



Collaboration as Health Indicator

CHAOScon EU 2022

Software
Development
Analytics

Daniel Izquierdo Cortázar
(in collaboration with Miguel
Ángel Fernández)

Collaboration (from Latin com- "with" + laborare "to labor", "to work") is the process of two or more people, entities or organizations working together to complete a task or achieve a goal.

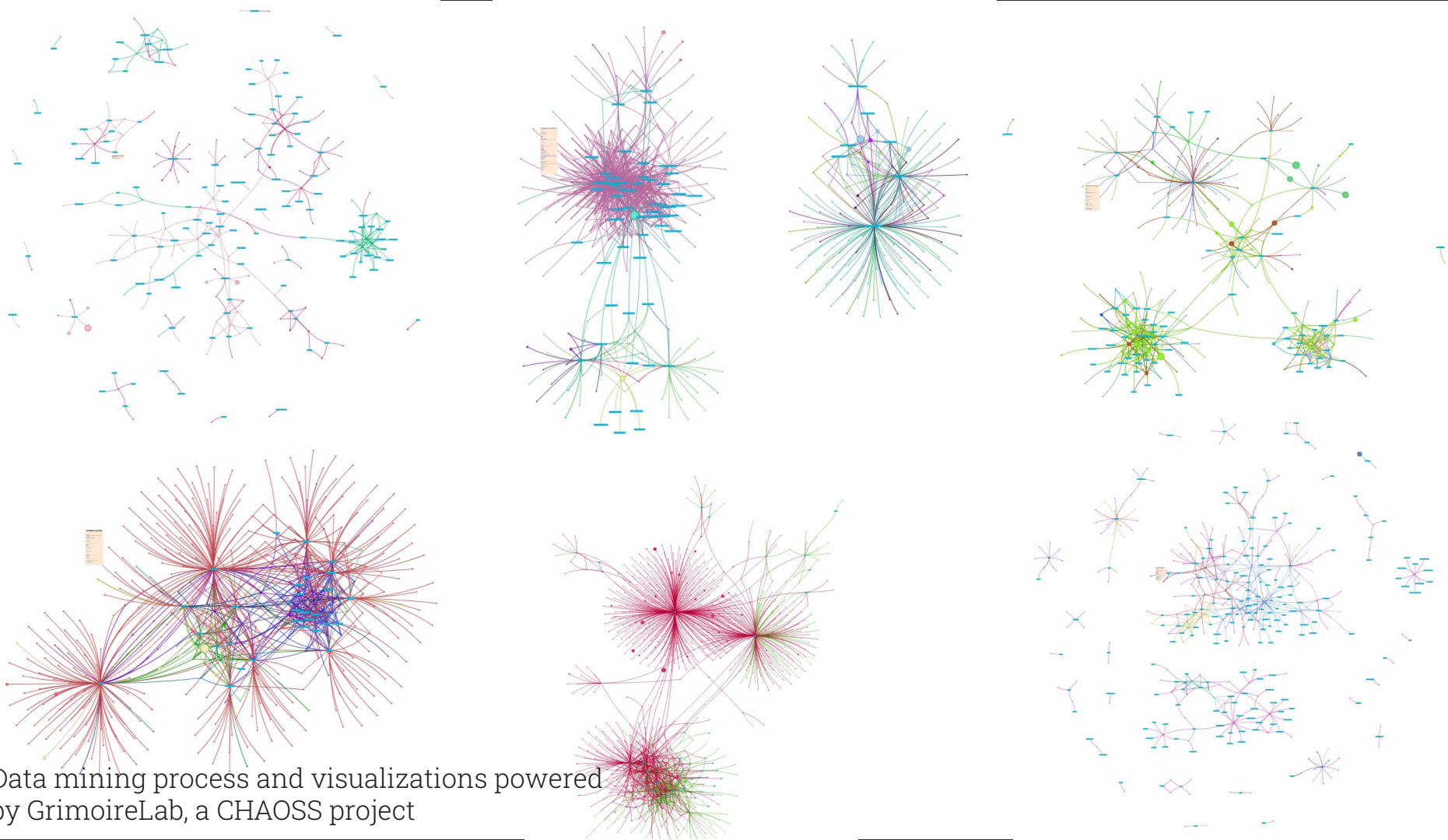
Wikipedia dixit



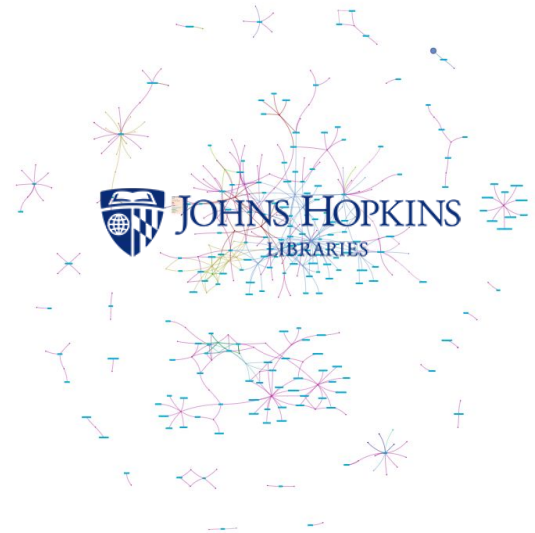
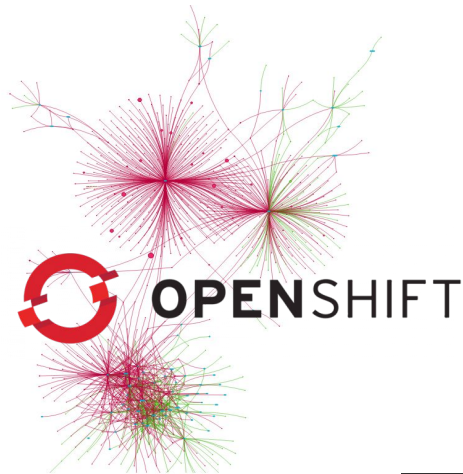
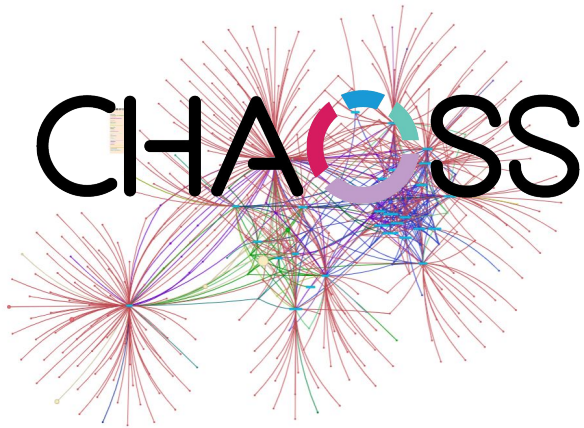
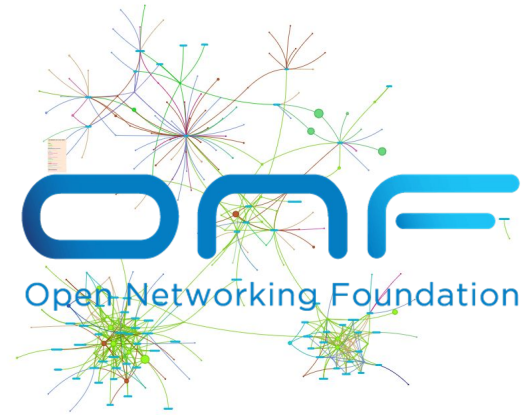
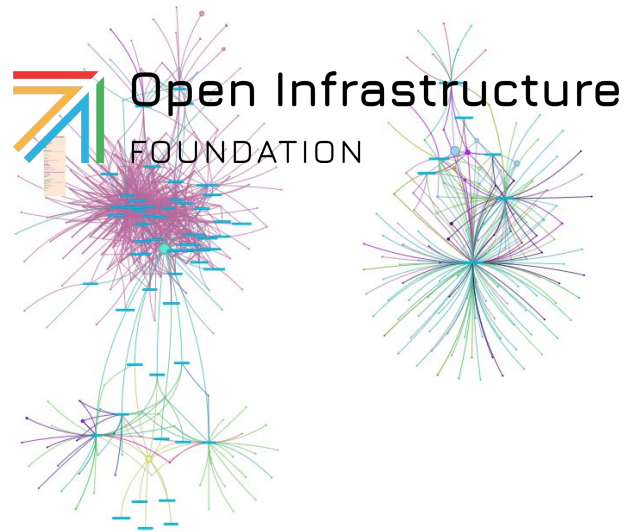
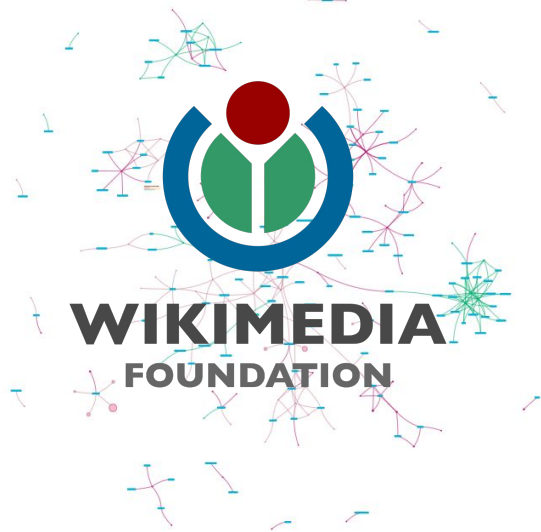


