Universiteit Leiden

Master Computer Science

Measuring architectural change and maintainability in networks of evolving software

Name:	Mona Matar
Student ID:	s2030309
Date:	01/09/2021
Specialisation:	Data Science
1st supervisor:	Frank Takes
2nd supervisor:	Michel Chaudron

Master's Thesis in Computer Science

Leiden Institute of Advanced Computer Science (LIACS)
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands

Abstract

The architecture and quality of software systems continuously change as software evolves. Traditional software metrics for measuring maintainability have been introduced to measure this evolution. Social network analysis techniques have been applied to software systems, but have not yet been investigated for measuring maintainability of evolving software systems. In this thesis we apply social network analysis to evolving software networks in order to understand architectural changes and identify network metrics that measure such changes. Having a broader understanding of changes to maintainability enables the improvement of all software quality attributes. In this exploratory study, we use software networks and network metrics to asses maintainability. We conducted an empirical study in which we modeled all versions of six open-source object-oriented programs as networks of different levels of granularity. We selected software metrics that quantify maintainability factors based on literature, and compared changes to these metrics against changes to the network metrics. We reached the conclusion that, for all levels of granularity, the best network metrics to measure changes to the maintainability of evolving software systems are communities, weakly connected components, motifs and density, and the best network granularity is the collaboration network.

Acknowledgements

First and foremost I would like to extend my gratitude to Dr. Frank Takes for his patience and support while working on my thesis, and Alexandra Blank for helping me get back on track during the COVID-19 lockdown. A special thanks to my family for their continuous support and encouragement during my studies. I would like to single out my husband for his unlimited encouragement and help during my study and specially during writing the thesis. Last but not least I would like to extend a heartfelt thanks to all the professors who taught me in LIACS , and my fellow students who made sure that I maintain my self confidence in pursuing my masters as a mature student.

Contents

List of Figures

List of Tables

1 Introduction

Software systems are constantly evolving to accommodate the ever-changing requirements (new technology, new customer and market requirements, performance improvement, extension of functionalities and error correction). Software maintainability is concerned with how efficiently a system can be modified to implement the required changes. Maintainability is about the ease and ability to implement required changes, while preserving the systems functionality and stability. As described in Lehman's laws of evolution, an evolving software program is continuously changed, its complexity increases, sometimes reflecting deteriorating structure, unless effort is applied to maintain or reduce it [1].

Concepts of complex network analysis have been an effective tool in understanding and analyzing the structure of complex systems such as the World Wide Web [2, 3], Internet [4], protein folding [5], and collaborations in science [6, 7]. A considerable number of researchers applied the network analysis concept within the software engineering discipline. Software systems are man made complex systems that can be represented as networks, [8, 9, 10, 11] known as software networks [12], where software components (classes/interfaces, methods/attributes, packages etc.) are nodes and their interactions (class relationships, method calls, etc.) are the edges.

Studies have shown that *static* software networks exhibit a small-world (SW) [8, 13, 14], and scale-free (SF) topology. One of the first works to introduce network theory into software analysis [15] found out that two software systems (JDK 1.2 and Ubisoft Prorally) have small-world structure (a high degree of clustering and a small average distance known as six degrees of separation) and scale-free (SF) properties where few nodes have high degree (connections) and the majority of the nodes have very few connections. Social network analysis (SNA) was employed, to identify important critical components [16, 17], defective modules [18], and to serve as quality indicator [19, 11]. This made it possible to gain further insight on into software structure with the aid of SNA.

There have been several studies analyzing software *evolution* using networks, for example by building a static snapshot of the software network for each software version. Analyzing changes in network topology over successive software versions, researchers applied different metrics to obtain detailed information about the effect of evolution on software structure. Studies were conducted to understand the dynamics of the relationships between evolution and system complexity, stability [20, 21], functionality [22] and maintenance effort and bug severity prediction [23].

Most studies on evolving software networks concentrate on bug prediction. However, there is a lack of studies on evolving software network and structure attributes (security, maintainability, portability, efficiency, usability, reliability, compatibility and functionality). Maintainability as defined by ISO 25010[1] is a less investigated area in the evolving software network field. Software systems and applications have become an integral part of daily life. As with any product software structure, a quality standard needs to be met, and maintenance tasks need to be performed. Maintenance cost is the highest percentage of the total project cost, so it is imperative to monitor maintainability after every evolution cycle.

This research is based on building successive snapshots of software networks, to help investi-

[1]ISO/IEC 25010 https://iso25000.com/index.php/en/iso-25000-standards/iso-25010?start=6

gate the structural and quality attribute changes (with focus on the maintainability attribute). We will model software systems using different levels of granularity (inheritance network, collaboration network -inheritance & dependency-, call network -method/global functions- and method to method call network). We will analyze and monitor six open source object-oriented software systems (OOSS) from the Apache system, with versions per software system ranging from 6 to 82 and live span ranging form 4 to 11 years.

The aim of this research is to monitor the changes in the software networks topology for successive versions in order to better understand software evolution and its effect on software structure. In this study we model successive versions of software systems as directed graphs using above mentioned four different levels of granularity and investigate the following questions:

1. *RQ1:* How can typical qualitative factors of maintainability from the software engineering literature be assessed using quantitative software metrics?

2. *RQ2:* What SNA measures and metrics capture software metrics that are strong indicators of changes in maintainability?

3. *RQ3:* How are the changes in SNA metrics different in the four levels of granularity for software networks, and which network is a better representation of structural change?

The rest of this thesis is organized as follows. First section 2 presents an overview of notations and definitions used. Then section 3 introduces previous and related work, followed by section 4 which describes the approach, data collection and analysis. Section 5 presents the experiments and results. Finally, a conclusion and future work are presented in section 6.

2 Preliminaries

In this section we will describe the network terminology, notations and metrics used throughout the thesis.

2.1 Networks

Networks are represented as graphs, in this thesis we will work with directed graphs. Unless mentioned otherwise, consider the directed network $G = (V, E)$ where V is a set of nodes and E is a set of ordered pairs of nodes, $n = |V|$ and $m = |E|$. We define an edge (u, v) to exist if $(u, v) \in E$ (with $u, v \in V$), the neighborhood of node v by $N(u) = \{v \in V : (u, v) \in E\}$. A path between two nodes u and v is a sequence of distinct edges from node u to node v. Figure 1 gives an example.

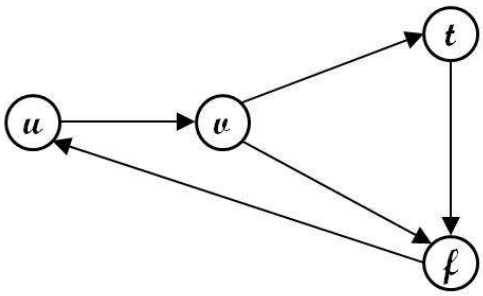


Figure 1: A directed network example

$V = \{u, v, t, f\}$,
$E = \{(u, v), (v, t), (t, f), (v, f), (f, u)\}$
$N(v) = \{t, f\}, m = 5, n = 4$
path from u to $t = ((u, v), (v, t))$

The following is a list of networks measures and metrics relevant to the remainder of this thesis:

- *Degree:* The number of edges regardless of direction connected to a node; k
- *Out-degree:* The number of outgoing edges from a node; k_{out}.
- *In-degree:* The number of incoming edges to a node; k_{in}.
- *Average degree:* In a directed network the sum of out-degree/in-degree divided by n. In a directed network the sum of the out-degrees is equal to the sum of the in-degrees.
- *Degree distribution:* In an undirected network it is the probability $P(k)$ that a randomly chosen node in the network has degree k. Directed networks have two degree distributions out-degree $P(k_{out})$ and in-degree $P(k_{in})$.
- *Scale-free networks:* Networks with *power-law* degree distribution. For undirected networks $P(k) \sim k^{-\gamma}$, where γ is a constant parameter known as the exponent, and $2 < \gamma < 3$.
- *Density:* The ratio between total number of edges in the network and the all possible edges.

- *Shortest path:* The shortest path between nodes u and v is the path with the least number of edges, its length is also referred to as the *distance* between u and v, denoted by $d(u, v)$.
- *Diameter:* The length of the longest shortest path.
- *Average shortest path length:* The average number of edges over all shortest paths for all pairs of nodes.
- *Average clustering coefficient:* The average of the clustering coefficient values of all nodes in the network, where the clustering coefficient of node u is the ratio of edges connecting the nodes in the neighborhood of node u and the number edges if the neighborhood was fully connected.
- *Weakly connected component (WCC):* A maximal subgraph where all nodes are connected to each other by some path, ignoring the direction of edges.
- *Strongly connected component (SCC):* A maximal subgraph such that for every pair of nodes u and v, there is a direct path from u to v and a path from v to u.
- *Community structure:* A community is a subset of nodes within a graph for which the connections between its members are denser than with the remainder of the network.
- *Modularity:* The extent to which a network can be divided as multiple communities. And thus a community structure.
- *Degree assortativity:* A measure of how nodes with similar degrees tend to associate together. A network is assortative when high degree nodes are, on average, connected to other nodes with high degree and low degree nodes are, on average, connected to other nodes with low degree. Assortativity measures the similarity of connections in the graph with respect to the node degree.
- *Motifs:* Subgraphs of two to five nodes that repeat themselves at high numbers compared to random networks. For example, see figure 2. Motifs are often called the building blocks of networks. Each type of network seems to display its own set of characteristic motifs.

