Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece

Kyriakos C. Chatzidimitriou, Michail D. Papamichail, Themistoklis Diamantopoulos, Napoleon-Christos I. Oikonomou, and Andreas L. Symeonidis

{kyrcha, mpapamic, thdiaman, napoleon.oikonomou}@issel.ee.auth.gr, asymeon@eng.auth.gr

npm Packages as Ingredients: a Recipe-based Approach

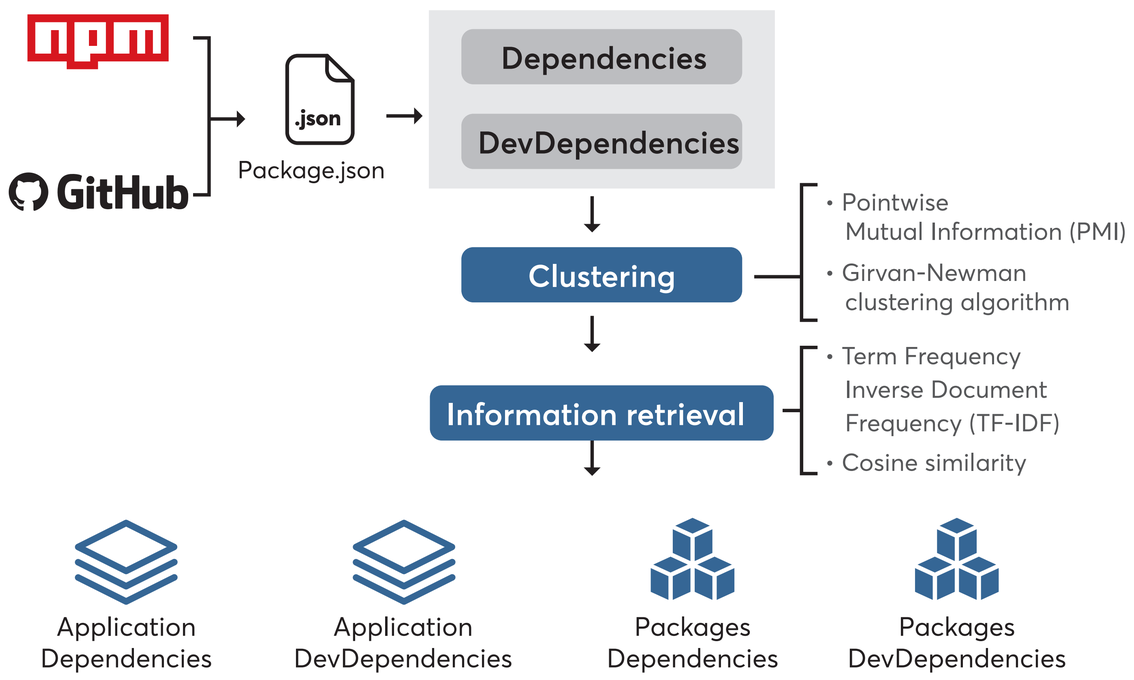
INTRODUCTION – MOTIVATION

- **JavaScript:** One Programming Language to Rule them All
- **Npm-registry** *(with more than 800K packages)* facilitates the reuse-oriented software development paradigm and constitutes micropackaging as state-of-the-practice approach.
- **80% of the source code** in today's applications comes from libraries and frameworks.
- There are one-liner libraries that have more that **70 dependencies.**

In this work, we mine software repositories in order to:

1. Take advantage of the collective knowledge residing in online software repositories (e.g. GitHub) and extract packages communities.
2. Treat dependencies as ingredients and packages (or applications) as recipes (inspired by Teng et al., 2012)
3. Investigate the scope of the dominant communities that appear in the npm ecosystem and analyze the similarities (or differences) between package recipes and application recipes.
4. Extract valuable information regarding the development trends as reflected in the dominant devDependencies communities.