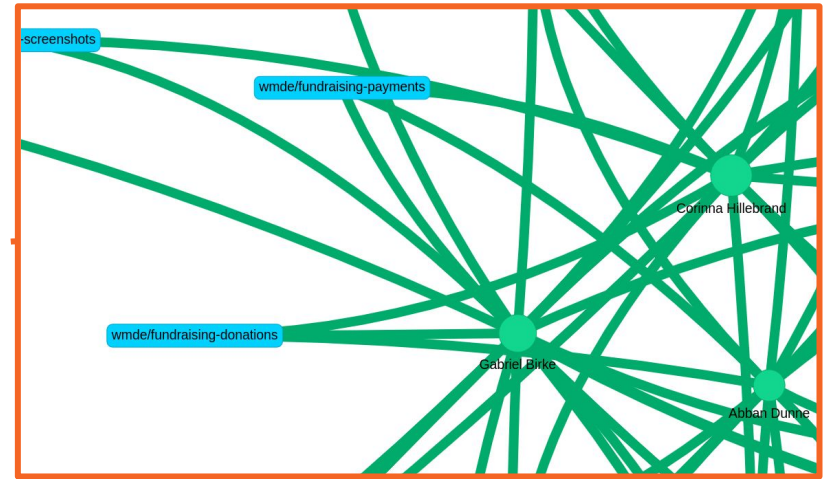
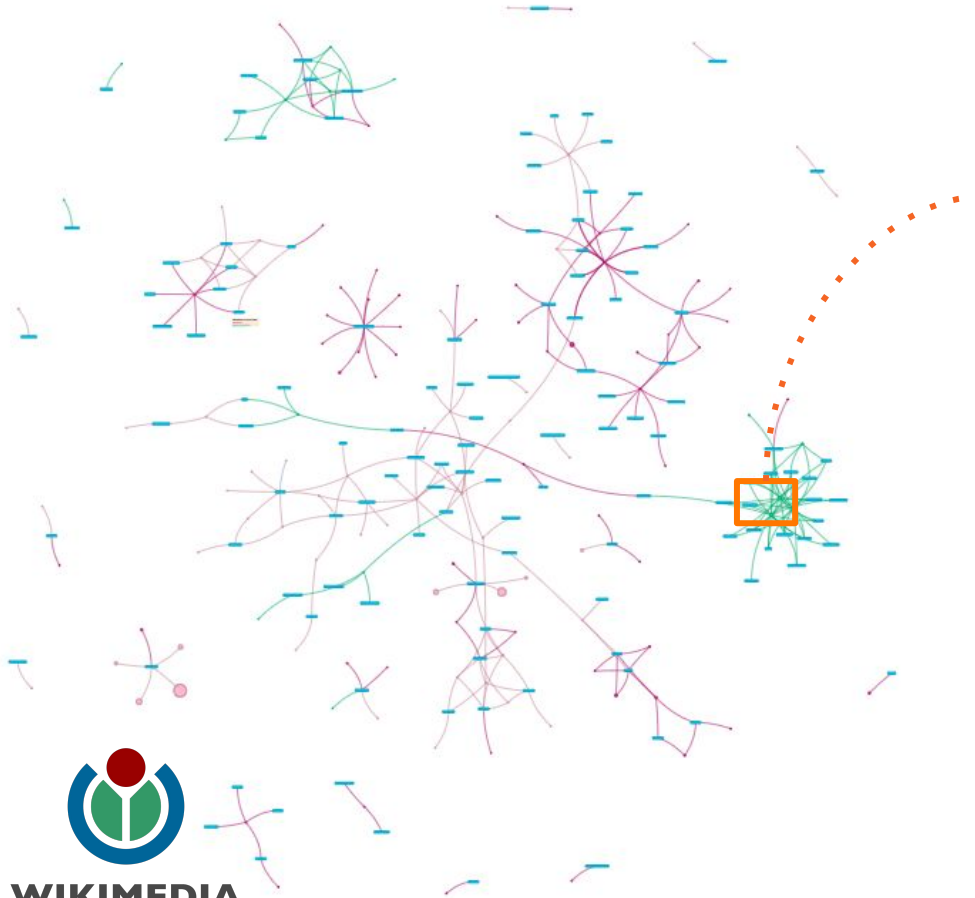
OPEN

Welcome to Open Source
Communities!

