

Measuring the Impact of Open Source Software Innovation Using Network Analysis on GitHub Hosted Python Packages

Team Members: *Derek Banks, Camille Leonard, Shilpa Narayan, Nick Thompson*
Client and Sponsor: *Gizem Korkmaz, Brandon Kramer*
Faculty Advisor: *Abbas Kazemipour*

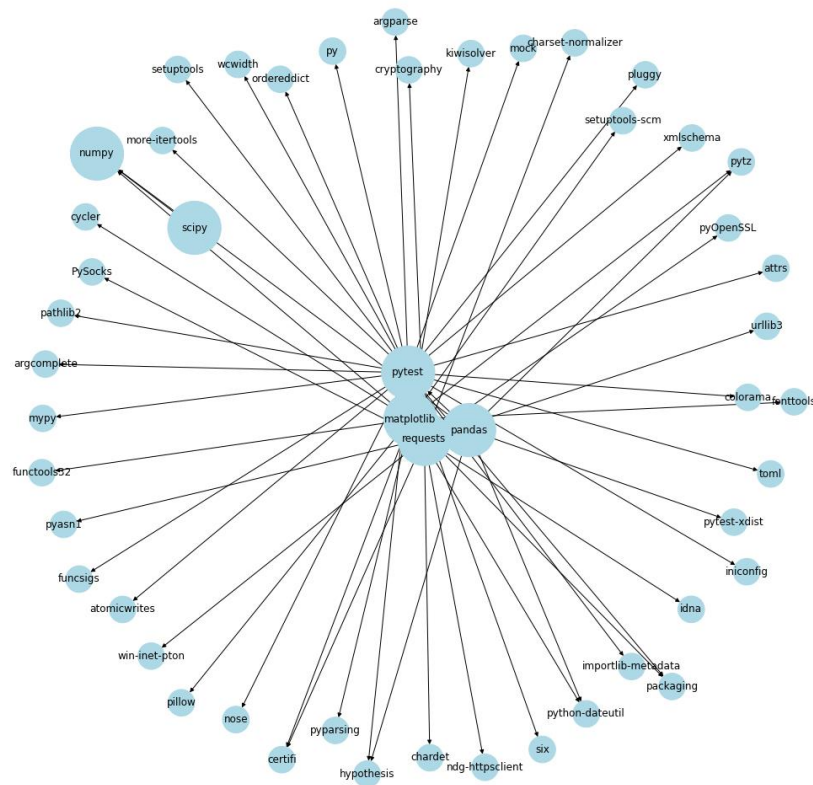
Background and Problem Statement

- NSF Focus - no established method to quantify the impact of open source ecosystems
- Develop a framework to measure the impact of Open Source Software (OSS) development which can be applied to any open source ecosystem eg: citations
- Goal is to identify influential actors (e.g. packages, developers, countries) by using the following impact measures:
 - Centrality - network measures
 - Downloads - OSS measure
 - Development cost leveraged through dependencies

Data

- Package and download information sourced from PyPI through Google Big Query
- Data limited to Python packages hosted on GitHub with an OSS license
 - 72% of the PyPI data
- Package information was joined with commits data scraped from GitHub
- 68,992 unique packages were analyzed
 - Reduced from 75,000 when joined with GitHub commits data