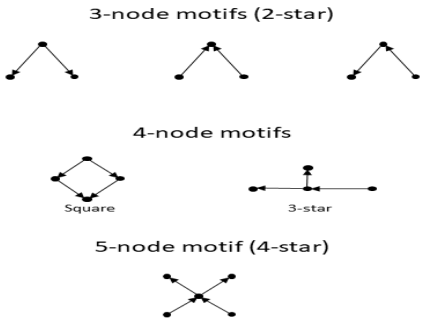


Figure 2: Example of motifs

2.2 Software metrics

Metrics give a quick overview of a project helping designers with the detection of flaws or degradation in their architecture. In this subsection we will give an overview the software metrics used in the thesis. Designed to provide a summary of the overall quality of object oriented projects, Chidamber and Kemerer proposed the first suite of object oriented systems design measures known as CK Metrics [24], Abreu defined metrics for OOSS design known as MOOD metrics [25], and McCabe's Cyclomatic Complexity index [26]. This results in the following list of metrics:

- *Coupling between objects* **(CBO)**: for a class, this is the total number of other classes to which it is coupled; two classes are coupled if methods of one use methods and/or instance variables of the other. It assesses the dependency of one class on other classes; high CBO indicates a complex design.
- *Lack of cohesion* **(LCOM4)**: measures the number of connected components in a class. A connected component is a set of related methods (and class-level variables). There should be only one such component in each class. If there are two or more components, the class should be split into smaller classes [27].
- *Depth of inheritance* **(DIT)**: the maximum path length from a class to the root class in the inheritance hierarchy. Measures the depth of class hierarchy, the deeper a class is in the hierarchy, the more methods and variables it is likely to inherit, making it more complex [28].
- *Response for a class* **(RFC)**: the total number of methods that can be executed in response to a message to a class. This count includes all the methods available in the whole class hierarchy.
- *Weighted methods per class* **(WMC)**: a weighted sum of methods in a class (tools calculate the WMC metric as simply the number of methods in a class). It measures the total complexity for all class methods.
- *Number of children* **(NOC)**: number of immediate sub-classes of a class. It measures the breadth of a class hierarchy, a high NOC might indicate high reuse.
- *Coupling factor* **(COF)**: the ratio of the actual couplings among all classes to the maximum number of possible couplings. Couplings due to inheritance are not considered.
- *Cyclomatic complexity* **(CC)**: a metric based on the control flow of the source code, calculating the number of independent paths through a procedure/function. A high CC means a program is hard to understand and hence to maintain. It is calculated from the control flow graph of the program using the formula:

  $CC = e - n + 2p$ where e is the number of edges, n is the number of nodes in a graph and p is the number of connected components, which for a single component is one, hence:
  $CC = e - n + 2$.

2.3 Software quality

Software structural quality refers to how a software system meets non-functional requirements that support the delivery of the functional requirements, such as robustness or maintainability. It has a lot to do with the degree to which the software works as needed.

The ISO/IEC 25010 quality standard also known as SQuaRE (Software product Quality Requirements and Evaluation)[2] classifies software according to following eight characteristics:

1. *Functional suitability:* the degree to which the system provides desired functions.
2. *Performance efficiency:* the system performance relative to the amount of resources used under stated conditions
3. *Compatibility:* the degree to which the system/components can exchange information with other systems/components, while working cross platform.
4. *Usability:* the degree to which the system can be used by specific users.
5. *Reliability:* the degree to which a system maintains the level of performance when used under specified conditions.
6. *Security:* the degree to which information and data protection is ensured by the system, so that persons or other products or systems have the degree of data access appropriate to their types and levels of authorization.
7. *Maintainability:* the degree of effectiveness and efficiency with which a product or system can be modified to improve it, correct it or adapt it to changes in environment, and in requirements
8. *Portability:* the ability of software to be transferred from one environment to another.

Each of the characteristics above are further subdivided. Since our focus in this thesis is on maintainability we will only describe maintainability in more detail.

There are many definition of maintainability in literature. One definition is *"The ease with which a software system can be corrected when errors or deficiencies occur and can be expanded or contracted to satisfy new requirements"* [29]. Maintainability is a key to other software quality attributes. High maintainability means less effort to understand and apply changes to the system, to improve performance or fix an issue.

Maintainability depends on the following factors [30]:

1. *Understandability/Analysability:* effort needed to analyze and identify which part to be modified for the maintenance task required. Complex architecture is harder to understand, hence the aim is to make it make it as simple as possible.
2. *Testability:* how testing scenarios are planned and carried out, this also depends on how simple or complex the system is.
3. *Modifiability:* degree to which a system can be modified without introducing defects or degrading quality.
4. *Modularity:* degree to which a system's components may be separated and recombined, such that a change to one component has minimal impact on other components.

[2]ISO/IEC 25010 https://www.iso.org/standard/35733.html

5. *Reusability:* the extent of which the software in question can be integrated it in a another software system.

Later, we use the description in section 4 to guide in selecting the appropriate software metrics that are indicators for the maintainability characteristic.

3 Related work

In this section we will discuss related work on software networks, software networks for evolving software, and the mapping of network and software metrics.

3.1 Software networks

Applying complex network theory in software engineering was introduced as early as 2002, where an undirected network was constructed (a class dependency network) using class diagram where nodes denote classes and edges denote relationships (inheritance, aggregation, etc), and concluded that software networks in the study exhibited small world and scale-free characteristic [15].

Emphasizing the importance of the control flow aspect of software systems *Myers* used directed software graphs with edges as class collaboration. The study concluded that irrespective of the programming language, software networks have small world and scale-free characteristics. Due to the directionality of the network, he was able to uncover the difference between in and out-degree distribution, the negative correlation between them, and the weak positive assortativity among out-degree as an indication of the hierarchical layering of functionality [8]. Using directed network for class diagram not only Small world and scale free characteristics were detected, but also the networks shared features of hierarchical networks(small average path, and higher than random expectation clustering coefficient) [13].

With increasing interest in software networks, various types of networks have been proposed and used to analyze and study diverse software engineering aspects. For instance using the original class dependency networks [31] revealing community structure, with the communities having scale free and small world properties. Using class, method and package collaboration graphs, to analyze the three levels of granularity through the development phase [32], their findings were in line with Myers's results in all levels. Empirical studies were conducted to validate the scale-free and small-world characteristics hold for different software systems and languages [33, 34, 35].

3.2 Software networks and metrics

Scholars started applying SNA metrics to software networks. Using the lowest level of granularity, [16] applied SNA metrics to a system wide binary dependency network. The conclusion of the study is that SNA network measures can identify critical binaries that are missed by complexity metrics, and can indicate and predict the number of defects. In a follow up study to validate the mentioned findings, applying SNA metrics to class/method dependency graph [18], the result was that SNA metrics are strong indicators of defective modules for large and complex systems, but not for small scale systems.

With their seminal work, Concas *et al.* studied the distribution of SNA metrics in software network then comparing them with CK metrics [19]. In [36], their findings showed negative correlation between structural metrics and the number of bugs, concluding that SNA metrics can be utilized as quality indicators.

3.3 Evolving software networks

Software evolution is inevitable, and monitoring software structure changes and measuring its quality is the way to ensure that the probability of introducing faults is eliminated and/or mini-

mized. Software networks can be used as means to facilitate understanding, monitoring structure changes due to system maintenance and evolution.

To understand how change is distributed in an evolving software, studying type dependency graphs of evolving software system over the span of two years was considered [20]. Applying degree-based metrics the results showed that around 20% of classes are changed and about 8% are added. This finding is useful in expecting where growth and change will occur in the software system.

Using complex networks to study the evolution of Linux kernel (223 versions) [37], directed call graphs of different components of the Linux Kernel (file system, Kernel , memory management and net) were constructed and analyzed. Preferential attachment growth was observed, and all call graphs showed scale-free and small-world characteristics. The researchers also proposed a method to find major structural change during system evolution.

To study and predict the severity of bugs an analysis of evolving software topology of eleven popular open source systems was performed [23]. This was done by building multi version software networks on three levels; function call graphs, module collaboration graphs on the system level, and bug based developer collaboration graphs on the development level. They also proposed the NodeRank metric similar to page rank that proved powerful in bug detection.

Applying motif frequency concept to study relationship between motifs and defects in evolving software has also been considered. In [38] software networks were represented by motif frequency occurrence, the study was conducted on the evolution of three systems with more than 30 releases. The conclusion was drawn that motifs are present through system evolution, and that defects and some motifs are correlated.

Constructing multi-version multi-granular software network (methods, class, package), Pan *et al.* traced network parameter throughout the software versions [39]. The findings showed that all the networks exhibited small world, scale free phenomenon and linear growth. In addition class and package networks are both disassortative, while method level network is assortative. Further examination of modularity helped in providing insight identifying where refactoring might be needed. In a following research they studied the communities in the evolving software [40, 41] and managed to identify where refactoring was needed at both the class and package level. Results were validated with software engineers.

We will be following the idea of tracing parameters throughout the software versions, working with directed networks, using a multi granularity software network, considering concept of network metrics as an indicator of software metrics.

4 Approach

This section describes our approach, subsection 4.1 explains our choice of software metrics answering *RQ1*. Subsection 4.2 describes the steps of data collection preparation, network construction, and computation of metrics and measure.

4.1 Metric selection

Software metrics

A considerable amount of literature has been published on the link between the software architecture and software quality attributes including maintainability, exploring methods to quantify them [30, 42, 43, 44, 45, 46, 47, 48].

Papers written on assessing maintainability using software metrics have been reviewed. An empirical study on the ability to predict class quality and fault proneness was conducted comparing metric suits (CK, MOOD) [49]. The results show that the CK metrics suite produced the best result for class quality and fault proneness prediction, while MOOD failed in class quality. Metrics that are associated with modularity and encapsulation were found to be good predictors for maintainability and fault proneness. High LCOM4 was linked to higher maintenance effort [50], WMC and LCOM4 were good predictors of maintenance effort [50, 51, 52]. The interaction between CBO and LCOM4 proved to be a good indicator of difficult to maintain classes [30], hence using the product of CBO and LCOM4 as the structural complexity metric [53].

We identified the following metrics as good indicators of the six maintainability factors as shown in table 1, addressing RQ1 of the thesis.

Below is a description of each metric and what it evaluates:

- CBO: low CBO means high modularity, less complexity, and easier reuse. High CBO means higher inter-dependency with other modules, making it harder to understand. CBO essentially evaluates reusability.
- LCOM4: high cohesion indicates modularity. Lack of cohesion increases complexity. This metric evaluates reusability.
- DIT: the deeper a class is within the hierarchy, the greater the number methods and variables it is likely to inherit making it more complex. Deeper trees also means the possibility of method reuse. DIT metric evaluates reuse but also relates to understandability and testability.
- RFC: a large number of methods are executed in response to a message, making it harder to understand and test. This metric evaluates testability and understandability. High RFC means more effort required for testing, greater design complexity and fault-proneness.
- WMC: classes with a larger number of methods are harder to understand, tend to be application-specific, making it harder to reuse. Changes applied on these classes have impact on children and thus have a direct effect on maintainability.

- NOC: the greater the number of children, the greater the reusability since inheritance is a form of reuse. If a class has a large number of children, it may require more testing of the methods of that class, thus increase the testing time. It primarily evaluates reusability and testability.
- COF: high value of coupling factor in the system leads to larger complexity, lower coupling means high modularity. This metric evalutes understandability, modularity, and testability.
- CC: high complexity number means greater probability of errors with increased time to maintain and troubleshoot. This metric evaluates understandability and testability.

Network metrics

The following network metrics have been selected for this study:

- *Density*: density measure the connectedness of a graph, high density would mean low modularity, a complex software system, hence low understandability and less reusability.
- *Degree distributions*: the relationship between in-degrees (as proxy for the number of called classes, methods or functions) and out-degrees (as proxy for the number of calling classes, methods or functions. Nodes with high degree k_{in} indicate a popular node could mean high reuse, modifiability and testability.
- *Communities*: high values indicates modularity which in turns means ease of modifiability and understandability.
- *Diameter*: measures the maximum distance between nodes, so a high value means higher runtime stacks affecting testablility, stability and modifiability.
- *Average clustering coefficient* ($\overline{CC}$): measures neighborhoods connectedness; high values indicates high coupling and lower modularity and stability.
- *Average shortest path length* ($\overline{d}$): measures how interconnected a network is, it could be used as an indicator for testability, modifiability and understandability.
- *Motifs*: high occurrence of certain motifs could mean high reusability, ease of testability, and understandability.

Table 1: Software and Network as Maintainability indicators

Maintainability Factors	**Software Metrics** *(literature)*	**Network Metrics** *(hypothesis)*
Understandability	WMC, RFC, CC, CBO, DIT, COF	Motifs, Communities, Density, $\overline{d}$, $\overline{CC}$
Testability	RFC, DIT, NOC, COF, CC	Diameter, Motifs, $\overline{d}$, γ_{in}, γ_{out}, γ
Reusability	WMC, CBO, NOC, DIT	Motifs, Density, γ_{in}, γ_{out}, γ
Modularity	CBO, COF, LCOM4	Density, Communities, $\overline{CC}$
Modifiability	WMC, RFC, LCOM4, CBO, CC	Diameter, Communities, $\overline{d}$, γ_{in}, γ_{out}, γ

4.2 Data collection and preparation

In this subsection we describe the steps followed to calculate the software and network metrics. The main steps are: collecting the software systems source code, constructing networks, and computing metrics. This is schematically shown in figure 3.

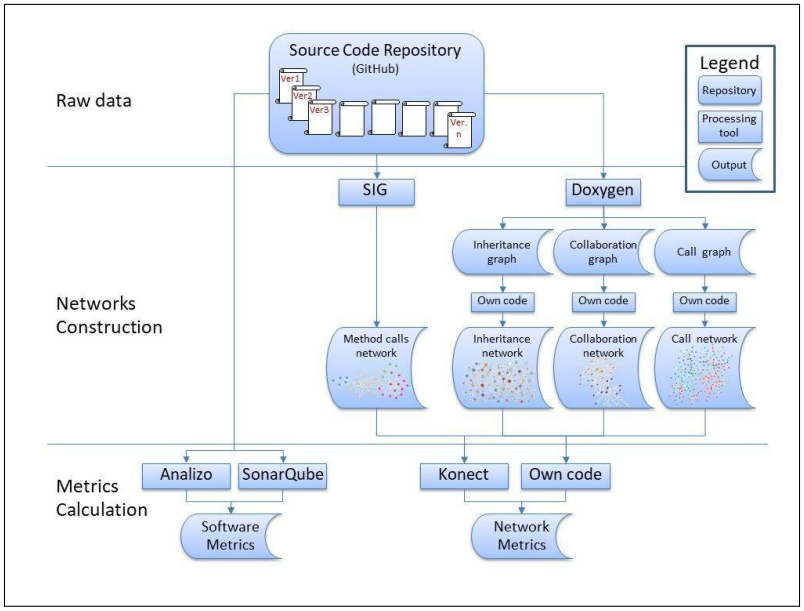


Figure 3: Data collection and preparation steps

1- Raw data: software system source code from Github repository

2- Network construction: SIG[3] proprietary tool, and Doxygen[4]

3- Metrics calculation: using Analizo[5] , SonarQube[6] and konect[7] toolbox

4.2.1 Software systems source code

We have selected different Apache software systems, which are object-oriented open-source systems. The selected systems provided different system sizes, are being actively updated, have wide range of versions from 6 to 82 (major, minor and revisions) and with life spans between 4 and 11 years.

The source code of the systems in the study was obtained from Apache releases in the Github repository[8]. As seen in table 2 they vary in number of versions, time span and number of contributors. Below is a short description of the systems.

Apache Cassandra is an open-source NoSQL distributed database management system, providing high availability, performance, and linear scalability with no single point of failure.

[3]SIG, Software Improvement Group https://www.softwareimprovementgroup.com/
[4]Doxgen https://www.doxygen.nl/index.html
[5]Extensible Multi-Language Source Code Analysis and Visualization Toolkit https://www.analizo.org/
[6]Software evaluation tool based on SQuaRE httphttps://www.sonarqube.org/
[7]The KONECT Project http://konect.cc/
[8]e.g ant-ivyhttps://github.com/apache/ant-ivy/releases

Apache Chukwa is an open source data collection system for monitoring large distributed systems.

Apache Hadoop is a collection of open-source software utilities that provides a software framework for distributed storage and processing of big data.

Apache HttpComponents is an open-source system for creating and maintaining a toolset of low level Java components focused on HTTP and associated protocols.

Apache Ant-Ivy is a tool for managing (recording, tracking, resolving and reporting) project dependencies.

Apache Jena is a framework for writing Semantic Web applications.

Table 2: Systems analyzed

System	**Versions**	**Time Span** *(mm/yy)*	**Contributors**
Cassandra	82	09/09 - 09/13	325
Chukwa	6	05/09 - 12/15	10
Hadoop	55	04/06 - 06/14	357
HttpComponents	33	07/08 - 05/15	53
Ant-Ivy	18	12/06 - 01/13	20
Jena	32	06/12 - 03/21	70

4.2.2 Network construction

As depicted in figure 3, after getting the software source code from the Github repository, the different networks are constructed. This was done using two approaches. One is to generate the inheritance, collaboration and call UML diagrams using Doxygen and then build the directed networks for each version using a program we developed. The other is using SIG proprietary software to create a network of method calls.

Doxygen is a document generator for object oriented software, it produces inheritance graphs, collaboration (inheritance & dependencies) graphs and call (method/global functions) graphs. We utilize the so- called graph dot file to build networks. For each software system version Doxygen will generate separate diagrams for each class inheritance.

Software Improvement Group (SIG) is a company specialised in improving the health and security of software. They have developed a proprietary software that constructs networks for method calls.

The next step was to construct the networks, we wrote a program that goes through the Doxygen dot graph files of the diagrams, extracts the edges (inheritance, dependencies, and method/-function calls) and the nodes (classes, method/global function) in edges and nodes files. This is demonstrated in figure 4, showing an example of two call graphs, and figure 5 showing the resulting call network.