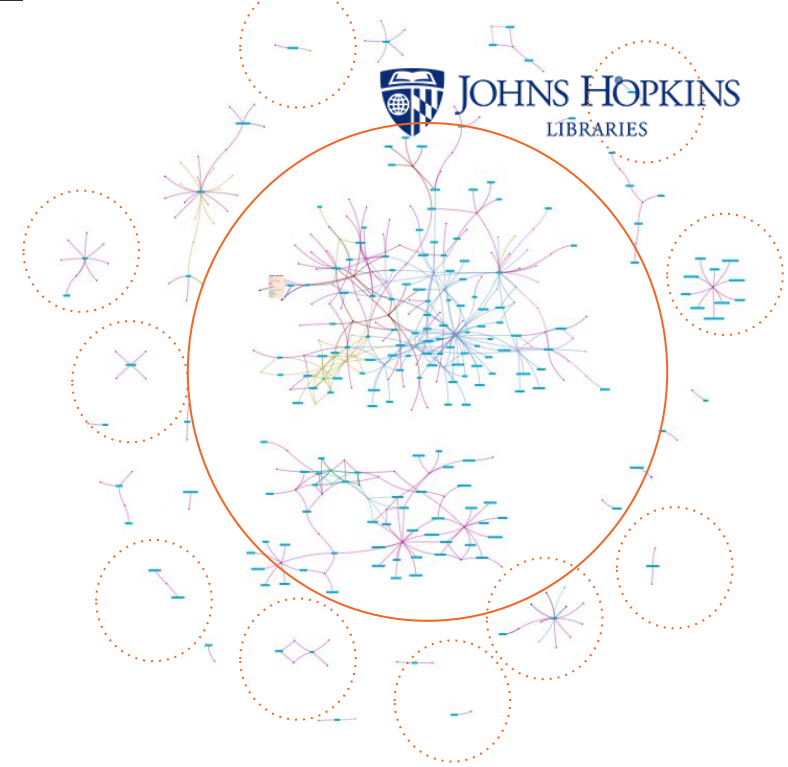
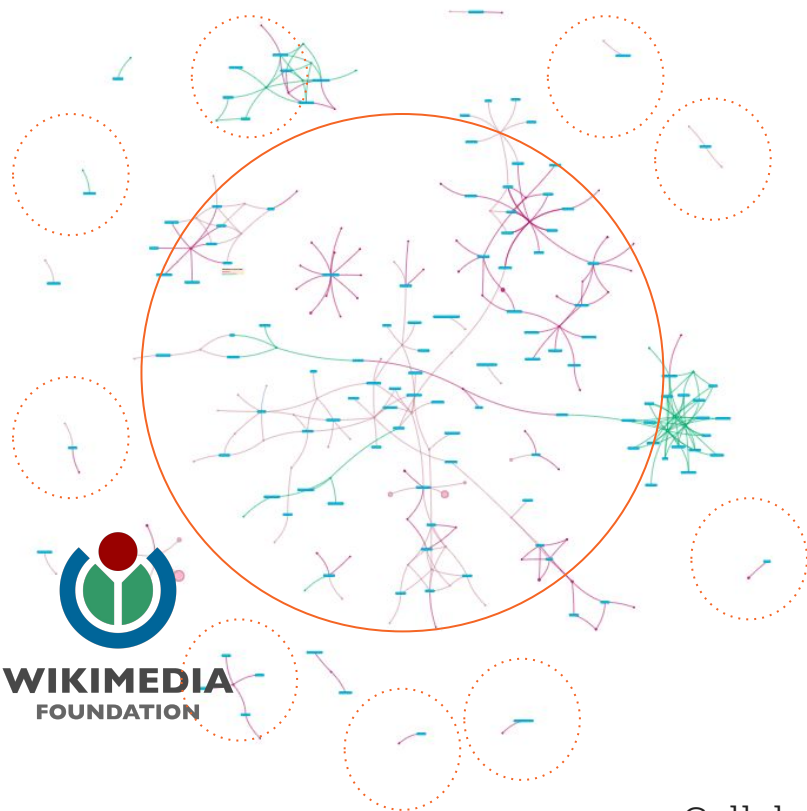


Data mining process and visualizations powered
by GrimoireLab, a CHAOSS project

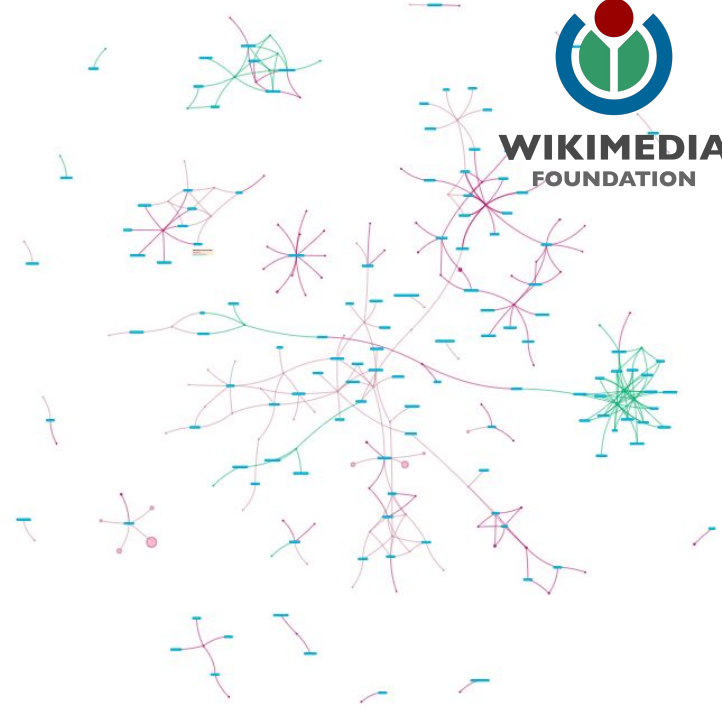
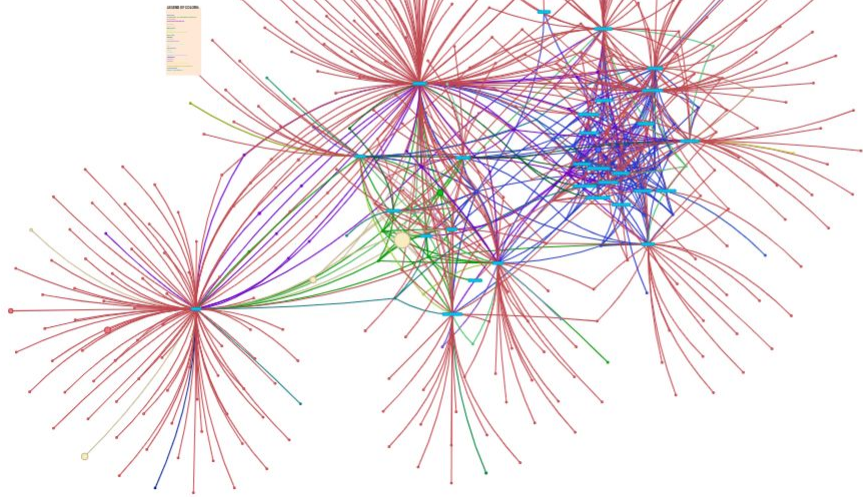




Dots are developers
Squares are repositories
Edge exists if a developer has
contributed to a repository

