

Evolution of New Open Source Communities

Oscar Castañeda

Introduction

New open source communities in the Apache Incubator are groups of individuals contributing to the evolution of a software codebase. Some contributors volunteer time and efforts while others are paid to develop software. They seldom meet but often interact and collaborate through shared infrastructure on the Internet, like email and revision control systems. Still they organize in communities to collectively develop enterprise grade software and eventually become full-fledged open source communities after graduation from the Apache Incubator. The question behind this project is: 'how?'

Burning questions...

New open source communities continue to become part of The Apache Software Foundation (ASF). The entry path for these communities is the Apache Incubator, a top-level project created by the ASF to develop new projects and communities to comply with legal standards and adhere to the guiding principles of 'The Apache Way' [1]. The enormous success enjoyed by the ASF, its open source communities, and the open source software they produce, leads to questions such as:

1. How do new open source communities evolve in the Apache Incubator?
2. How are the processes of 'The Apache Way' related to the organization of new open source communities?
3. How does voting influence community structure over time?

Answering these questions requires an understanding of how open source communities are organized and how they sustain themselves. It also requires an understanding how open source community development works in practice.

How are open source communities organized and how do they sustain themselves?

Recent organizational research into open source communities by Van Wendel de Joode (2005), showed that open source communities are self-organizing, which limits their malleability and critically questions whether it is possible to purposefully design and create new open source communities. To come to this conclusion, Van Wendel de Joode (2005) adopted a framework from research on community- managed common pool resources.

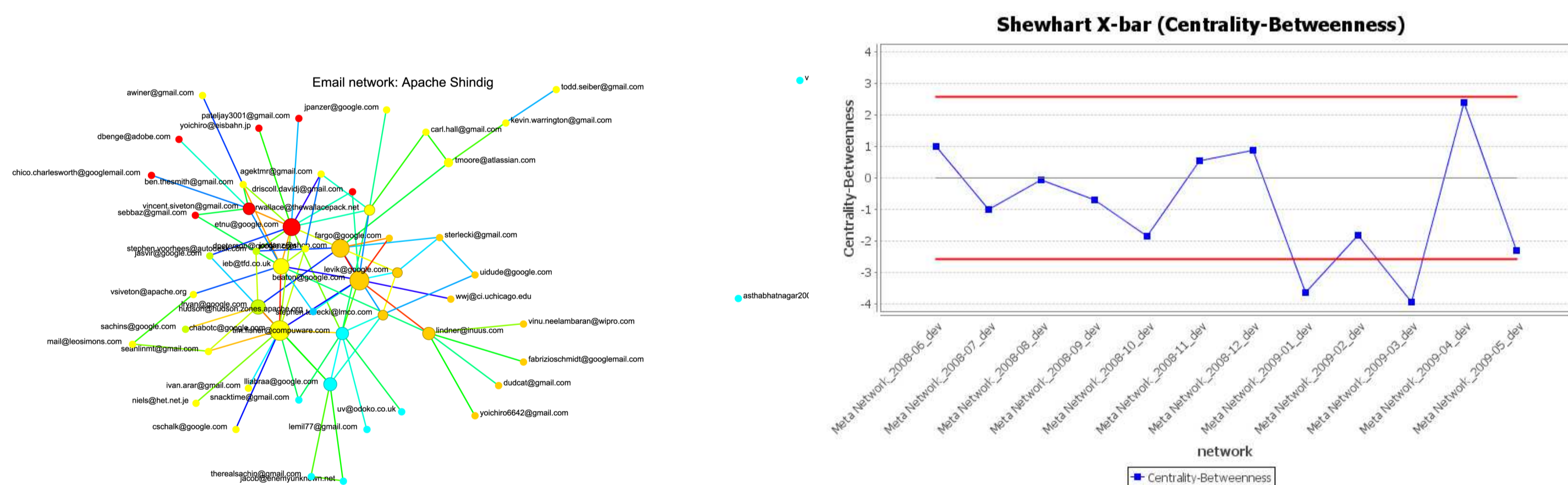
The framework consists of eight design principles that reconcile the two main strands of state of the art research into the organization of open source communities: *self-organization* and *institutions*. It has been used to study other communities managing common pool resources; from farming communities sharing irrigation systems for more than 1,000 years, to global commons like telecommunications providers that transcend national boundaries, and virtual commons like communities found on the Internet. Importantly for this research, the framework shows 'where to look' to study a key aspect in the evolution of new open source communities.

How does voting influence community structure?

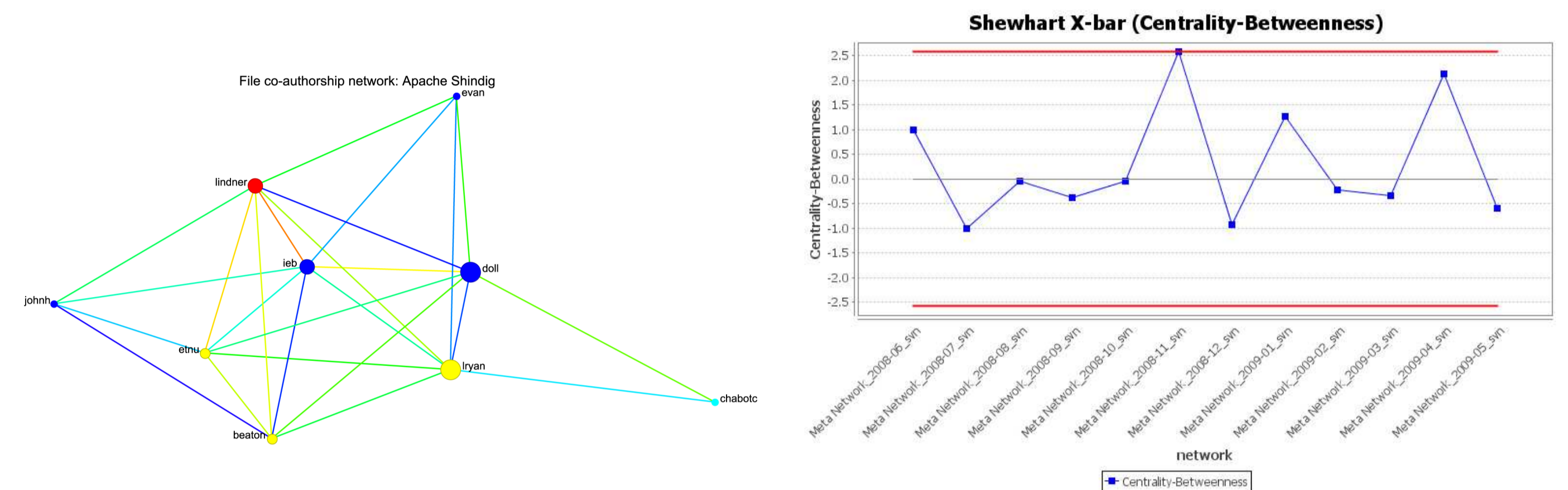
To help answer this question, social network analysis will be performed on new open source communities using the *ORA tool developed at Carnegie Mellon University (CMU). Social networks over time are obtained from the publicly available email archives and source code repository (SVN) logs of incubator projects of The Apache Software Foundation. Statistical change detection techniques are then applied on different social network measures over time.

Observations in a new open source community currently under incubation

Social network analysis of email message interaction as a proxy for voting...

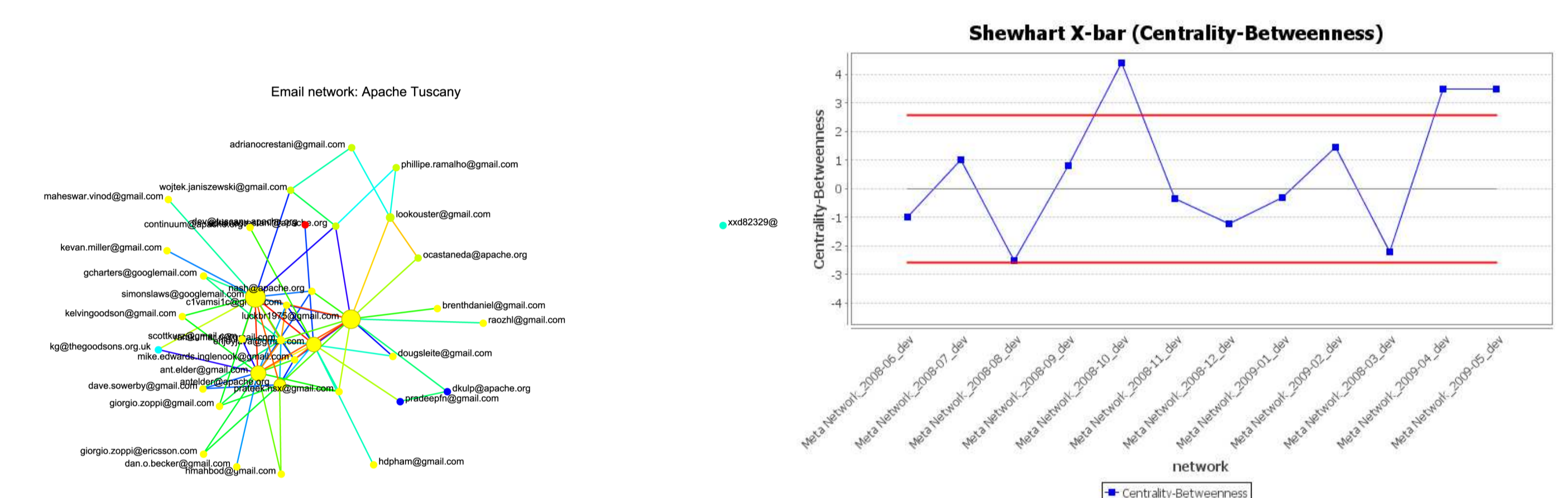


...and the corresponding file co-authorship social networks over time.

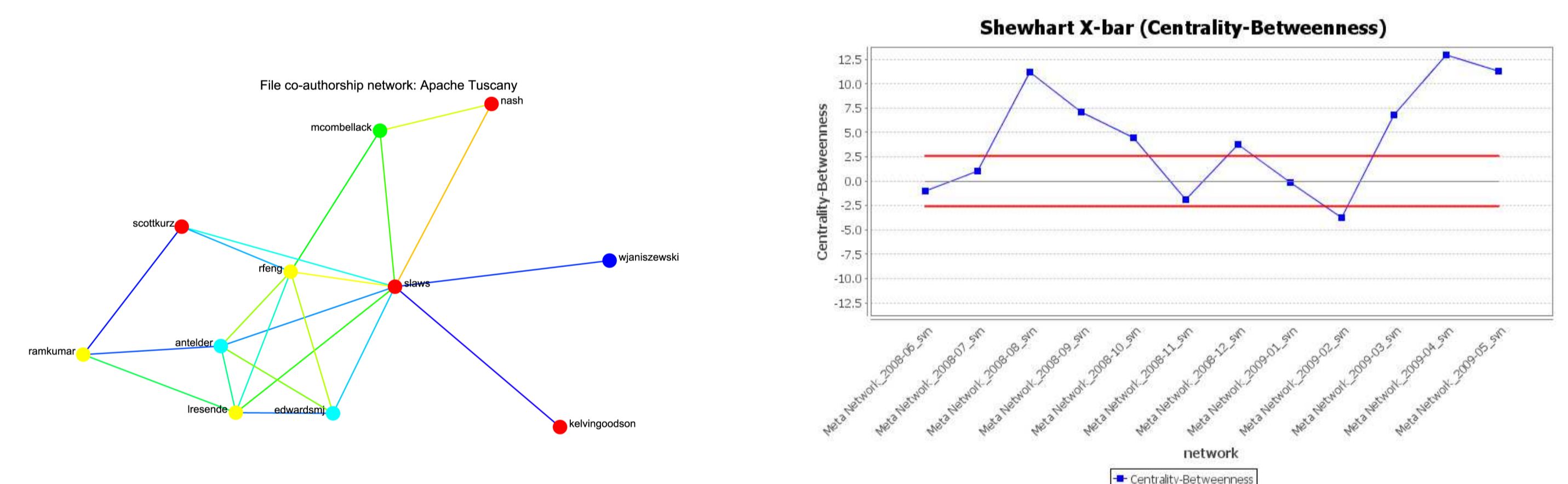


Observations in a new open source community after incubation

Social network analysis of email message interaction as a proxy for voting...



...and the corresponding file co-authorship social networks over time.



Next steps

The next steps for this project are to analyze social network measures over time, at different levels of analysis and for different (new) open source communities. These communities are either currently under incubation or have recently graduated from the Apache Incubator. One of the major goals of this project is to identify and reason on the basis of patterns that may explain the influence of institutions like voting on community structure. Such patterns are believed to be important in the design and evolution of new open source communities.

References

[1] The Apache Software Foundation - The Apache Way - <http://www.apache.org/foundation/how-it-works.html>

R. Van Wendel de Joode. Understanding Open Source Communities, an organizational perspective. PhD thesis, Delft University of Technology, 2005.

Acknowledgements ¹

Supported by the Programme Alβan, the European Union Programme of High Level Scholarships for Latin America, scholarship No. E07M401381GT.

Prof. dr. Michel van Eeten and Dr. Victor Scholten, MSc. thesis supervisors at Delft University of Technology, Faculty of Technology, Policy and Management.

The Apache Software Foundation and Stone Circle Productions.

Google Summer of Code Program, Google Inc.

Center for Computational Analysis of Social and Organizational Systems (CASOS), Carnegie Mellon University.

¹Developed on L^AT_EX