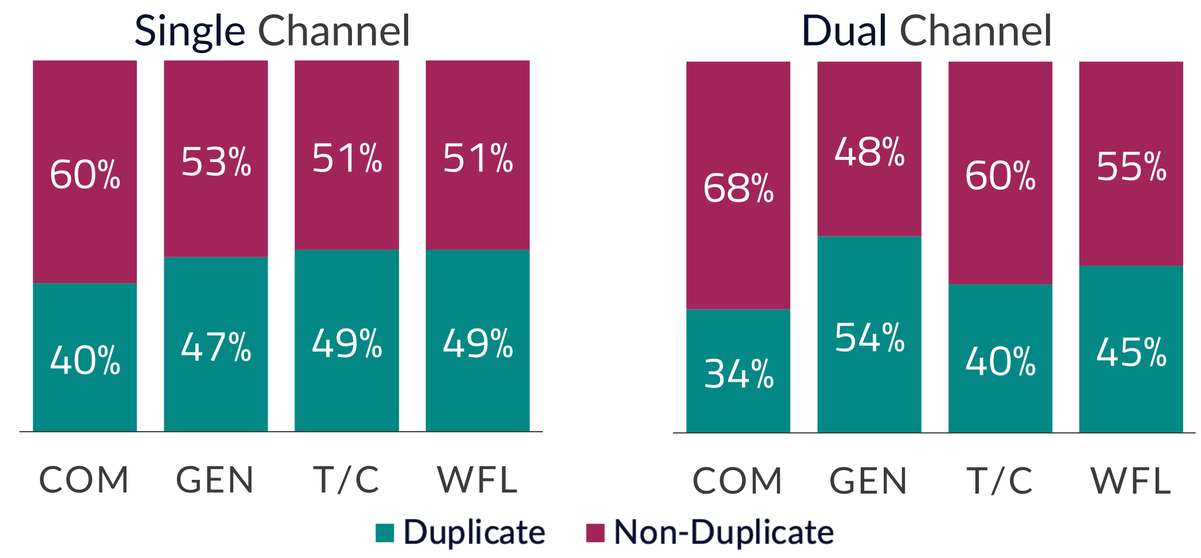
Category	Complexity					3C+ Shape		Asso.	Tran.
	%Iso.	%2C	%3C+	Den.	#Links	%Tree	%Star		
Rel.	85.6	70.7	29.3	**.562**	218833	**81.9**	**62.8**	**.025**	**.121**
Dup.	92.6	82.8	17.2	.558	98178	88.5	76.4	.003	.050
T/C.	94.7	73.5	26.5	.496	84236	84.4	63.7	-.049	.048
Com.	85.0	**41.2**	**58.8**	**.394**	190913	**97.7**	**90.1**	**-.260**	**.002**
WFl	98.9	**84.5**	**15.5**	.490	15048	82.5	75.6	-.098	.003

RQ2 - DETECTION

Are state-of-the-art models [3] for duplicate detection able to distinguish between Duplicate and other links?