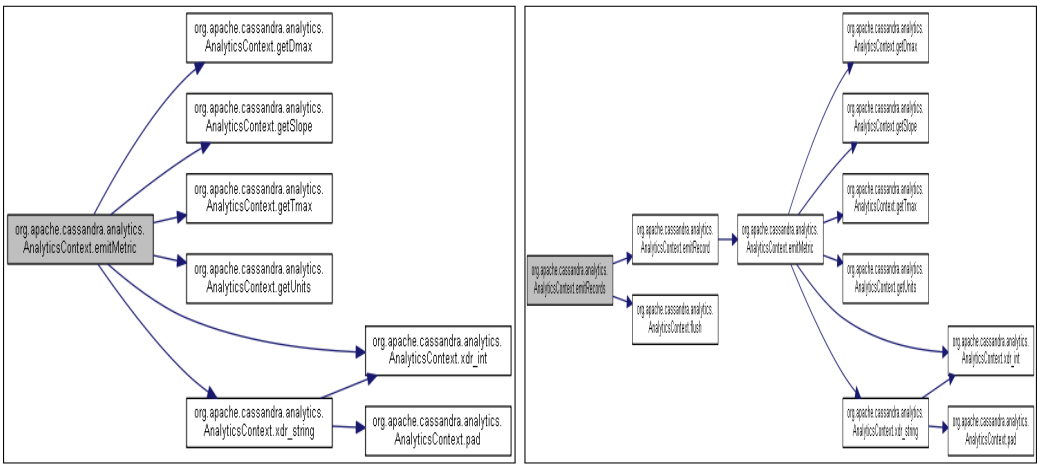


Figure 4: Two call graphs
generated by Doxgen, using Graphviz[9] to visualize the Dot files

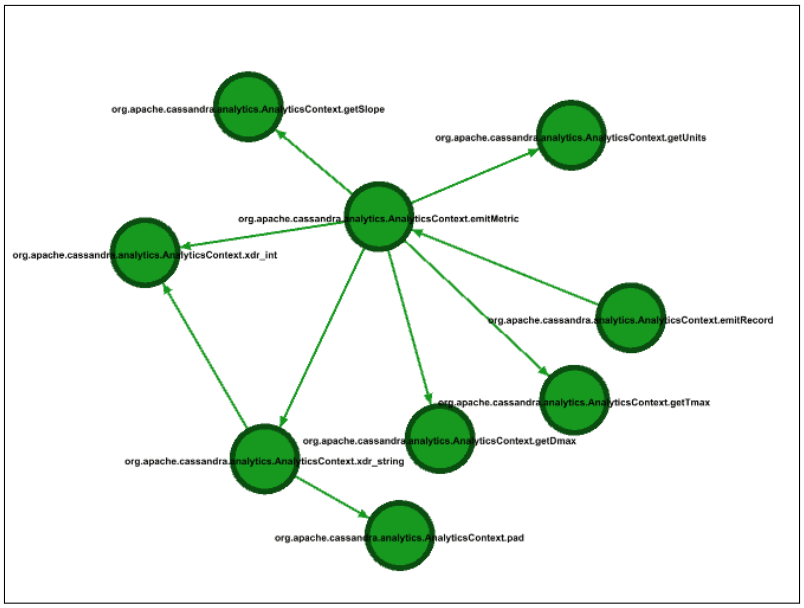


Figure 5: Constructed network from dot files
created from dot files of figure 4

Below are descriptions of constructed software networks and their level of granularity:

- Class level
  - inheritance network: nodes are classes and the edges are inheritance relations.
  - Collaboration network: nodes are classes and the edges are the class dependencies (inheritance, aggregation, composition and class reference variables).
  - Method-call network: nodes the are class methods and the edges are the method calls (generated by SIG).
- Subroutine level
  - Call network: nodes are global functions or class methods and the edges are the calls.

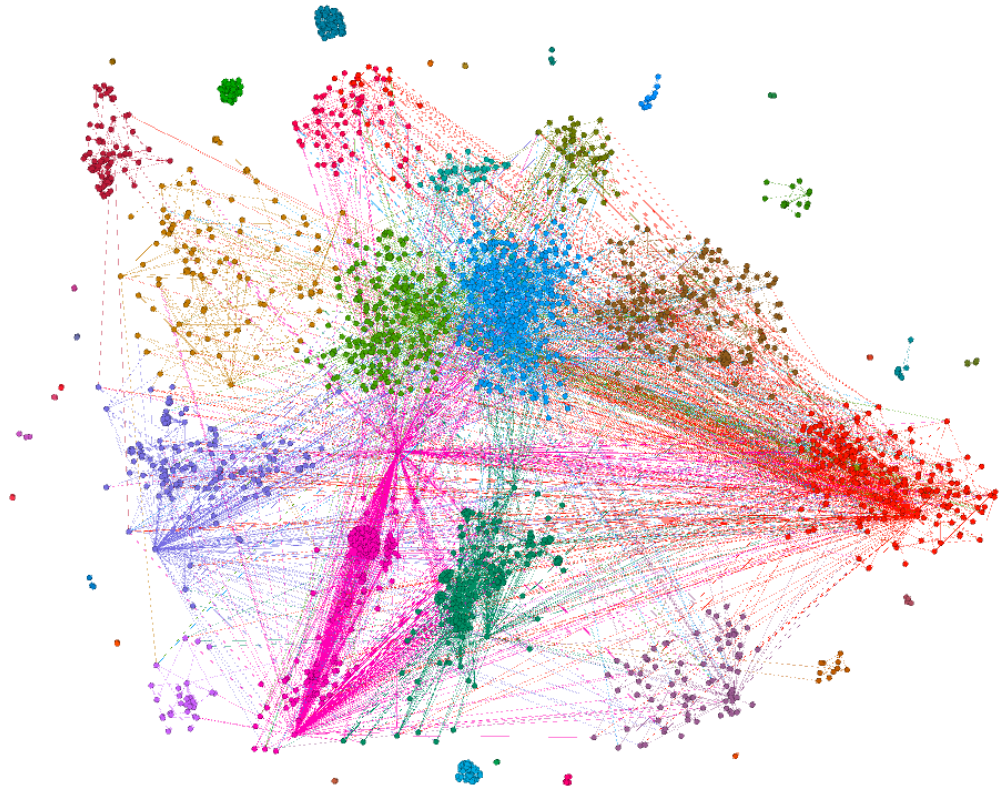


Figure 6: Cassandra-2.0.0 collaboration network - 2213 nodes and 5560 edges

An illustration of a software network is shown in figure 6, created using Gephi[10] and coloring the nodes according to communities.

4.2.3 Metrics computation

Network metrics

We calculate the network metrics using the NetworkX package in python, and Konect toolbox in matlab. The following network measures and metrics are calculated:

1. *NetworkX:* number of nodes, average degree, in and out degree distributions and their respective powerlaw parameter γ, average shortest path (based on the undirected largest WCC), network density number of SCC, number of WCC, largest WCC (nodes, degrees), average clustering coefficient, number of communities.

2. *Konect:* Network diameter, motifs (two ,three, and four stars - see figure 2), degree distribution exponent γ (treating the network as undirected).

[10]Gephi is an open-source visualization software for networks. https://gephi.org/

Software metrics

We used Analizo to compute software metrics. Analizo provides project-level metrics (calculated for the entire project) and module-level metrics (calculated individually for each module). Below are the metrics produced by Analizo. It should be noted that we will use only the underlined metrics as explained in subsection 4.1.

- Project level metrics: Total Coupling Factor (COF), Total Lines of Code, Total number of methods per abstract class (WMC), Total Number of Modules/Classes, Total number of modules/classes with at least one defined attributes, Total number of modules/classes with at least one defined method, Total Number of Methods.

- Module level metrics: Afferent Connections per Class, Average Cyclomatic Complexity per Method(CC), Average Method LOC, Average Number of Parameters per Method, Coupling Between Objects (CBO), Depth of Inheritance Tree(DIT), Lack of Cohesion of Methods (LCOM4), Lines of Code, Max Method LOC, Number of Attributes, Number of Children (NOC), Number of Methods, Number of Public Attributes, Number of Public Methods, Response For a Class (RFC).

On a project-level, Analizo also provides basic descriptive statistics for each of the module-level metrics: sum, mean, median, mode, standard deviation, variance, skewness and kurtosis of the distribution, minimum, and maximum value. Analizo also provides a change cost metric which calculates the percentage of software components affected when one component in the system is changed - a measure of the coupling degree, and structural complexity which is the product of CBO and LCOM4 which is used as measure of how attractive is the software system to contributors [54].

SonarQube is another tool that evaluates the eight software characteristics explained in subsection 2.3. SonarQube provides a score for each attribute along with a value cognitive complexity (Cog.c) as a measure for understandability.

5 Experiments

In this section we examine the data with the aid of plots and tables to track changes in the software systems, and describe methods utilised to analyze the metrics mentioned in section 4.2.3. In subsection 5.1 we describe the experimental setup. Subsection 5.2 presents the result and data descriptives. Subsection 5.3 attempts to identify SNA metrics that are good indicators of changes in maintainability, providing an answer to RQ2. Finally, subsection 5.4 identifies which network granularity best captures the maintainability factors, addressing RQ3.

5.1 Experimental setup

A number of tests were performed following the computation of the metrics to validate that results are aligned with the expectations as per literature.

We fitted the degree distributions of the networks to powerlaw distribution, and performed a comparative analysis against lognormal and exponential distributions using the *powerlaw* python package. We also compared plots across system versions of basic network metrics and corresponding software metrics (e.g., nodes vs. functions/classes) to confirm consistency of the network representation of the software systems.

Moreover, we computed both Pearson correlation and the non parametric Spearman rank correlation between software and network metrics on all networks. Figure 8 is an example of a resulting heatmap of such a correlation. Finally, we identified the highly correlated metrics and drew scatter plots of the correlated metrics preserving the time factor (versions), for example as in figure 7 (for more see Appendix D).

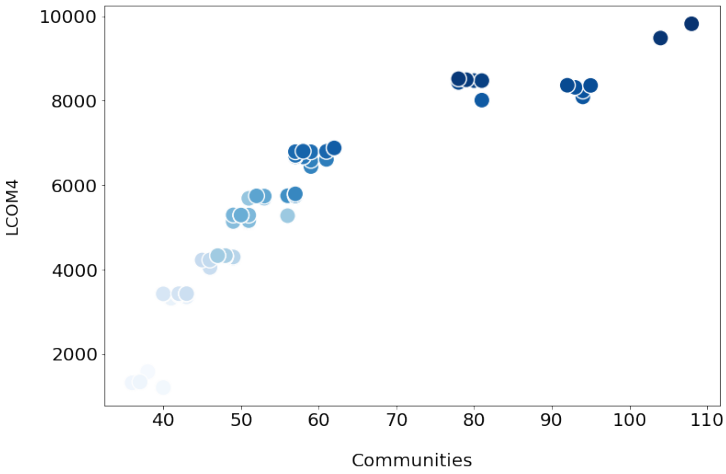


Figure 7: Scatter plot - Network vs. software metrics

Cassandra collaboration network
the color darkness increases with version number

5.2 Descriptives of data

To confirm that constructed software networks follow a powerlaw degree distribution, we fitted the degree distributions of the networks to powerlaw distribution, and compared it to lognormal

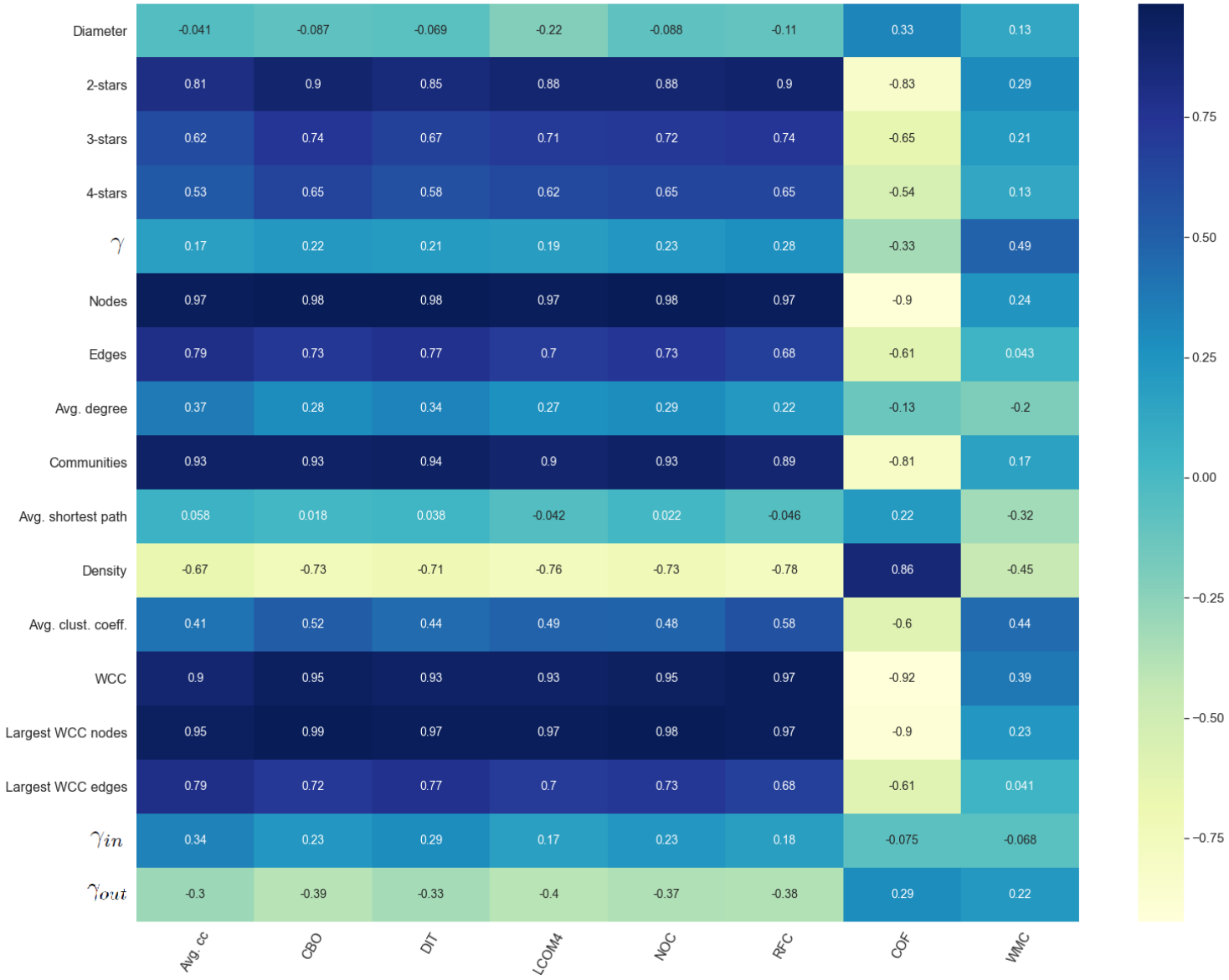


Figure 8: Collaboration network with software metrics heatmap
Cassandra Pearson correlation

and exponential distributions. Results always supported the powerlaw distribution. Most versions had a positive value for R, and three versions had a negative value for R combined with a p value > 0.05 (even though some versions had values less than two for γ, γ_{in} and γ_{out}). As example of such a distribution is shown in figure 9.

Network metrics were compared to software metrics and changes in software architecture. This comparison confirmed consistency of the network representation for the software systems. As an

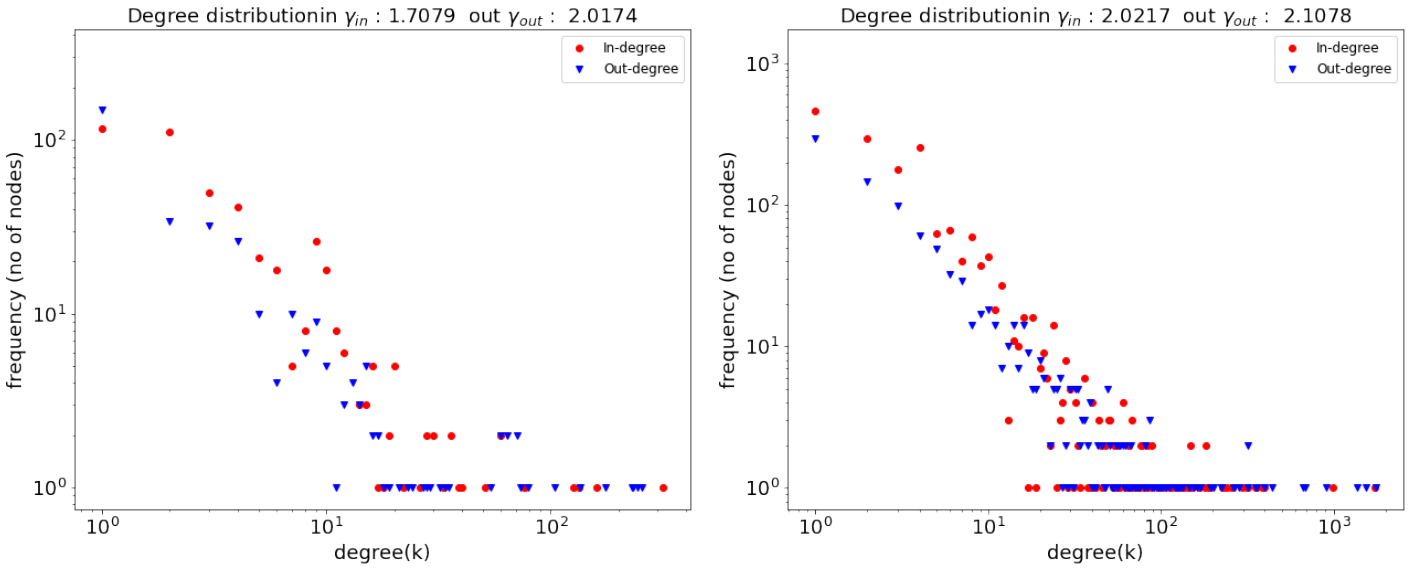


Figure 9: Collaboration network in and out degree distributions
for the first and last versions of Cassandra

example, the graphs shown in figure 10 compare the changes in number of functions to changes in the number of nodes in the call networks, demonstrating that they follow the same pattern. The same applies to collaboration networks with number of files. Both inheritance and method network nodes show the same pattern as the number of classes (except for Cassandra in method networks). The full set of graphs is to be found in Appendix A.

We observed from the graphs of software metrics across versions that NOC increases at a slow rate in all systems except for Hadoop. CBO, LCOM4, average cyclomatic complexity and DIT follow the same pattern. WMC increases, except for the Chukwa data set. COF decreases and RFC increases in general.

Analizo calculates software metrics per module and aggregates them at project level. Some of the values, like CBO, in some systems had mode of zero, median of 1, skeweness of 6 and Kurtosis of 64. For this reason we have used *total* instead of *mean*.

Similarly, network metrics for the same software systems show that the number of nodes, edges, communities, average clustering coefficient, WCC and size (nodes & edges) of the largest WCC increase in all network as they evolve. We also notice that density and average shortest path decrease over time (except for the HttpComponents and Ant-Ivy data sets).

Average degree slightly increases over time for call and collaboration networks suggesting that functions get called slightly more as the software evolves. In the inheritance and method networks average degree does not change much indicating that the class inheritance is stable. We observe these changes through the versions in all granularity.

It should be noted that the average shortest path length did not exhibit correlation with any of the software metrics; this could be due to the fact it was calculated using the undirected largest weakly connected component. Tables 3 to 7 show changes to network and software metrics between first and last versions of each of the software systems. These results demonstrate the growth of each evolving software system as expected.

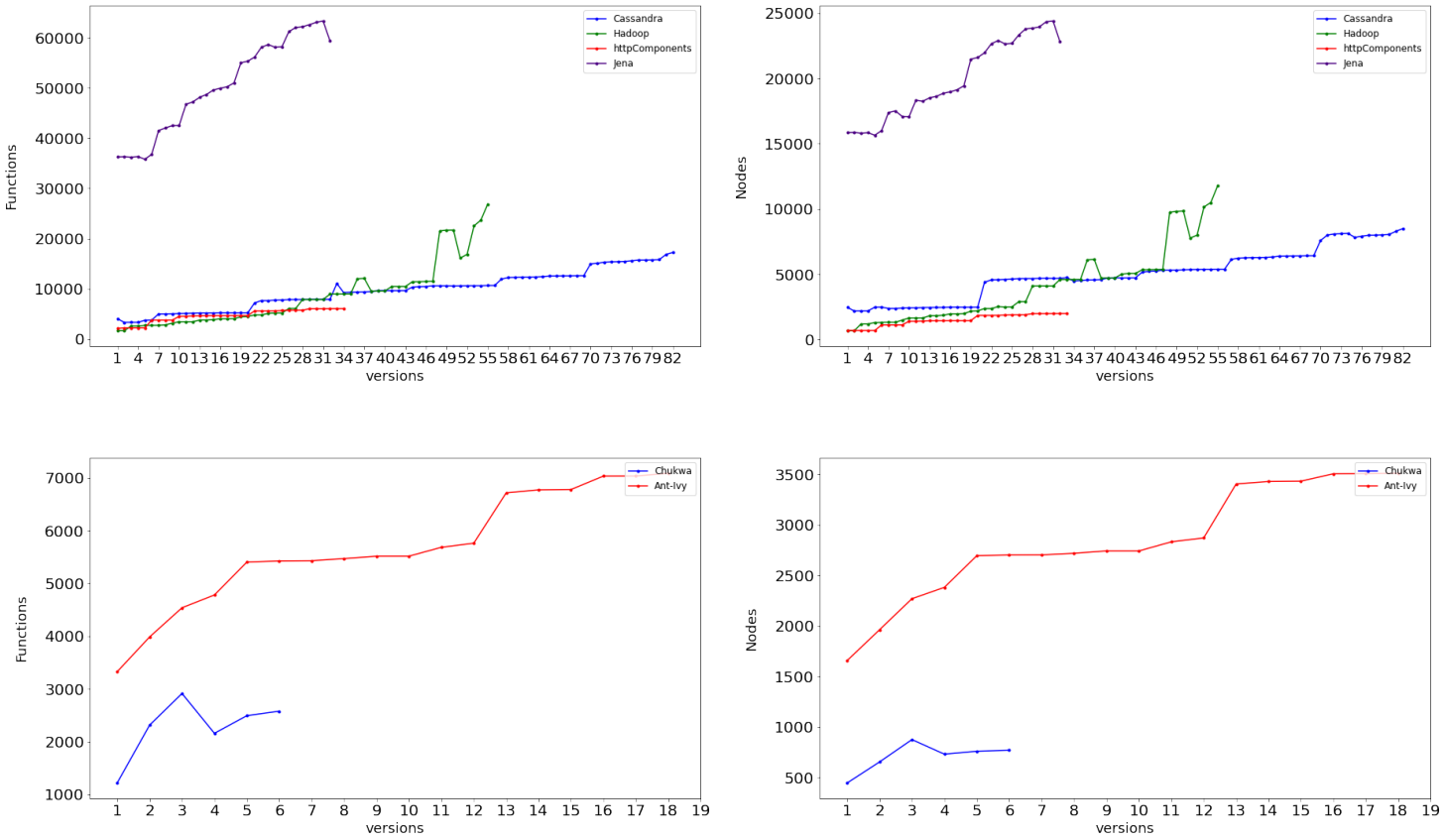


Figure 10: Software and call network metrics change as software evolves
Functions on the left vs. Nodes on the right

Table 3: Collaboration network metrics of first and last version of each considered software systems

Software system	Cassandra		Chukwa		Hadoop		HttpComponents		Ant-Ivy		Jena	
Version	First	Last	First	Last	First	Last	First	Last	First	Last	First	Last
Diameter	11	13	12	15	10	23	11	8	9	10	13	12
2-star motifs	39652	315148	12562	39756	624	24014	2761	66272	12612	60145	669856	919150
3-star motifs	984630	24739164	188533	1146088	2652	387355	11406	7228585	309908	3954509	165396905	211036242
4-star motifs	22964186	1897523572	2871700	32375828	11632	7142165	49908	628957445	8031742	254111772	37433388165	45590011797
Nodes	656	2242	486	1012	285	3340	329	653	337	746	3405	4665
Edges	3285	22244	1840	4147	2213	32192	1089	1999	4040	8525	31880	35848
Avg. degree	5.0076	9.9215	3.7860	4.0978	7.7649	9.6383	3.3100	3.0613	11.9881	11.4276	9.3627	7.6845
Communities	38	108	43	78	39	148	32	42	41	71	138	155
Avg. shortest path	3.7951	4.0982	3.5740	3.6828	3.6095	3.8555	4.9436	3.0172	3.4399	3.4189	4.0468	4.2332
Density	0.0076	0.0044	0.0078	0.0041	0.0273	0.0029	0.0101	0.0047	0.0357	0.0153	0.0028	0.0016
Avg. clust. coeff.	0.0511	0.0290	0.0263	0.0232	0.0368	0.0468	0.0226	0.0401	0.0551	0.0795	0.0508	0.0344
SCC	652	2229	486	1008	283	3311	329	653	337	741	3399	4655
WCC	11	29	23	45	9	72	9	10	7	9	38	54
Largest WCC nodes	617	1964	309	574	243	2983	303	622	321	676	3272	4382
Largest WCC edgeds	3256	21969	1530	3233	2172	31845	1071	1977	4027	8433	31716	35499
γ	2.0534	1.9959	2.1368	2.1576	4.0006	3.5802	2.2884	2.2618	2.0070	2.0160	2.1135	2.1380
γ_{in}	1.7079	1.9616	1.6230	1.6201	2.0476	1.8789	1.7777	1.7835	2.1603	2.1840	1.9829	1.8696
γ_{out}	2.0174	1.8149	1.8577	1.5845	2.2605	2.0086	1.7852	1.8667	2.5811	2.1185	1.8740	1.6973

Table 4: Call network metrics of first and last version of of each considered software systems

Software system	Cassandra		Chukwa		Hadoop		HttpComponents		Ant-Ivy		Jena	
Version	First	Last	First	Last	First	Last	First	Last	First	Last	First	Last
Diameter	19	24	25	18	10	23	20	25	14	19	28	28
2-star motifs	33411	338467	3344	11200	624	24014	3580	13912	38558	94636	380899	671434
3-star motifs	356434	24845408	13286	200715	2652	387355	10741	53619	372313	1368114	8750948	24836252
4-star motifs	5404181	2300625150	55157	4534881	11632	7142165	37313	228517	4264660	27178772	320096001	1716183736
Nodes	2456	8488	447	771	671	11778	698	1988	1657	3509	15838	22805
Edges	18182	51864	1408	2475	2048	67422	1579	8088	14948	33231	173263	205010
Avg. degree	7.403094463	6.1103	3.1499	3.2101	3.0522	5.7244	2.2622	4.0684	9.0211	9.4702	10.9397	8.9897
Communities	239	622	54	63	96	683	75	189	100	206	875	1138
Avg. shortest path	5.35734778	6.0927	7.8454	5.7950	6.7487	7.2354	7.7528	8.8761	4.9172	5.3570	7.6360	7.2796
Density	0.003015517	0.0007	0.0071	0.0042	0.0046	0.0005	0.0032	0.0020	0.0054	0.0027	0.0007	0.0004
Avg. clust. coeff.	0.016337946	0.0259	0.0234	0.0343	0.0164	0.0201	0.0096	0.0149	0.0327	0.0313	0.0314	0.0279
SCC	2452	8446	445	768	670	11721	697	1984	1636	3421	15545	22587
WCC	165	497	31	36	66	533	56	127	38	109	610	886
Largest WCC nodes	1376	6842	346	652	497	9221	423	1482	1542	3191	14195	20197
Largest WCC edgeds	16456	50302	1322	2356	1911	63624	1090	7238	14857	32983	171321	199978
γ	2.2040	2.1797	2.4507	2.3646	4.0006	3.5802	2.5473	2.4207	2.0558	2.0773	2.0930	2.1106
γ_{in}	1.71087038	1.6009	1.8391	1.6937	1.7346	1.5702	1.6841	1.7643	1.8363	1.6399	1.6524	1.5961
γ_{out}	1.654887728	1.5627	1.6731	1.6131	1.6749	1.5654	1.6751	1.7448	1.7058	1.7541	1.5938	1.5768

Table 5: Inheritance network metrics of first and last version of software systems