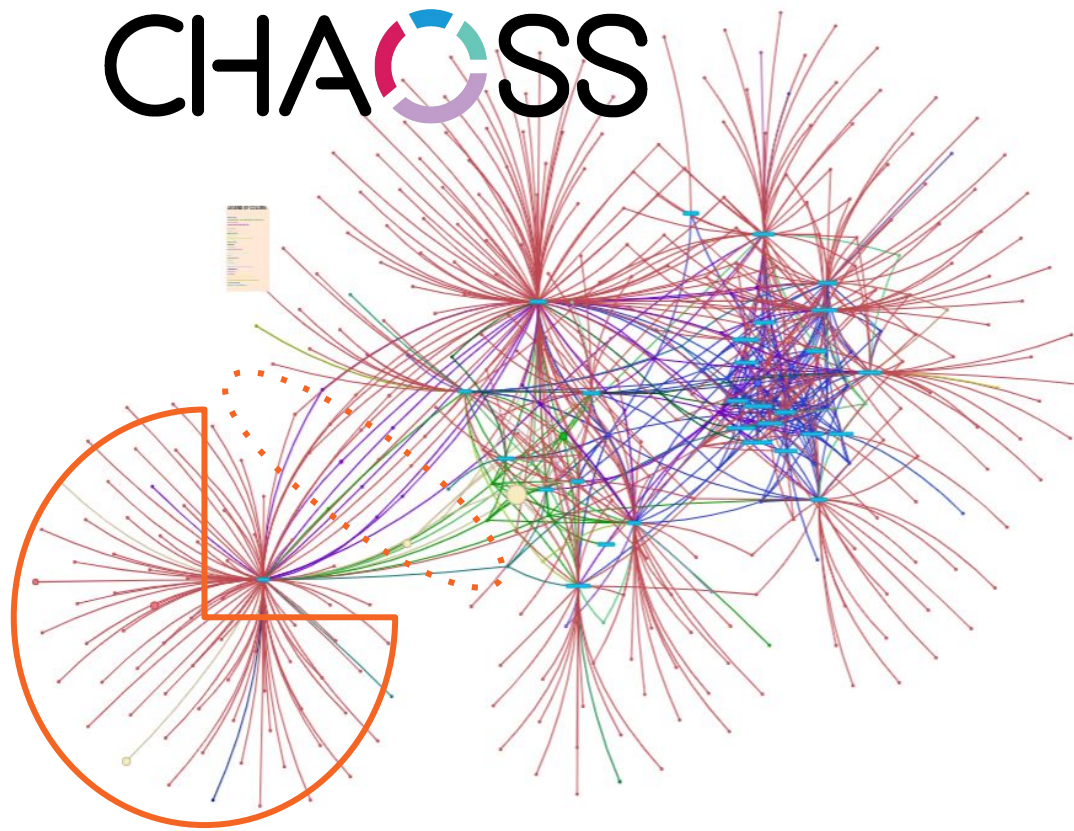


Collaboration vs Isolated Projects

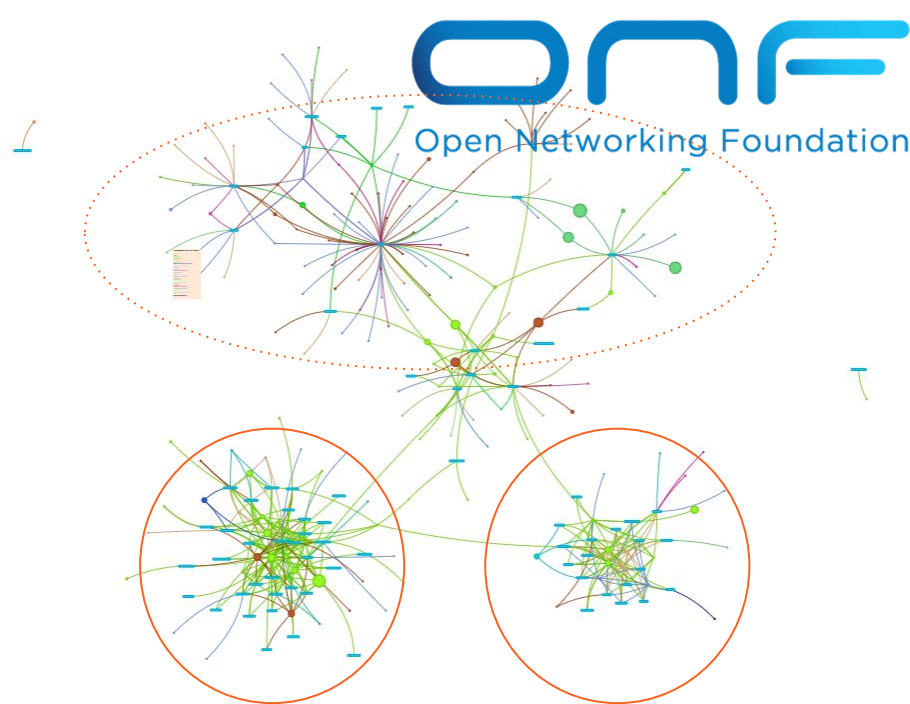
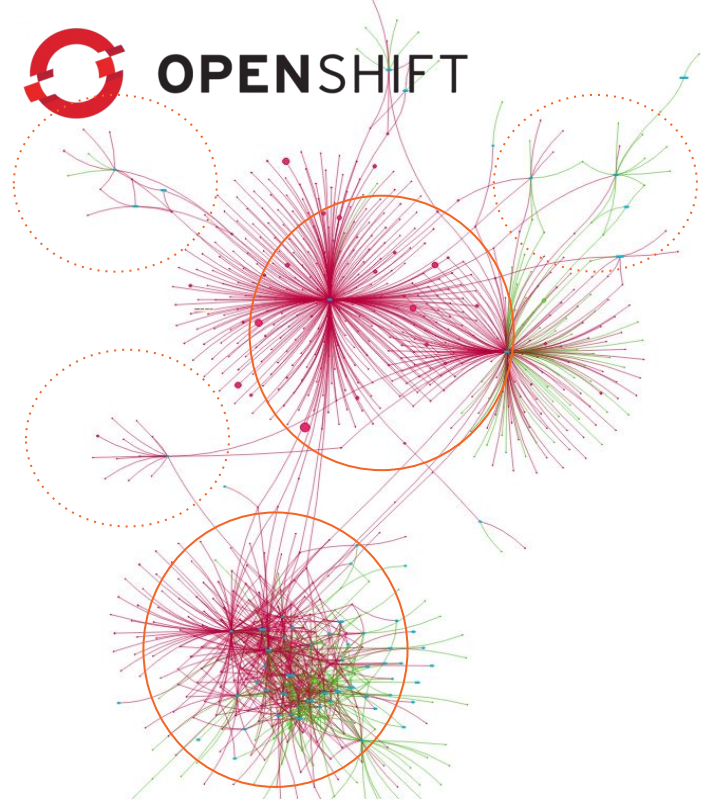