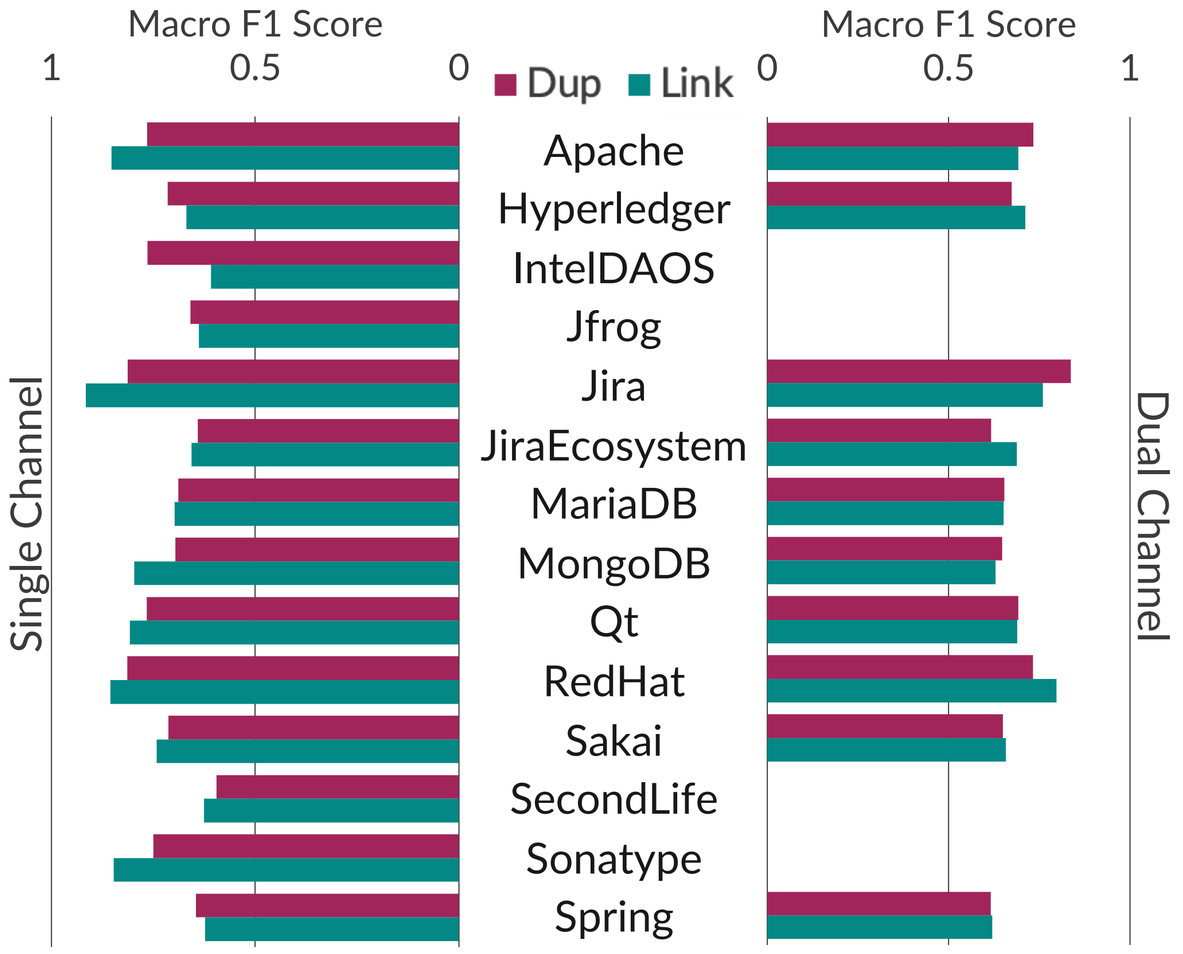
ROBUSTNESS

Do SotA models correctly classify other link types?



ADAPTABILITY

Do SotA models detect links or duplicates?



CONTRIBUTIONS

- **Taxonomy** of Link Types
- **Usage trends** of link types in Issue Tracking Systems
- Insights for **data quality** checking for Issue Tracking Systems
- Evidence that current **duplicate detection models** should be trained differently to distinguish other link types

[1] Lloyd Montgomery, Clara Lüders, and Walid Maalej. 2022. *An Alternative Issue Tracking Dataset of Public Jira Repositories*. In 2022 IEEE/ACM 18th International Conference on Mining Software Repositories (MSR '22). https://doi.org/10.1145/3524842.3528486
[2] Davide Fucci, Cristina Palomares, Xavier Franch, Dolors Costal, Mikko Raatikainen, Martin Stettinger, Zijad Kurtanovic, Tero Kojo, Lars Koenig, Andreas Falkner, Gottfried Schenner, Fabrizio Brasca, Tomi Männistö, Alexander Felfernig, and Walid Maalej. 2018. *Needs and Challenges for a Platform to Support Large-Scale Requirements Engineering: A Multiple-Case Study*. In Proceedings of the 12th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM '18). https://doi.org/10.1145/3239235.3240498
[3] Jianjun He, Ling Xu, Meng Yan, Xin Xia, and Yan Lei. 2020. *Duplicate Bug Report Detection Using Dual-Channel Convolutional Neural Networks*. In Proceedings of the 28th International Conference on Program Comprehension Association for Computing Machinery (ICPC '20). https://doi.org/10.1145/3387904.3389263

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG