On the Understanding of Complex Systems: the Linux Kernel as Use Case

Daniel Izquierdo, dizquierdo@bitergia.com

openEuler Summit, 2023

CEO at Bitergia

Board Member at CHAOSS

President at the InnerSource Commons Foundation

Proud father of an 18th months little boy :)

Daniel Izquierdo Cortázar

dizquierdo@bitergia.com

Definition

What is a complex system?

A complex system is a system composed of **many components** which may **interact** with each other. [...]

[...] systems whose behaviour is intrinsically difficult to model due to the **dependencies**, **competitions**, **relationships**, or other types of **interactions** between their parts. [...]

In many cases, it is useful to **represent** such a system as a **network** where the nodes represent the components and links to their interactions

https://en.wikipedia.org/wiki/Complex_system

Motivation

Open Source Software Ecosystems as complex systems

Open source projects are a compendium of people with different interests working together in a common mission.

They create and evolve different assets and by bringing all of those different interests, they are able to build a set of assets (source code, documentation, etc) useful for others.

Motivation

Why understanding complex systems matter

The more we are able to understand and model complex systems, in a better position we'll be to potentially predict catastrophes and decaying ecosystems.

And this has a direct application in the open source sustainability and health analysis.

Motivation

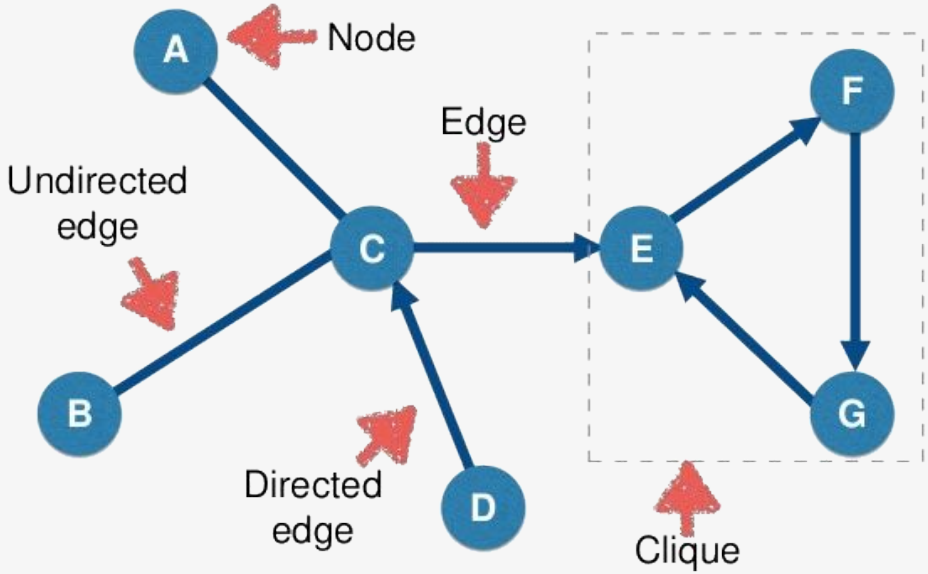
The Linux kernel is a massive collaborative effort across regions, industries, and organizations with a common vision and mission.

They work without consensus, but with a set of pre-defined rules.

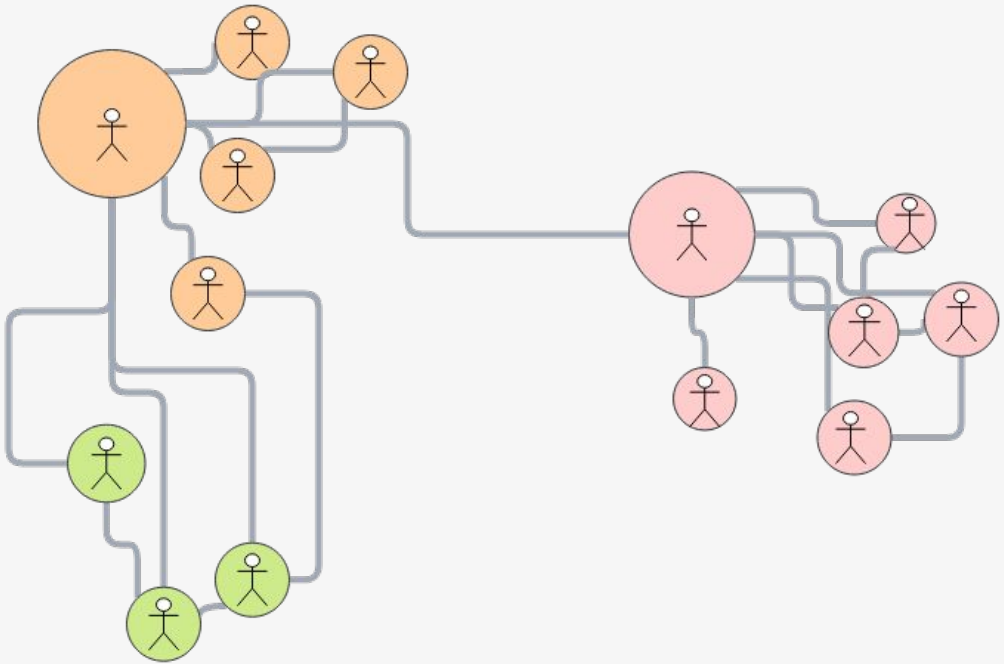
Crash course on networks

What is a graph?

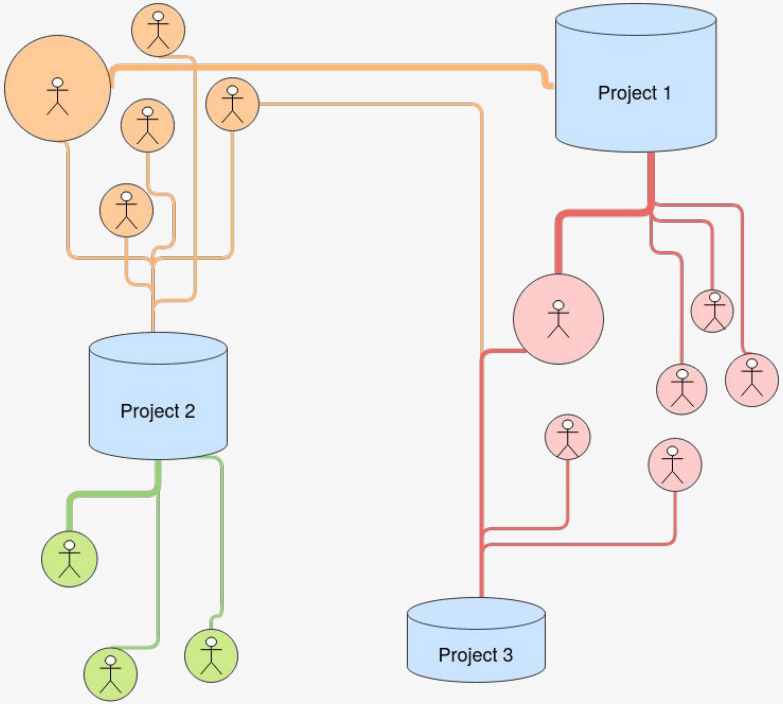
Representationof a network as a set of connected elements



Creating collaboration networks (I)



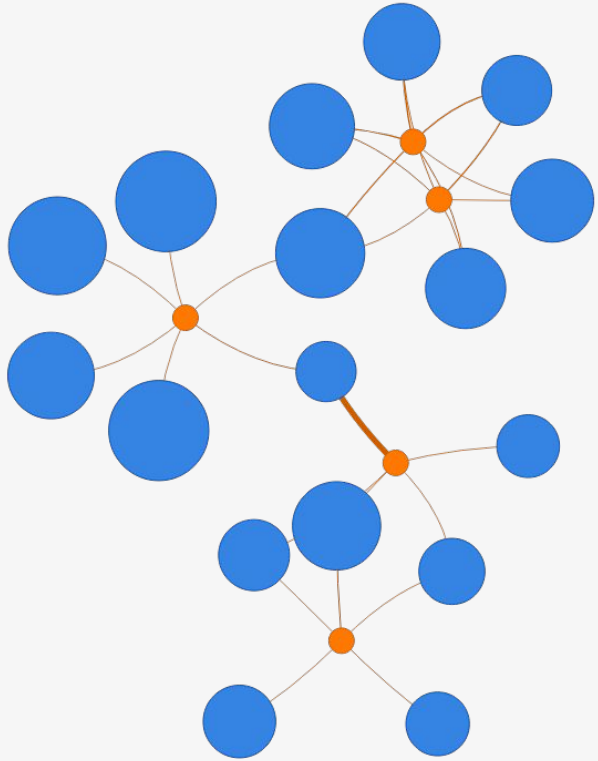
Creating collaboration networks (II)



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

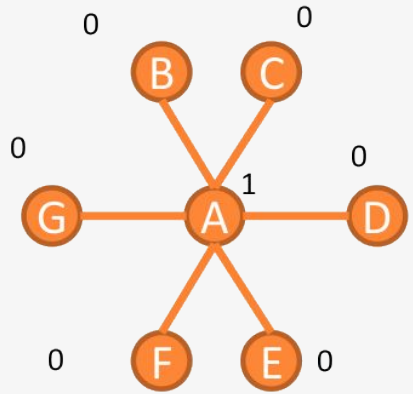
A node is reachable from another node if there is a path between them. A graph is **connected** if there is a path for every pair of nodes in the graph.

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.

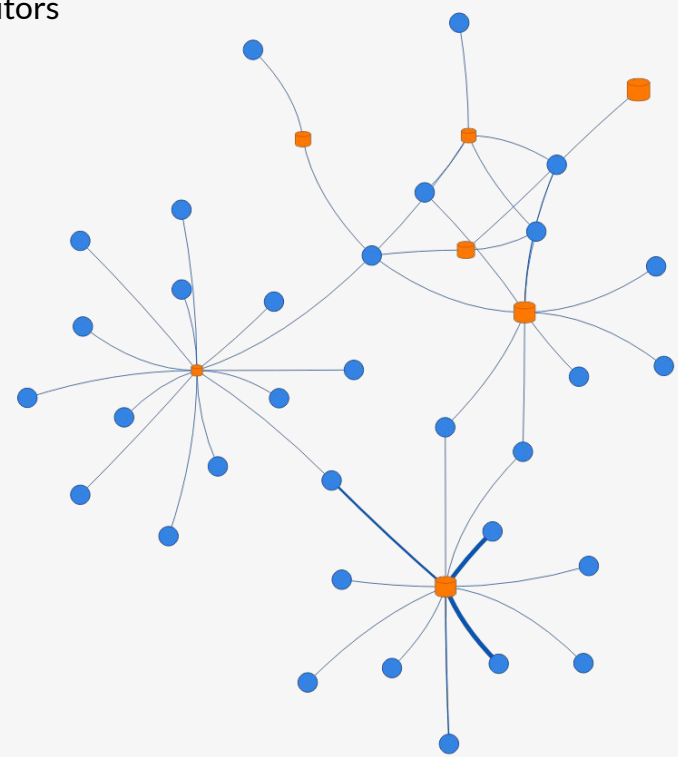
It is often used to **find nodes that serve as a bridge** from one part of a graph to another.



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.

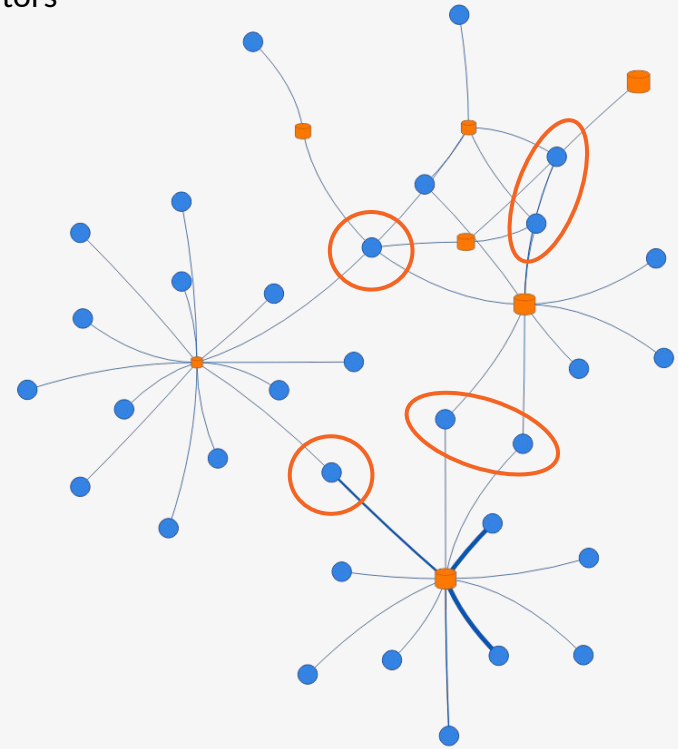
Connectivity

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

Analyzing a Real Network (II)