'Continent' communities vs Archipelago

CHAOS



1 Project Developer vs Many Projects Developer

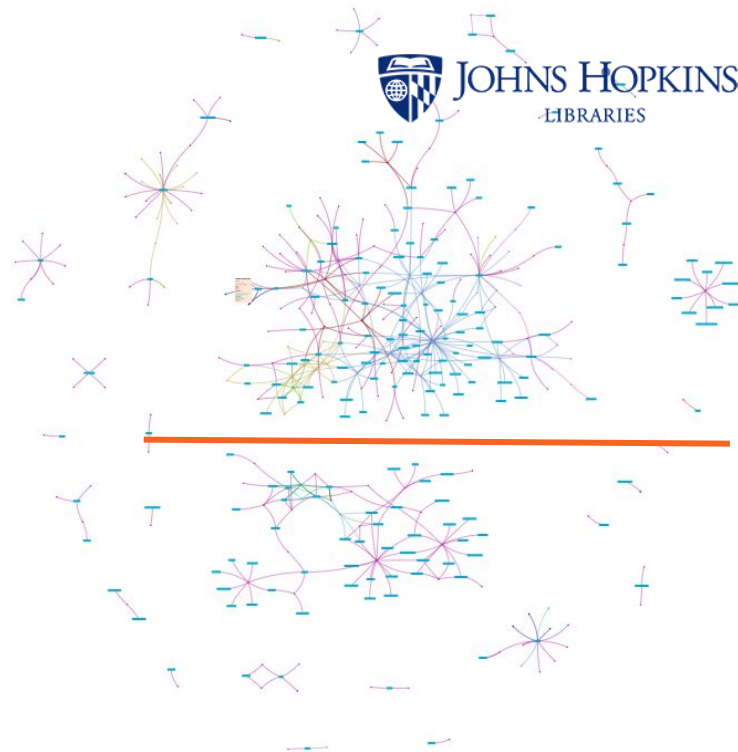
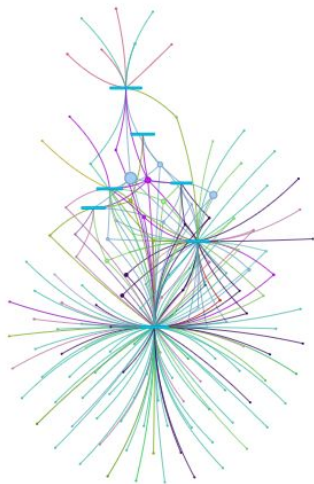
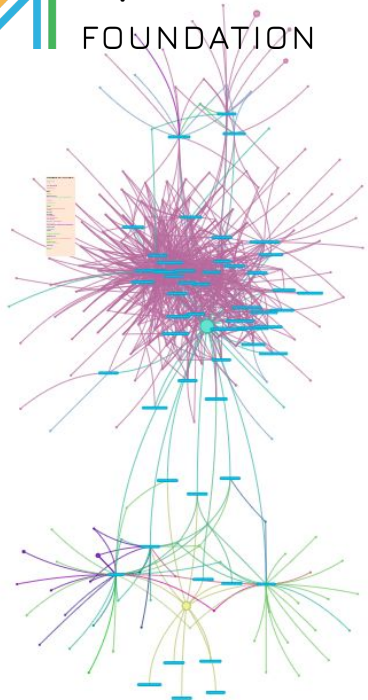


High density areas vs lighter ones



Open Infrastructure

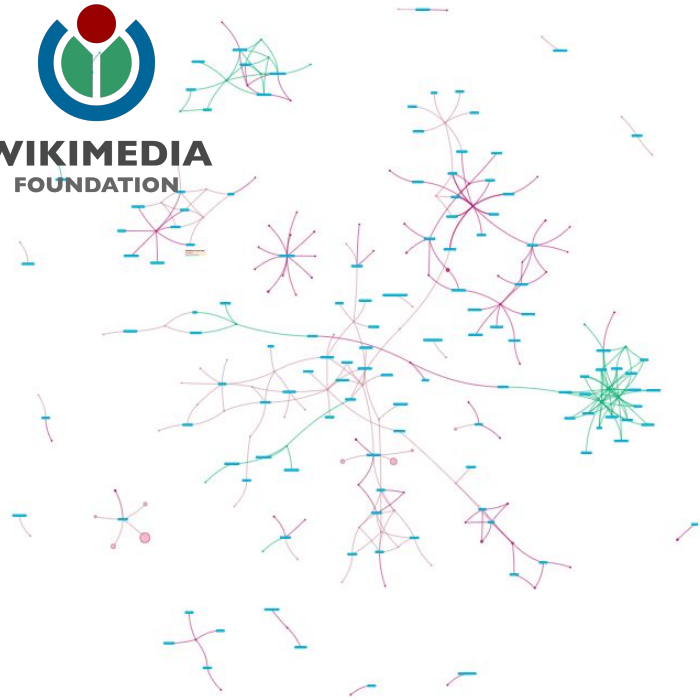
FOUNDATION



Knowledge silos, continent communities



WIKIMEDIA
FOUNDATION

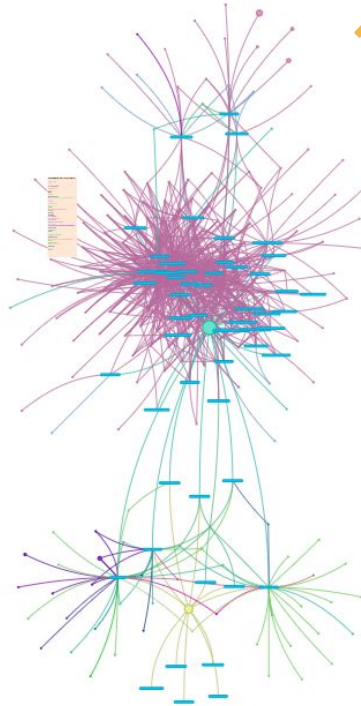


LEGEND OF COLORS:

Independent
Wikimedia Foundation
Wikimedia Deutschland
Speed & Function



Open Infrastructure
FOUNDATION



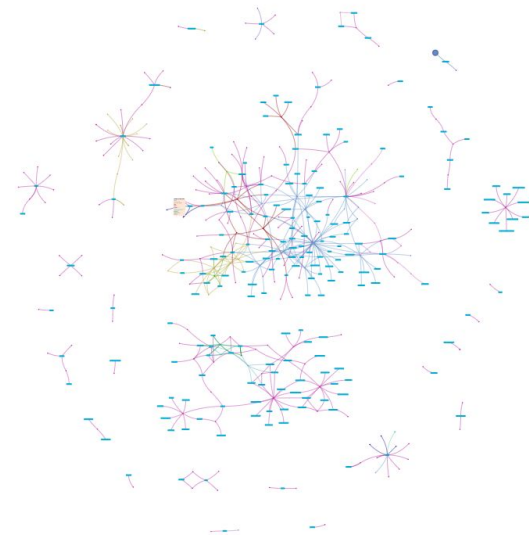
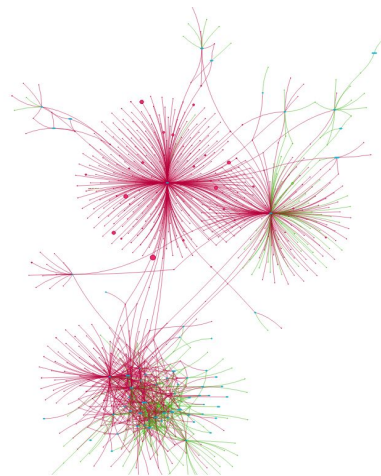
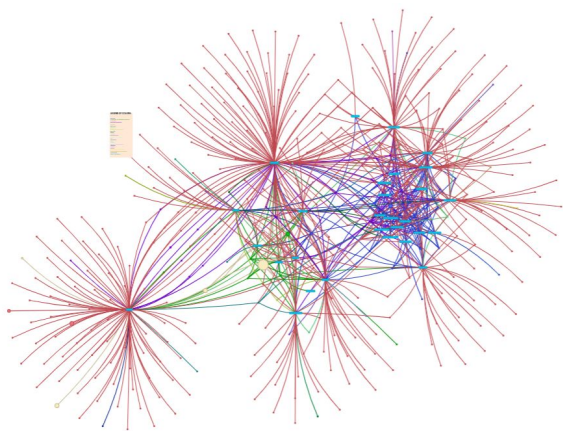
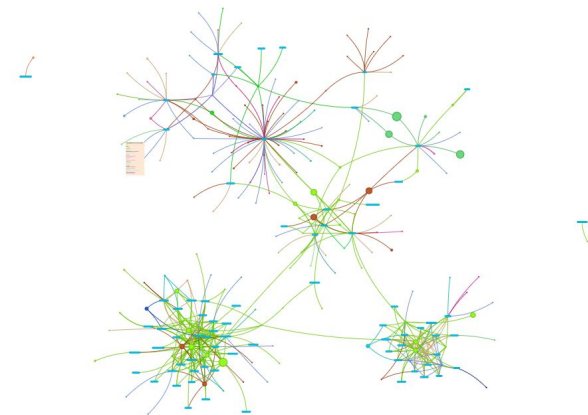
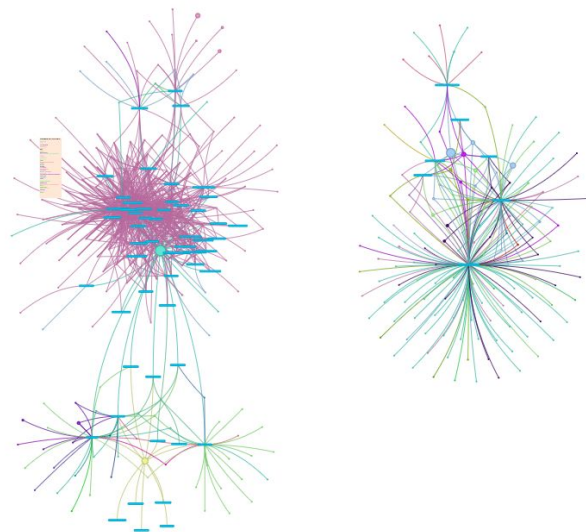
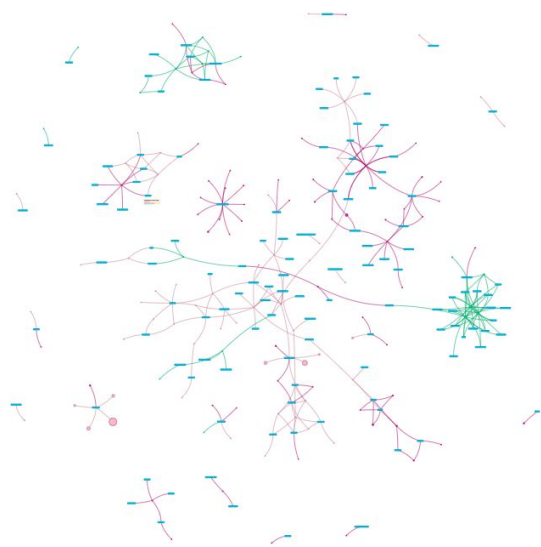
Organizational Diversity, areas of expertise



LEGEND OF COLORS:

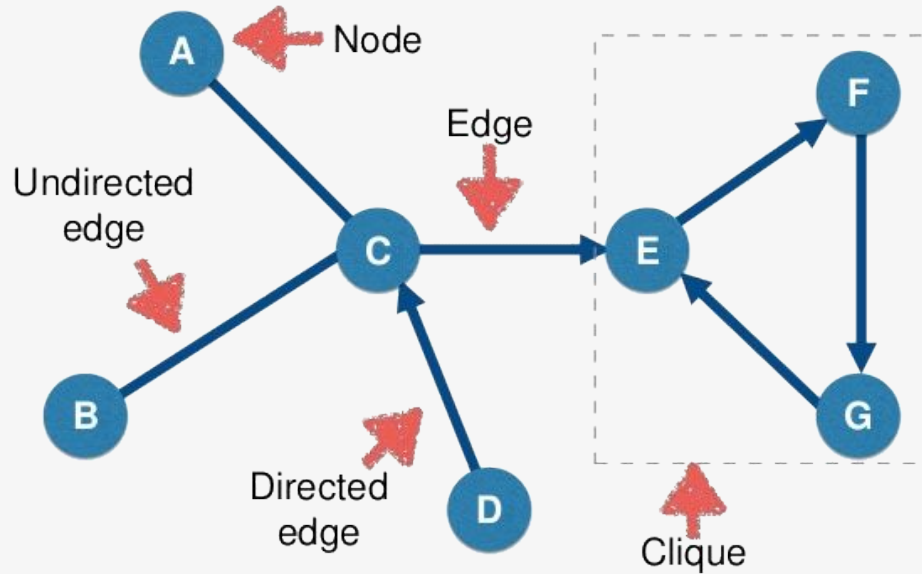
Unknown
Intel
Acme Group
Ant Financial
Wind River
Red Hat
IBM
Apple
BMW Group
Open Infrastructure Foundation
ARM
Alibaba
Databricks
EasyStack
Sony
OpenStack Foundation
Inspur
Nvidia
Garmin
T-Systems
OSISM GmbH
VexxHost
Debian GNU/Linux
Nubifocus
The Apache Software Foundation
Ibboncoin
SendGrid
AMD

ZTE Corporation
Spotify AB
BBC Research & Development
Bosch
Volvo Cars
Branch Metrics
99cloud
Microsoft
Nokia

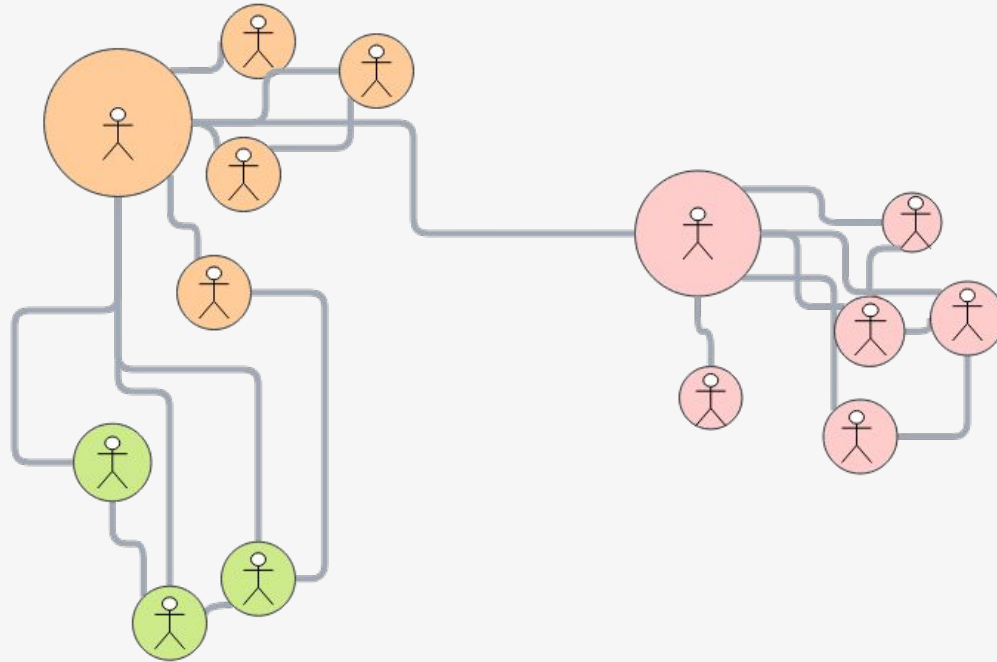


What is a graph?