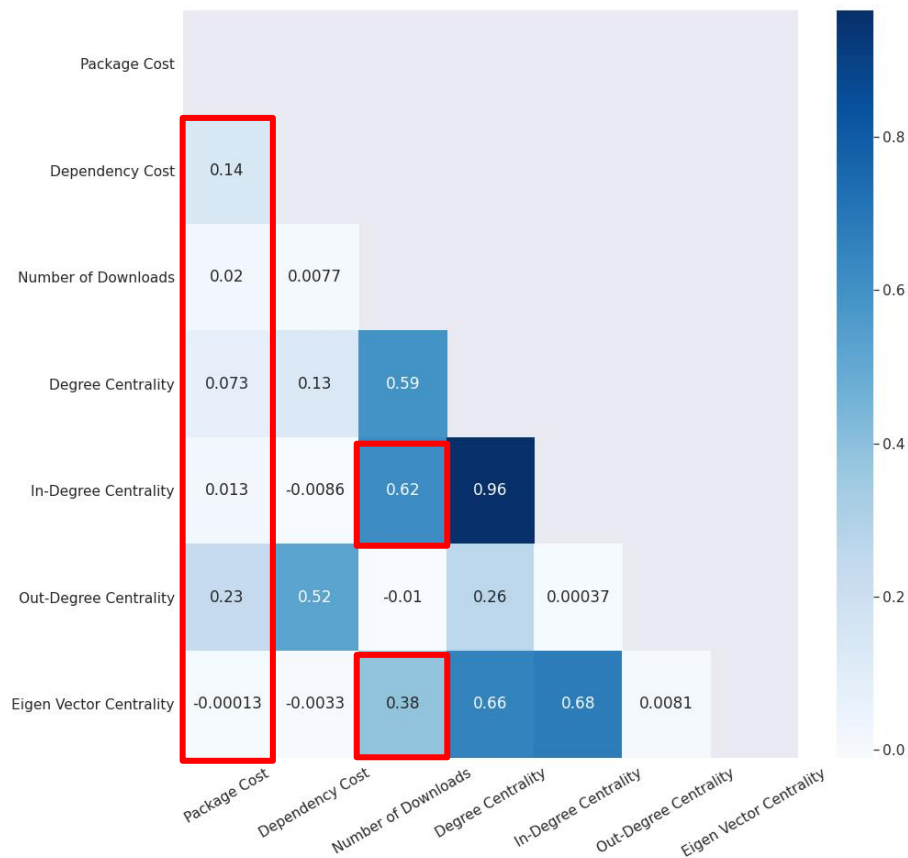
Created Dependency Network Model



Methods

- Calculated in-degree, out-degree, and eigenvector centrality measures
- Calculated package cost using breadth first search and the COCOMO II model
 - $\text{Effort} = 2.4(\text{KLOC})^{1.05}$
 - $\text{Nominal development time} = 2.5(\text{Effort})^{0.38}$
 - $\text{Development cost} = (\text{Monthly Resource Cost}) (\text{Nominal development time})$
- Aggregated downloads per package for use as an impact measure

Metric Correlations



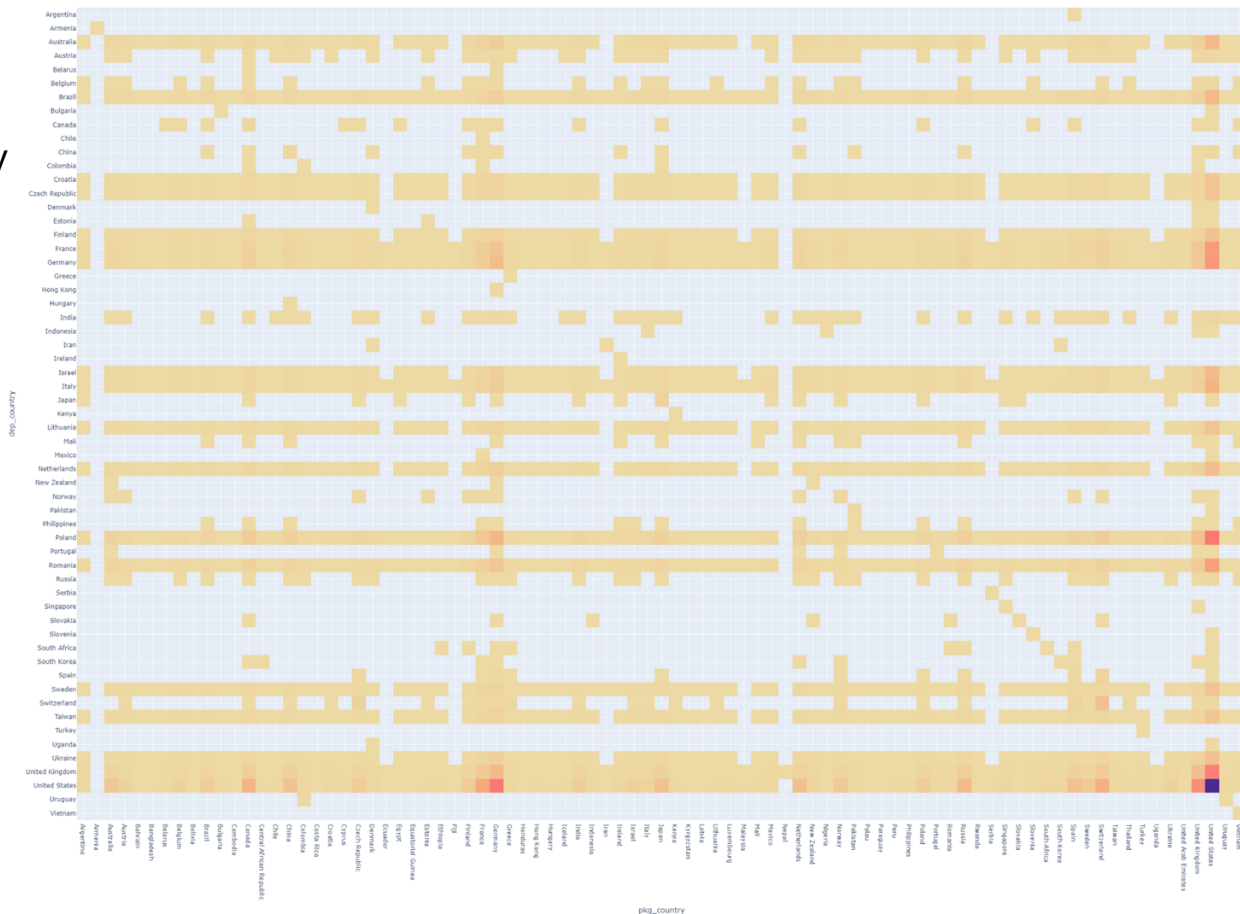
Influential Packages

MOST INFLUENTIAL PACKAGES RANKED BY CENTRALITY AND
DOWNLOADS

Package	Degree Centrality	In-Degree Centrality	Eigenvector Centrality	Downloads Ranked
six	1	1	1	1
pyyaml	2	2	33	5
pytest-cov	3	3	3	9,810
coverage	4	4	8	5,678
flask	5	5	154	8,477
twine	6	6	2	9,757
toml	21	20	4	28,586
releasecmd	936	273	5	46,570

Future Work - Technology diffusion

Country where the dependency package is developed



cost



Country where the package is developed



Thank you!

Questions ?