Contributors

Projects

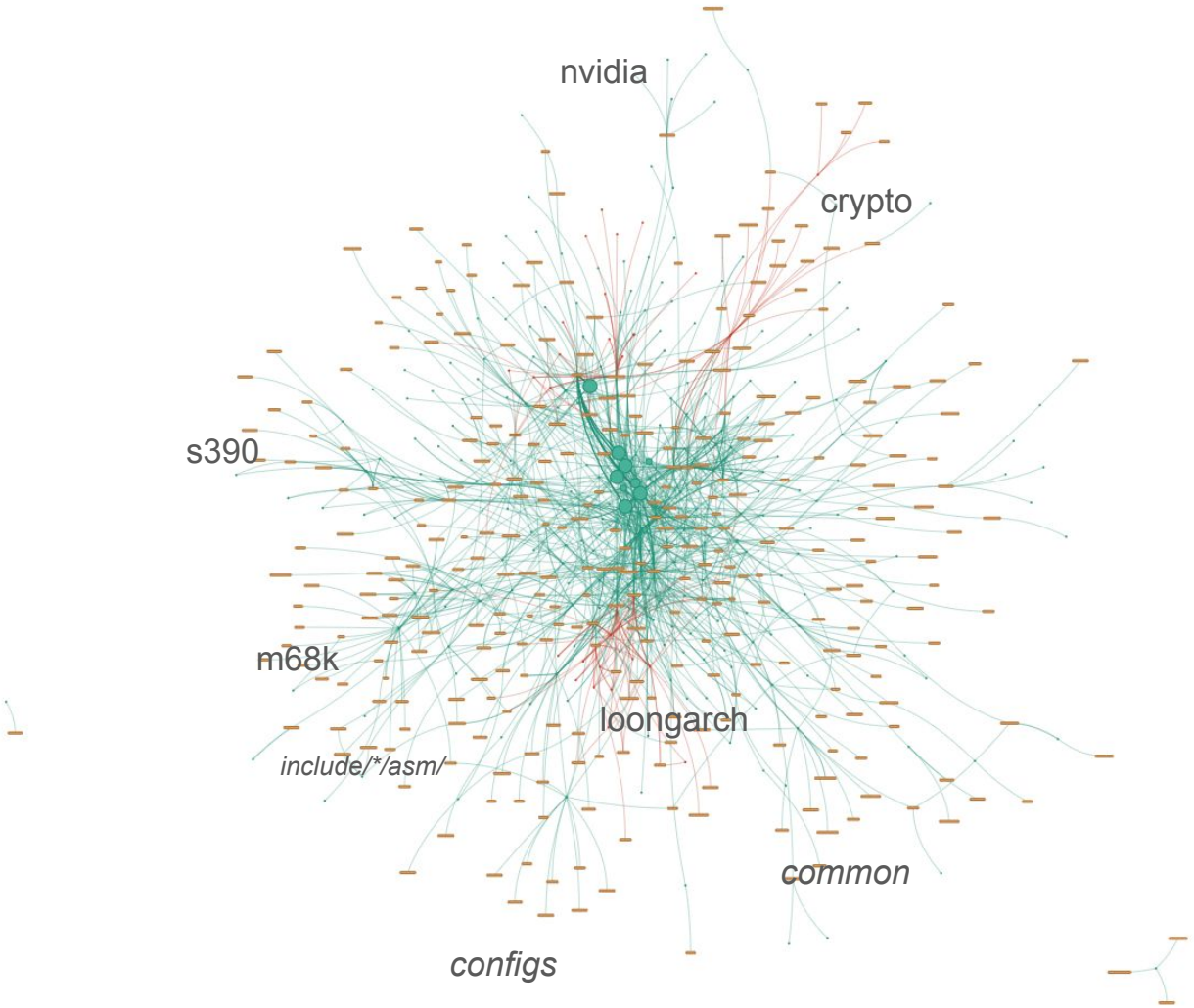


Betweenness Centrality

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

Linux Kernel

- Developers in China
- Developers in Europe

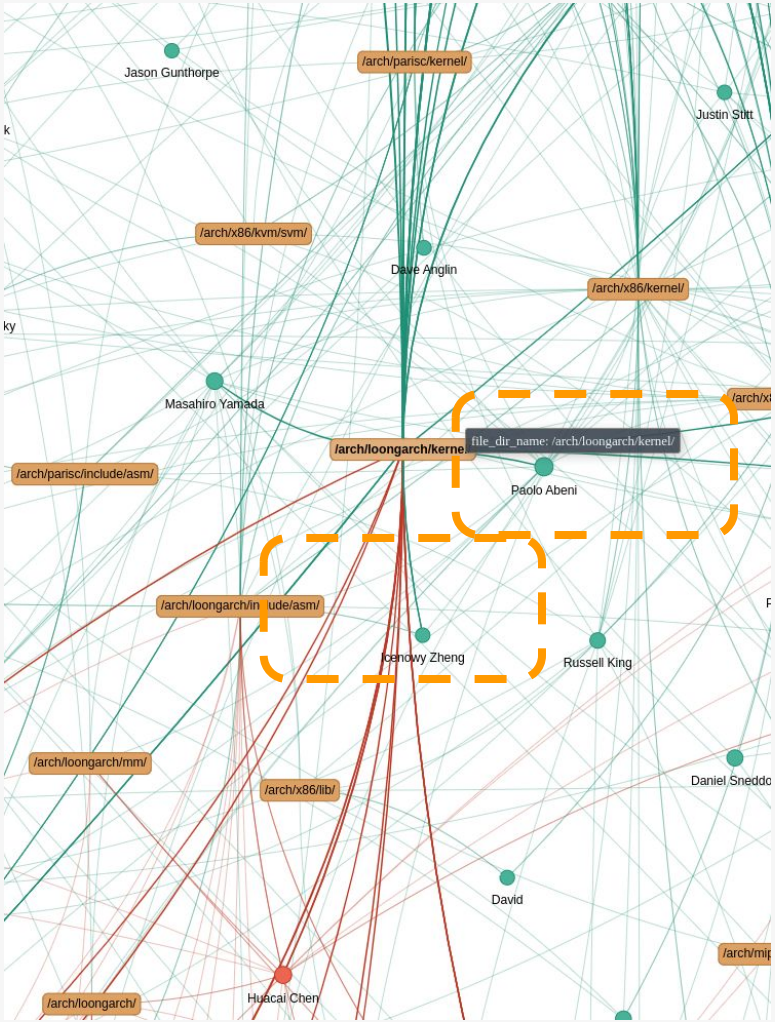


Linux Kernel