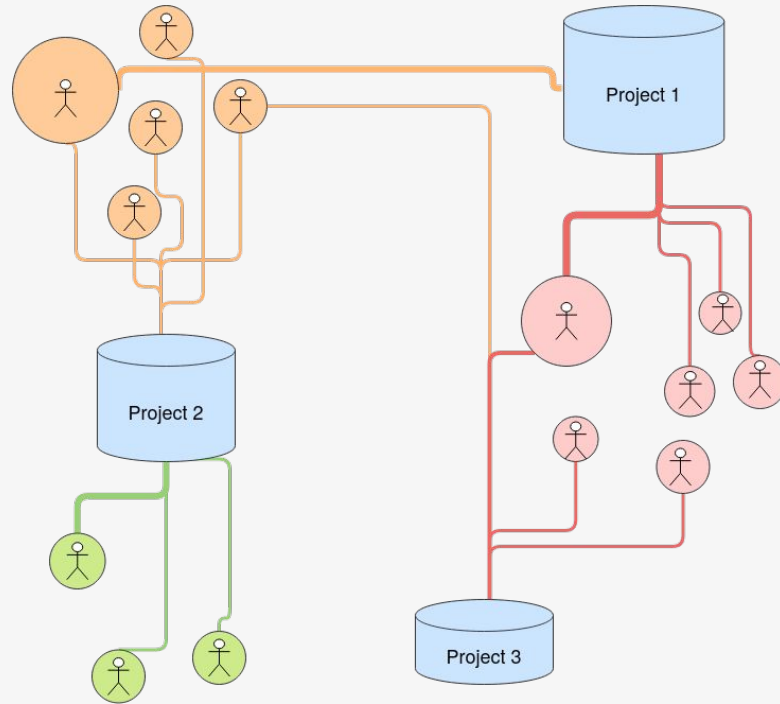
Representation of a network as a set of connected elements



Creating collaboration networks (I)



Creating collaboration networks (II)



How to measure collaboration from a network?

Which properties of the network can help us to measure collaboration?

Which metrics should we consider?



Applying Graph theory: Network properties

Adjacency

Two **nodes** are **adjacent** if there is an edge between them.
Two **edges** are **adjacent** if they share one of their ends.

Degree

The **degree** of a **node** is the number of connections that it has to other nodes in the network.

Connectivity

A node is reachable from another node if there is a path between them. A graph is **connected** if there is a path for every pair of nodes in the graph.

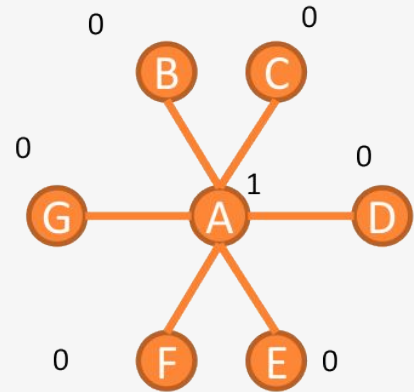


Applying Graph theory: Centrality metrics

Betweenness centrality

A way of detecting the **amount of influence** a node has over the flow of information in a graph.

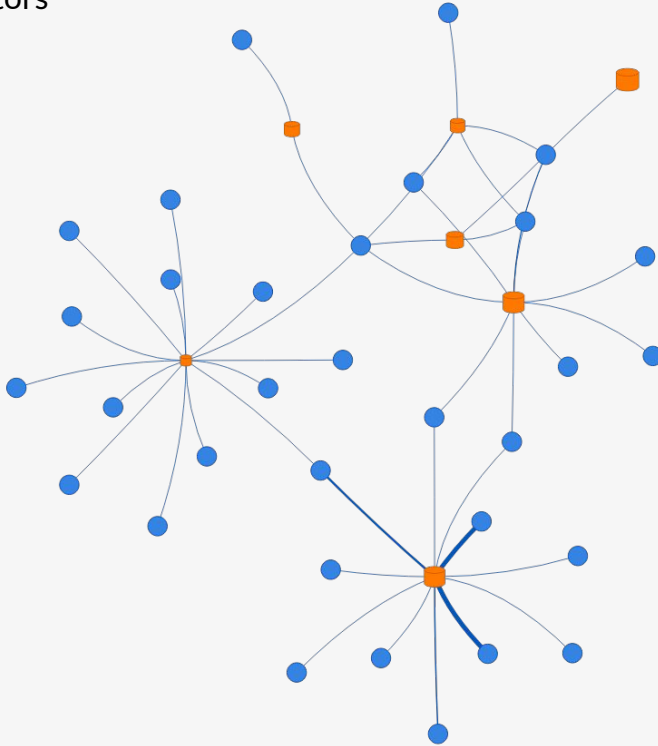
It is often used to **find nodes that serve as a bridge** from one part of a graph to another.



Analyzing a real network (I)

● Contributors

■ Projects



Adjacency

Contributor nodes sharing edges to *Project* nodes indicate collaboration among these people.

Degree

The amount of connections from a *Contributor* node indicates they collaborate in many projects.

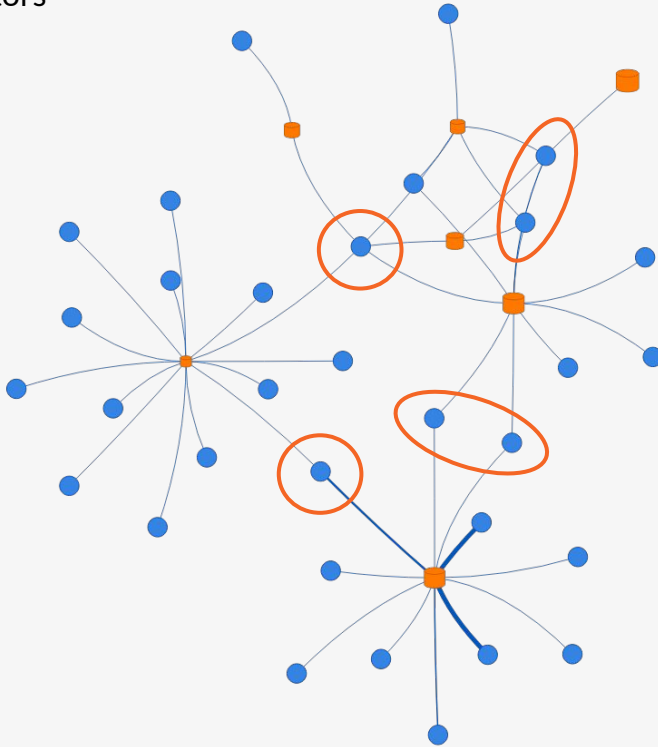
Connectivity

A **highly-connected network** indicates a more collaborative community.

Analyzing a real network (II)

● Contributors

■ Projects



Betweenness Centrality



Finding the contributors connected to a greater number of projects help us find the people acting as bridges in the community.