Software system	Cassandra		Chukwa		Hadoop		HttpComponents		Ant-Ivy		Jena	
Version	First	Last	First	Last	First	Last	First	Last	First	Last	First	Last
Diameter	10	17	5	8	10	23	8	18	9	8	19	30
2-star motifs	8464	51357	356	538	624	24014	313	1137	485	2113	17909	29966
3-star motifs	184736	1830213	1465	1812	2652	387355	416	2319	1457	21517	162269	379465
4-star motifs	3784540	54824289	6091	6648	11632	7142165	546	5491	5103	235237	1701357	5420002
Nodes	472	1672	163	246	182	2489	232	525	220	465	2900	3702
Edges	834	3463	223	318	459	5484	508	1343	721	1591	13182	15612
Avg. degree	1.7669	2.0712	1.3681	1.2927	2.5220	2.2033	2.1897	2.5581	3.2773	3.4215	4.5455	4.2172
Communities	69	208	40	58	40	284	63	112	51	85	336	371
Avg. shortest path	4.6128	5.0560	1.9091	1.9921	2.9463	8.2514	3.2277	8.0944	4.6090	2.3005	6.9136	11.4854
Density	0.0038	0.0012	0.0084	0.0053	0.0139	0.0009	0.0095	0.0049	0.0150	0.0074	0.0016	0.0011
Avg. clust. coeff	0.0000	0.0007	0.0000	0.0000	0.0000	0.0018	0.0000	0.0004	0.0094	0.0170	0.0105	0.0098
SCC	472	1670	163	246	182	2489	232	525	220	465	2900	3702
WCC	60	192	39	56	31	246	55	96	37	66	275	300
Largest WCC nodes	235	391	22	23	41	1405	26	151	40	61	711	1384
Largest WCC edges	479	1060	76	72	136	3856	149	668	312	270	4148	8335
γ	4.0787	3.8515	4.8824	4.4234	4.0006	3.5802	4.2667	3.5787	4.3326	4.1310	3.5129	3.4835
γ_{in}	1.3831	1.4696	1.4606	12.2442	1.5936	1.5851	2.3550	1.6499	1.6323	1.5348	1.6862	1.6291
γ_{out}	1.9702	1.7869	1.8349	1.6911	2.0449	1.7109	1.8719	1.8818	2.5619	1.8063	1.7855	1.8564

Table 6: SIG network metrics of first and last versions of each considered software systems

Software system	Cassandra		Chukwa		Hadoop		HttpComponents		Ant-Ivy		Jena	
Version	First	Last	First	Last	First	Last	First	Last	First	Last	First	Last
Diameter	22	22	18	19	23	28	23	24	13	16	28	28
2-star motifs	27039	92889	5503	22120	93121	249139	4263	18796	44294	100275	458878	631508
3-star motifs	115998	839080	21778	493487	1268853	4576701	15537	86859	480933	1245735	9529807	21096755
4-star motifs	698906	11946072	97019	15164787	24754218	115647533	67057	563656	6182577	21157184	315373466	1137843955
Nodes	3456	7216	695	1355	5153	11343	844	2603	1937	4191	21229	23506
Edges	5067	11683	1007	2148	9214	20768	1005	3610	3899	8467	39013	43831
Avg. degree	1.4661	1.6190	1.4489	1.5852	1.7881	1.8309	1.1908	1.3869	2.0129	2.0203	1.8377	1.8647
ommunities	331	640	64	97	274	548	96	198	94	194	891	927
Avg. shortest path	7.9726	7.5160	6.8427	6.1630	6.2838	6.8396	9.2643	9.1069	5.0093	5.6031	8.1573	8.1131
Density	0.0004	0.0002	0.0021	0.0012	0.0003	0.0002	0.0014	0.0005	0.0010	0.0005	0.0001	0.0001
Avg. clust. coeff	0.0106	0.0099	0.0211	0.0225	0.0211	0.0194	0.0113	0.0121	0.0250	0.0259	0.0194	0.0186
SCC	3451	7188	695	1354	5142	11316	844	2603	1923	4166	21009	23294
WCC	290	564	50	76	206	434	82	166	60	144	733	766
Largest WCC nodes	2275	5213	487	1140	4273	9762	560	1863	1753	3840	19062	21263
Largest WCC edges	4085	10179	809	2001	8354	19366	780	2924	3769	8241	37489	42278
γ	2.3526	2.3005	2.4071	2.3286	2.1688	2.1626	2.7697	2.4273	2.0806	2.0934	2.1271	2.1260
γ_{in}	1.4323	1.4441	1.6220	1.6229	1.4572	1.4645	1.5200	1.4839	1.6243	1.5609	1.4676	1.4848
γ_{out}	1.7575	1.4106	1.4761	1.6773	1.4436	1.4567	1.5004	1.6018	1.6009	1.5326	1.3393	1.3290

Table 7: Software metrics of first and last versions of each considered software systems

Software system	Cassandra		Chukwa		Hadoop		HttpComponents		Ant-Ivy		Jena	
Version	First	Last	First	Last	First	Last	First	Last	First	Last	First	Last
Average CC	956.4944	4911.9057	785.0489	1424.8883	299.8619	1115.3034	684.3593	1364.2508	634.2953	1372.2409	2238.4826	2304.2789
Var	1.8112	6.5208	4.1187	4.5228	1.0381	1.5214	1.7473	1.6236	2.2473	1.4813	3.2371	2.7931
Mean	1.6212	2.0255	2.5655	2.4738	1.5378	1.6499	1.8153	1.6457	1.6605	1.6817	1.4621	1.5230
CBO	1019	5606	409	850	263	875	529	1382	875	2276	4003	3793
Var	8.8134	20.8128	2.5191	3.0498	4.8056	4.6614	4.5232	6.4784	14.9573	21.2831	15.9769	14.8771
Mean	1.7271	2.3118	1.3366	1.4757	1.3487	1.2944	1.4032	1.6671	2.2906	2.7892	2.6146	2.5069
DIT	396	2227	252	477	179	616	379	677	404	909	3017	2709
Var	0.3501	0.4999	0.4671	0.5635	0.5499	0.7151	0.7659	0.8770	1.3353	1.3649	3.0991	2.7861
Mean	0.6712	0.9184	0.8235	0.8281	0.9179	0.9112	1.0053	0.8166	1.0576	1.1140	1.9706	1.7905
lcom4	1587	9821	594	1323	909	2531	1427	3356	950	2061	6600	7592
Var	11.0734	49.1506	3.5441	8.8004	41.4931	26.7033	24.4245	48.9373	6.2715	8.6742	72.4301	90.4937
Mean	2.6898	4.0499	1.9412	2.2969	4.6615	3.7441	3.7851	4.0483	2.4869	2.5257	4.3109	5.0178
NOC	191	723	80	156	116	304	152	348	178	395	1167	1098
Var	3.5741	4.4222	2.0036	2.0483	4.3247	2.7664	1.3104	2.0313	2.2285	4.3433	11.1735	8.7931
Mean	0.3237	0.2981	0.2614	0.2708	0.5949	0.4497	0.4032	0.4198	0.4660	0.4841	0.7622	0.7257
RFC	12258	51750	4180	9168	3507	13234	5937	20005	11848	27729	43930	43573
Var	1193.5662	2665.3632	276.4808	454.7791	529.1389	1090.7659	720.4390	4985.9187	5951.0129	8536.2193	3355.4950	3160.1117
Mean	20.7763	21.3402	13.6601	15.9167	17.9846	19.5769	15.7480	24.1315	31.0157	33.9816	28.6937	28.7991
COF	0.0034	0.0011	0.0052	0.0030	0.0105	0.0026	0.0050	0.0027	0.0076	0.0041	0.0027	0.0025
WMC	7.3889	7.9339	13.6667	7.7778	9.6000	11.9400	12.6000	14.6154	13.0870	12.3111	11.7586	12.8947

5.3 SNA metrics for maintainability (RQ2)

In this section we address RQ2, on how SNA measures and metrics capture software metrics that are strong indicators of changes in maintainability.

After examining the heatmaps (see figure 8 as an example and Appendix B for full set of heatmaps), and correlation tables (see tables 8 and 9 as example and Appendix C), the results indicate that the strongest correlation across all software systems are the *density, motifs communities* and *weakly connected components(WCC)*. This applies to all networks across all software systems.

Correlation was observed between the network metrics mentioned below and all software metrics except for WMC. A strong positive correlation is observed between communities, WCC and motifs and software metrics except for COF, which shows a negative correlation. Strong negative correlation is observed between density and software metrics except for COF, which shows a positive correlation. The correlation between WCC and software metrics is positive, but not strong.

Diameter is also a strong indicator for software metrics. Jena, as an outlier, was not considered in arriving to this observation. This is explained by the limitation of Analizo described in section 5.5.

A negative correlation has been observed between γ_{out} for call networks and all software metrics (except for COF where this correlation is positive). This correlation was not strong for the Ant-Ivy data set. The reverse of this correlation is observed for the Average clustering coefficient in call networks.

Finally, γ_{out} in collaboration networks demonstrates negative correlation with WMC, RFC, LCOM4 and NOC.

Table 8: Correlation between CBO and inheritance network metrics

	Cassandra	Chukwa	Hadoop	Http-Components	Ant-IVY	Jena
Diameter	0.6734	0.9132	0.6558	0.8192	0.1277	-0.4956
2-star motifs	0.9849	0.3821	0.6803	0.8976	0.8404	0.3157
3-star motifs	0.9796	0.0971	0.5565	0.7613	0.7729	0.3721
4-star motifs	0.9663	0.0547	0.5290	0.6296	0.7394	0.3886
Nodes	0.9945	0.8250	0.7970	0.9847	0.9650	0.2477
Edges	0.9888	0.7424	0.8041	0.9498	0.9542	0.1872
Avg. degree	0.7924	-0.9385	-0.4834	0.6968	0.7405	-0.3306
Communities	0.9969	0.9593	0.8004	0.9956	0.9630	0.3234
Avg. shortest path	0.7264	0.7634	0.3795	0.8261	-0.7878	-0.5071
Density	-0.8840	-0.8692	-0.7453	-0.9600	-0.9925	-0.2184
Avg. clust. coeff.	0.4955		0.6234	0.5528	-0.1561	-0.1847
WCC	0.9955	0.9490	0.8019	0.9954	0.9880	0.3665
Largest WCC nodes	0.9352	0.4336	0.6972	0.8364	0.7598	0.1009
Largest WCC edges	0.9550	-0.0548	0.7383	0.8032	0.3630	-0.0823
γ	0.6678	-0.8232	-0.5600	-0.8667	0.3383	0.7086
γ_{in}	-0.1611	0.6082	0.3020	-0.1584	-0.4431	-0.0462
γ_{out}	-0.5114	-0.6798	-0.6342	-0.1015	-0.4780	-0.0120