Where collaboration happens

/arch/loongarch/include/asm/

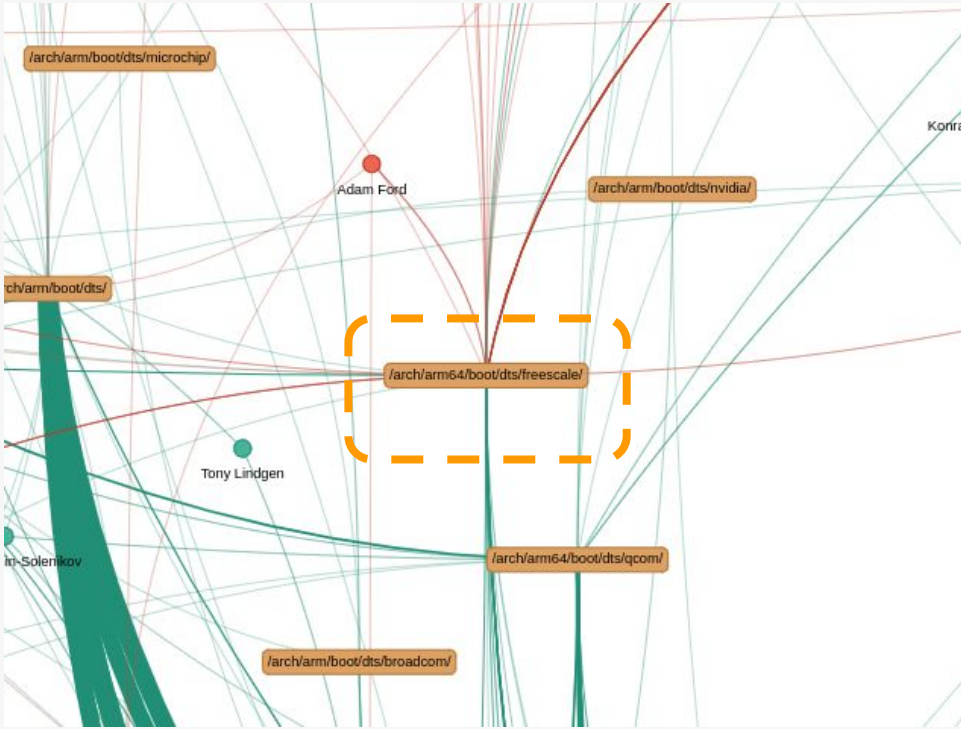
/arch/loongarch/kernel/



Linux Kernel

Where collaboration happens

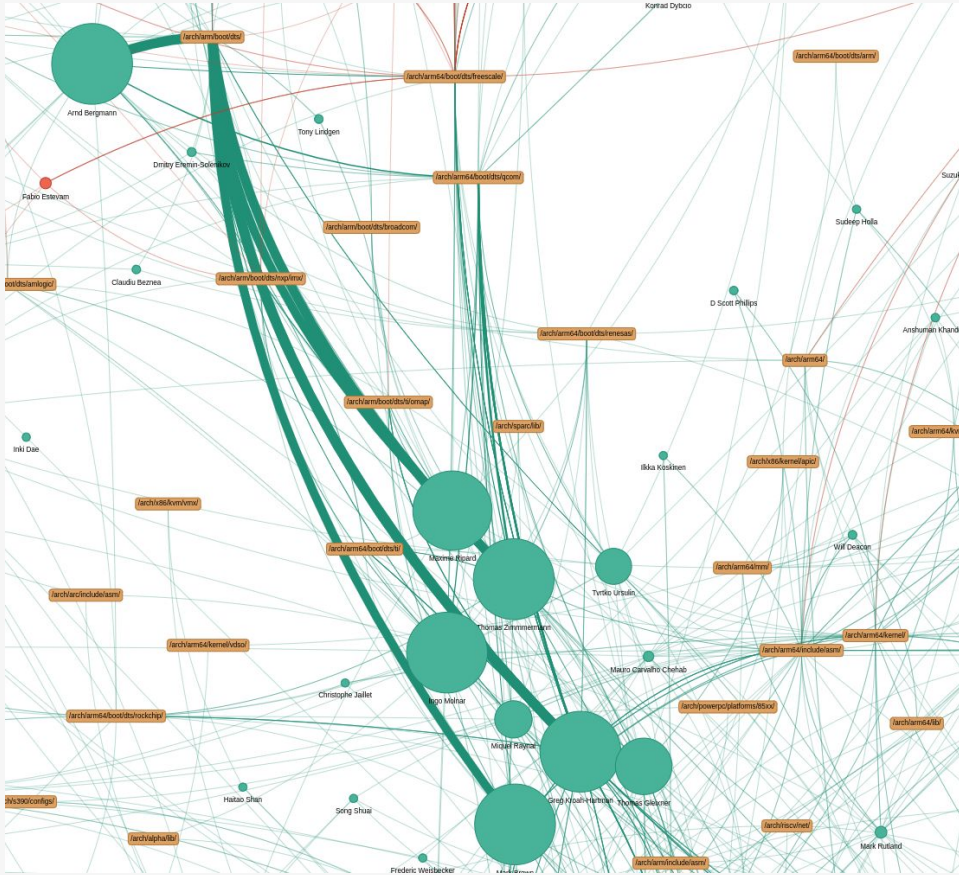
/arch/arm64/boot/dts/freescale



Linux Kernel

Where high specialization takes place

/arch/arm/boot/dts



Why is important to understand complex systems?

Digital infrastructure nowadays run on complex systems

The sustainability of the several pieces of the software supply chain is critical for the sustainability of projects as openEuler