METHODOLOGY



$$PMI = log\frac{p(a,b)}{p(a) * p(b)} \quad (1)$$

$$p(a,b) = \frac{\#\ of\ packages\ containing\ a\ and\ b}{\#\ of\ packages} \quad (2)$$

$$p(a) = \frac{\#\ of\ packages\ containing\ a}{\#\ of\ packages} \quad (3)$$

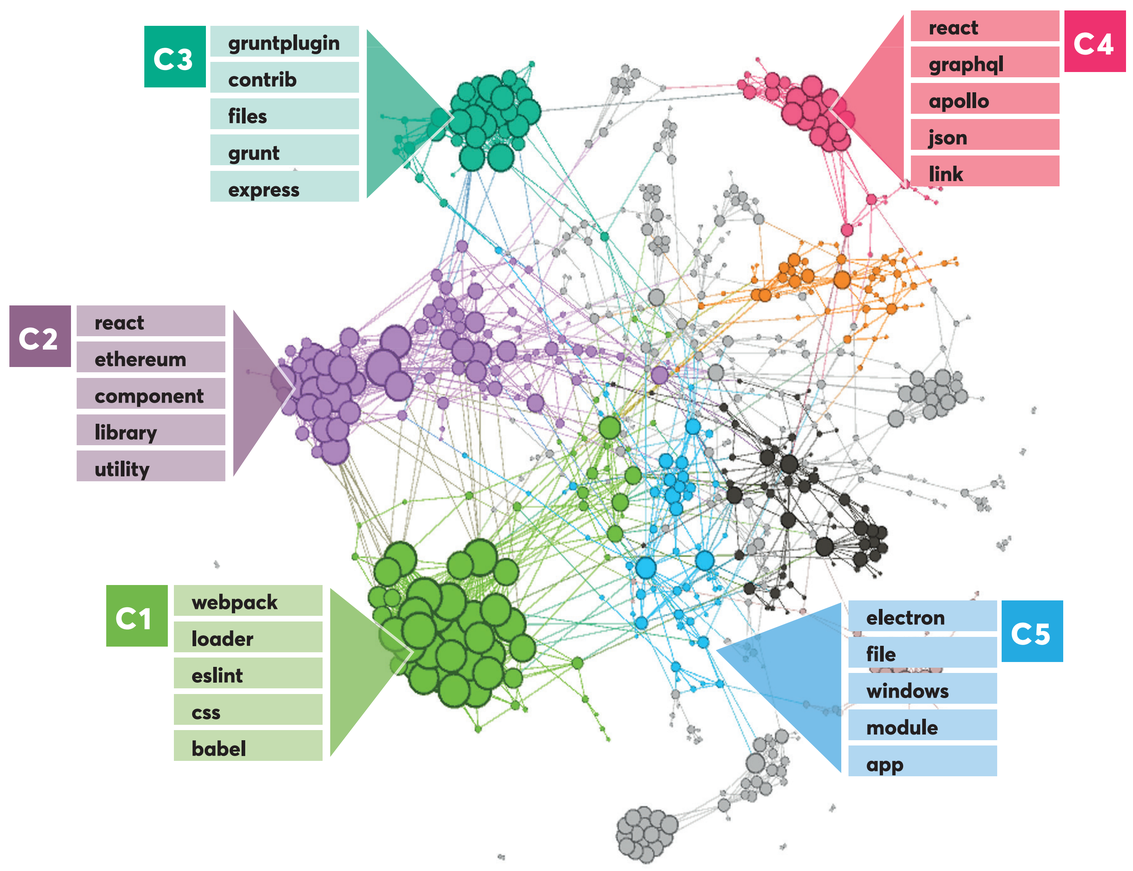
$$p(b) = \frac{\#\ of\ packages\ containing\ b}{\#\ of\ packages} \quad (4)$$

ICSOFT 2019 : 14th International Conference on Software Technologies

DATASET

Metric	Value
Number of packages	8,732
Number of applications	13,884
Mean number of Dependencies per package	6.3
Mean number of DevDependencies per package	11.8
Mean number of Dependencies per application	5.6
Mean number of DevDependencies per application	9.7

APPLICATIONS DEPENDENCY NETWORK



EVALUATION

- ✓ The identified communities illustrate the latest trends in the development of both packages and applications.
- ✓ We apply graph analysis and information retrieval techniques (TF-IDF, cosine similarity) in order to examine whether the formulated clusters are reasonable from a software engineering perspective.
- ✓ The graph analysis is based on the **degree**, the **betweenness centrality** and the **closeness centrality** of the packages included in all communities.
- ✓ **webpack-dev-server** and **react-hot-loader** packages exhibit the highest degree values on packages dependencies network, while **object.values** package exhibits the highest betweenness centrality.

ACKNOWLEDGEMENTS

This research has been co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project code: T1EDK-02347).

issel
Intelligent Systems & Software Engineering Labgroup

CONTACT INFO
Website: https://issel.ee.auth.gr
e-mail: mpapamic@issel.ee.auth.gr