Metrics

Implementation
agnostic community
development metrics

4. Definition and toolkit

4.1 Definition

- Influence
 - Influence uses a centrality algorithm to identify the influence of the specific node (repo or developer) in a graph created by basic CHAOSS metrics as relationships or node properties. Influence integrates CHAOSS metrics not by linear algorithm but by an open source graph.
- Centrality algorithm
 - A PageRank-like algorithm can be used as the centrality algorithm. Assume the whole open source world network constructs with project and developer nodes. The algorithm has following basic assumptions:
 - Developers with more influence tend to contribute to projects with more influence.
 - Projects with more influence tend to attract developers with more influence.
 - Projects with more influence tend to be used by more projects.

4.2 Graph construction

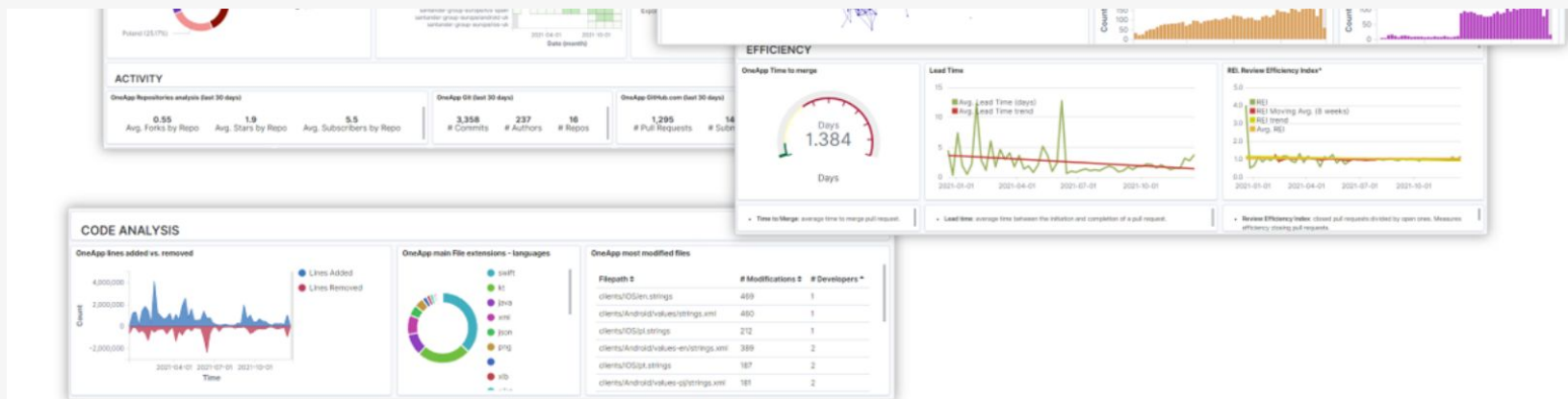
In the influence metrics model. Rather than using the count numbers of each metric, it splits the relationships into more detailed granularity.

Work in Progress @ Metrics Models Working Group

Join #wg-metrics-models @ CHAOSS Slack



Collaboration in Action: Santander InnerSource Metrics



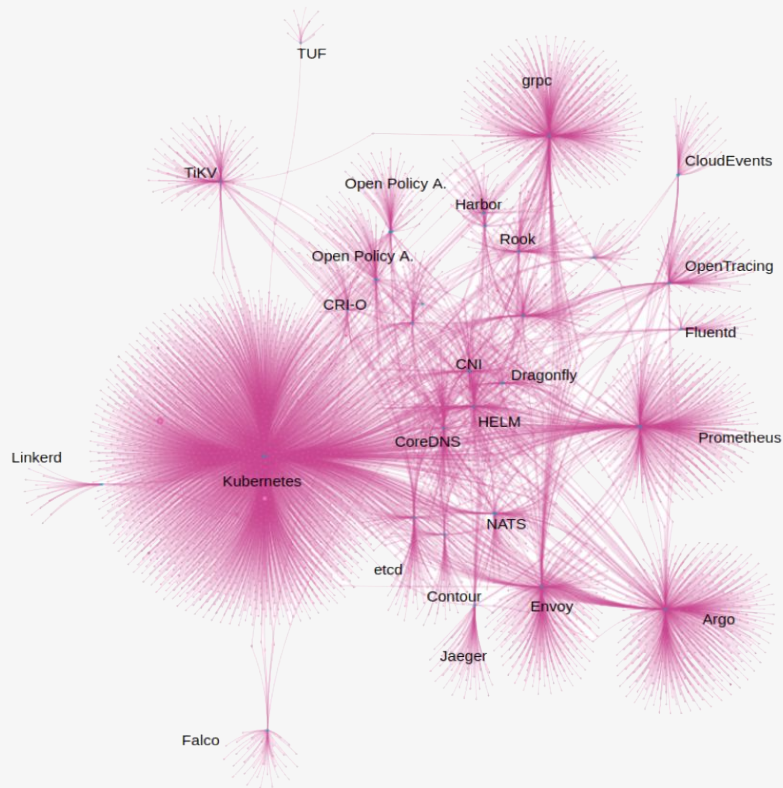
powered by  Bitergia

Take a look at our Metrics Analysis Strategy
METRICS THAT MATTER. CLICK HERE



<https://innersourceportal.santander.com>

Collaboration in Action: Red Hat and CNCF



Discover developer interrelations, onboard newcomers faster, and align project expectations and releases.

“Scaling management skills by 10x thanks to data insights”

From Art to Science: The Evolution of Community Development. Diane Mueller and Daniel Izquierdo.

IEEE Software Volume: 36, Issue: 6, Nov.-Dec. 2019

<https://www.cncf.io/blog/2020/08/04/a-guide-to-untangling-the-cncf-cross-community-relationships/>

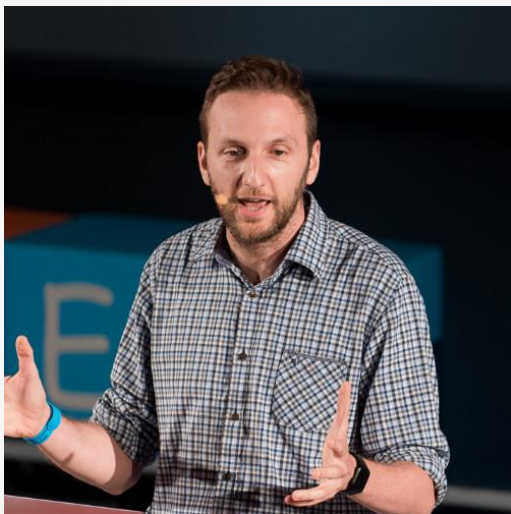
Collaboration in Action: Mozilla Rebel Alliance



“[...] holistic view of our contributor ecosystem’s network structure, health and impact [...]”

“[...] we’re able to visually describe these distinct contributor communities as well as how they are interconnected [...]”

<https://report.mozilla.community/>



CEO @ Bitergia
Governing Board @ CHAOSS
VP @ InnerSource Commons Foundation

dizquierdo@bitergia.com
@dizquierdo





Collaboration as Health Indicator

CHAOScon EU 2022

Software
Development
Analytics

Daniel Izquierdo Cortázar
(in collaboration with Miguel
Ángel Fernández)