Highly collaborative environments are more resilient

Next Steps

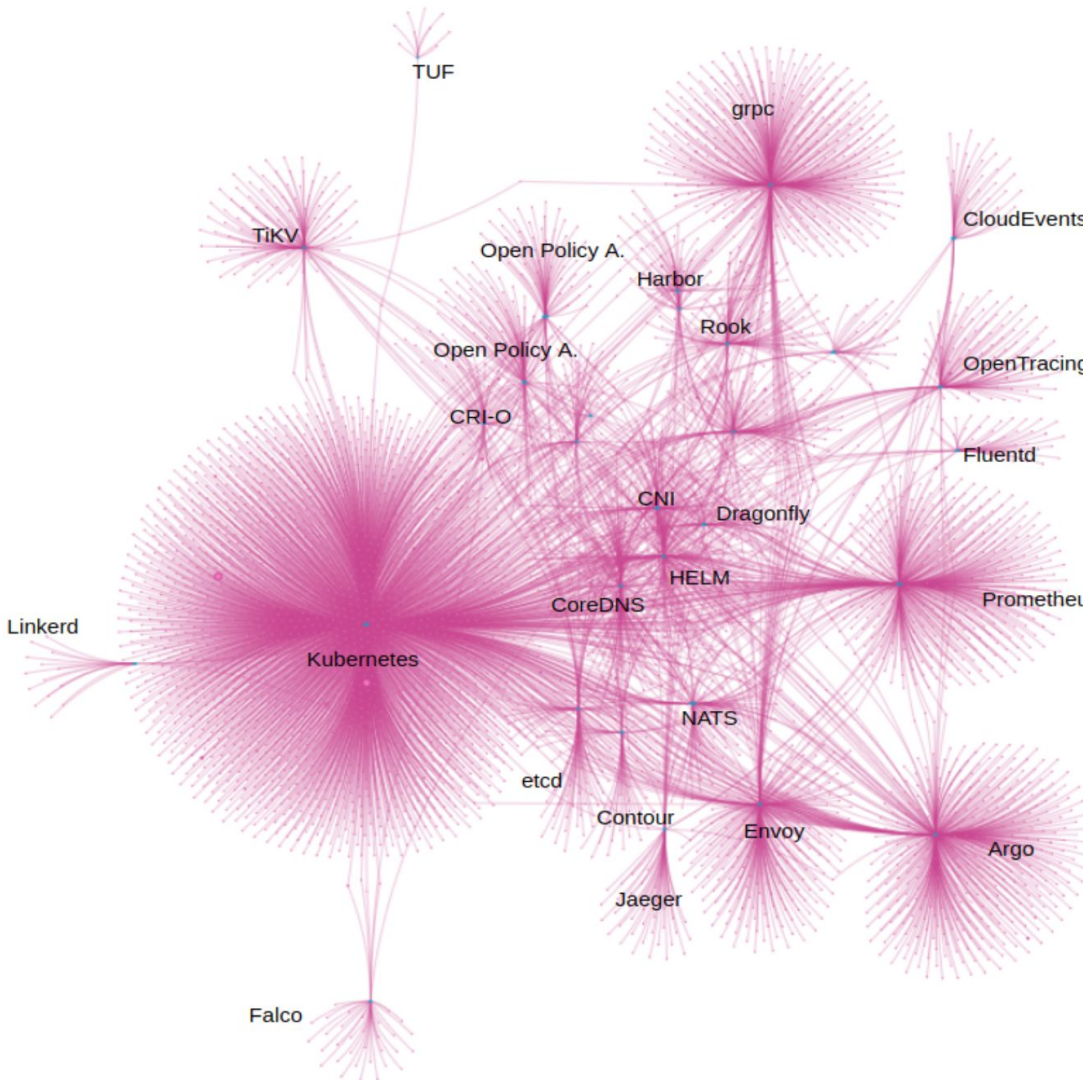
openEuler and CHAOSS are starting a collaboration on this

You can join us at the Metrics Models meeting

https://chaoss.community

Extra Ball I

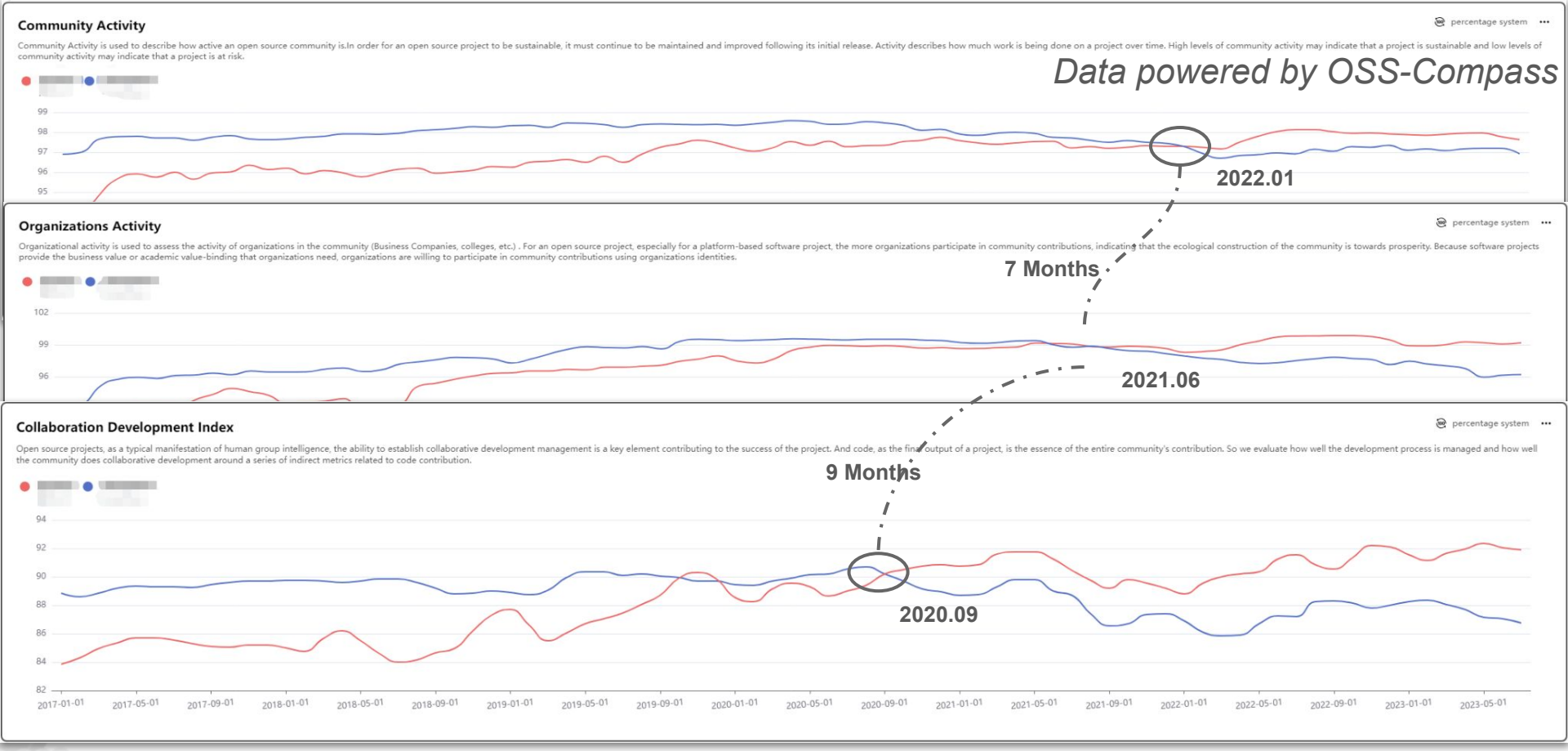
Having a look at the CNCF ecosystem



Data powered by Bitergia Analytics

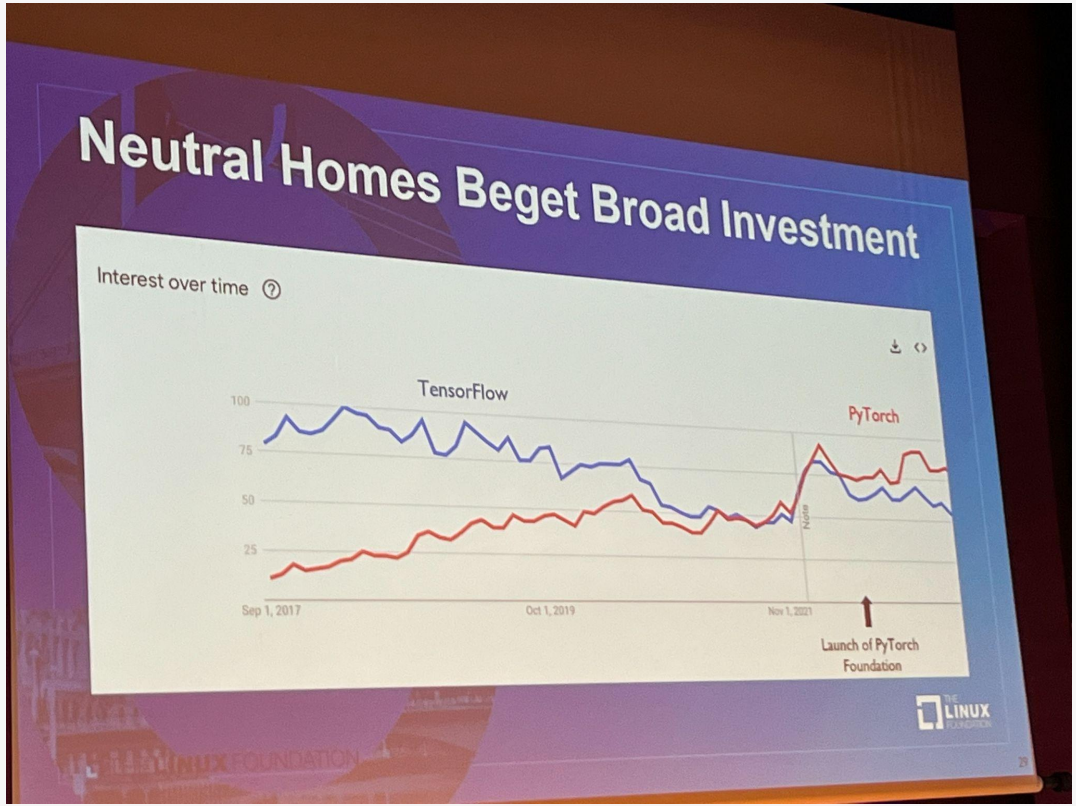
Extra Ball II

COMPASS



Extra Ball II

Jim Zemlin's Keynote in Bilbao on Pytorch and Tensorflow



On the Understanding of Complex Systems: the Linux Kernel as Use Case

Daniel Izquierdo, dizquierdo@bitergia.com

openEuler Summit, 2023