Table 9: Correlation between Average Cyclomatic complexity and collaboration network metrics

	Cassandra	Chukwa	Hadoop	Http-Components	IVY	Jena
Diameter	-0.041	0.876	0.683	-0.527	0.727	-0.662
2-star motifs	0.809	0.995	0.737	0.736	0.988	0.627
3-star motifs	0.619	0.977	0.626	0.681	0.971	0.487
4-star motifs	0.532	0.953	0.601	0.672	0.959	0.373
Nodes	0.968	0.988	0.832	0.985	0.995	0.536
Edges	0.795	0.990	0.801	0.875	0.997	0.348
Avg. degree	0.375	0.952	0.304	-0.148	-0.853	-0.568
Communities	0.930	0.942	0.834	0.907	0.936	0.459
Avg. shortest path	0.058	0.855	0.460	-0.811	-0.720	0.099
Density	-0.665	-0.998	-0.768	-0.923	-0.977	-0.522
Avg. clust. coeff.	0.408	-0.517	0.092	0.752	0.886	-0.351
WCC	0.895	0.885	0.772	0.412	0.601	0.452
Largest WCC nodes	0.953	0.997	0.831	0.983	0.996	0.548
Largest WCC edges	0.792	0.998	0.799	0.877	0.997	0.345
γ	0.173	0.158	-0.568	-0.036	0.721	0.544
γ_{in}	0.338	-0.336	-0.670	-0.157	-0.169	-0.053
γ_{out}	-0.303	-0.401	-0.581	0.723	-0.720	-0.384

5.4 Representation of structural changes by different types of networks (RQ3)

In this section we address RQ3, on how the changes in SNA metrics are different for the four levels of granularity for software networks, aiming to understand which network is a better representation of structural change.

The correlation between changes to each of the network metrics identified in section 5.3 and changes to software metrics has been compared for each of the different network granularity levels. Collaboration network has demonstrated consistently the strongest level of correlation in this comparison.

It should be noted that the inheritance network was the least to resemble a real world network. The diameter clearly correlates with DIT and NOC as demonstrated in the heat maps (see Appendix B). These observations are inline with our original expectations.

Furthermore, a comparison between the characteristics of the different networks has been carried out, see table 10. This demonstrates that collaboration networks best represent the characteristics of real world networks: small-world ($4 < average\ \ shortest\ \ path\ \ length < 6$) , scale-free ($2 < \gamma_{in}$,γ_{out} ,$\gamma < 3$) and average clustering coefficient between 0.02 and 0.083 [55].

Therefore, we concluded that changes in collaboration networks metrics best reflect changes in software maintainability metrics.

Table 10: Networks metrics for all network granularities

			Cassandra	Chukwa	Hadoop	Http-Components	Ant-Ivy	Jena
Inhertance	Avg. shortest path	min	3.8777	1.9091	2.9463	3.2277	2.2836	6.8939
		max	6.5374	1.9938	8.2514	8.1287	4.6090	12.2975
	Avg. clust. coeff.	min	0.0000	0.0084	0.0000	0.0000	0.0094	0.0086
		max	0.0036	0.0055	0.0101	0.0010	0.0281	0.0113
	γ_{in}	min	1.3331	1.4076	1.4729	1.5330	1.5334	1.5216
		max	4.7987	12.2442	1.7045	2.3550	1.7229	1.7053
	γ_{out}	min	1.6921	0.1847	1.6933	1.8189	1.7732	1.6872
		max	2.0893	0.2410	2.1495	2.0778	2.7721	1.8564
	γ	min	3.3235	4.4234	2.8771	3.5787	3.7454	3.4613
		max	4.0787	5.2579	3.8743	4.3710	4.3326	3.5403
Collaboration	Avg. shortest path	min	3.3311	3.5740	3.6062	3.0094	3.3833	4.0137
		max	4.2224	3.6828	4.2329	4.9436	3.5002	4.4592
	Avg. clust. coeff.	min	0.0238	0.0232	0.0283	0.0207	0.0514	0.0333
		max	0.0511	0.0317	0.0468	0.0439	0.0831	0.0533
	γ_{in}	min	1.4642	1.6201	1.6704	1.6689	2.0695	1.5469
		max	2.0217	1.7699	2.0476	1.8005	2.2738	2.0605
	γ_{out}	min	1.6243	1.5845	1.6309	1.7507	2.0764	1.6665
		max	2.1743	1.9508	2.2605	2.0048	2.5811	1.8779
	γ	min	1.9117	2.0745	1.8446	2.1182	1.9719	2.0901
		max	2.2944	2.1576	1.9993	2.3280	2.0323	2.1603
Call	Avg. shortest path	min	5.3573	5.7881	6.3903	7.1153	4.9172	7.2171
		max	8.2380	7.8454	8.1607	8.8956	5.4715	7.6397
	Avg. clust. coeff.	min	0.0145	0.0161	0.0163	0.0096	0.0283	0.0258
		max	0.0287	0.0343	0.0301	0.0174	0.0327	0.0314
	γ_{in}	min	1.5495	1.6841	1.5441	1.5515	1.6399	1.5796
		max	1.8154	2.1933	1.8394	1.8306	1.8363	1.7352
	γ_{out}	min	1.5753	1.5692	1.5265	1.6092	1.6165	1.5486
		max	1.7791	1.8135	1.8668	1.7965	1.9649	1.6173
	γ	min	2.1032	2.3641	2.0994	2.3902	2.0539	2.0849
		max	2.2126	2.4972	2.3696	2.5473	2.1013	2.1249
Method	Avg. shortest path	min	6.9281	6.1026	6.2269	8.7976	5.0093	8.1131
		max	8.2782	7.7431	6.8572	10.4079	5.6166	8.1573
	Avg. clust. coeff.	min	0.0091	0.0173	0.0194	0.0162	0.0244	0.0186
		max	0.0134	0.0302	0.0227	0.0162	0.0264	0.0194
	γ_{in}	min	1.4039	1.5680	1.4572	1.4238	1.4615	1.4676
		max	1.5592	1.6833	1.6116	1.5200	1.6243	1.5401
	γ_{out}	min	1.3619	1.3809	1.3891	1.4346	1.5001	1.3290
		max	1.7575	1.6773	1.4610	1.6038	1.6009	1.3530
	γ	min	2.2463	2.2427	2.1418	2.4273	2.0655	2.1236
		max	2.4109	2.4071	2.4146	2.7697	2.1129	2.1296

5.5 Possible limitations

Visual assessments of plots for metrics across versions (e.g. figure 10) has been used in this study rather than treating software versions as time series. This choice was made due to the complexity of handling major versions, minor versions and revisions (e.g. major to major vs. major to minor vs. revision to minor, ...etc.). The used software has some technical limitations. The software metrics program Analizo produces accurate results when the main program is in the root directory and remaining modules are located in sub-directories. It skips modules that are located in the same directory of the main program (as with Jena) resulting in less accurate metric values.

SonarQube uses the CK suite to calculate maintainability metrics, but it only provides the score for maintainability as a whole as its output, together with CC and cognitive complexity for software systems and versions. All versions of the six software systems selected for this study were given the same maintainability score, indicating that the sample may not be representative of the full spectrum.

The timestamps associated with some versions of the selected software systems are inaccurate (i.e., some versions have got earlier timestamps than earlier ones). Versions were sequenced manually based on version numbers and last commitment date (as per Github and Apache official site).

6 Conclusion and future work

In this exploratory study, we have identified social network analysis metrics that provide an indication of changes to maintainability of software systems as they evolve. We generated different types of networks for a sample of six evolving open-source software systems and compared their SNA metrics against conventional software metrics for maintainability. We have demonstrated that communities, weakest connected components, motifs and density are strong indicators of changes in the maintainability of evolving software systems. Moreover, we concluded that the collaboration network provides the best representation of software systems for this purpose.

We suggest two directions future work. First, combining collaboration and call networks into an aggregate network and evaluating if the SNA metrics for that network will provide better representation of changes to maintainability of software systems as they evolve, might give insightful results. Second, it may be interesting to develop an approach for representing software versions as a time series in order to enable a more in-depth statistical analysis over time.

References

[1] M. M. Lehman, "Programs, life cycles, and laws of software evolution," *Proceedings of the IEEE*, vol. 68, no. 9, pp. 1060–1076, 1980.

[2] R. Albert, H. Jeong, and A.-L. Barabási, "Diameter of the world-wide web," *nature*, vol. 401, no. 6749, pp. 130–131, 1999.

[3] R. Kumar, P. Ragbavan, S. Rajagopalan, and A. Tomkins, "The web and social networks," *Computer*, vol. 35, no. 11, pp. 32–36, 2002.

[4] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the internet topology," *ACM SIGCOMM computer communication review*, vol. 29, no. 4, pp. 251–262, 1999.

[5] A. Scala, L. N. Amaral, and M. Barthélémy, "Small-world networks and the conformation space of a short lattice polymer chain," *EPL (Europhysics Letters)*, vol. 55, no. 4, p. 594, 2001.

[6] M. E. Newman, "Scientific collaboration networks. i. network construction and fundamental results," *Physical review E*, vol. 64, no. 1, p. 016131, 2001.

[7] M. E. Newman, "The structure of scientific collaboration networks," *Proceedings of the national academy of sciences*, vol. 98, no. 2, pp. 404–409, 2001.

[8] C. R. Myers, "Software systems as complex networks: Structure, function, and evolvability of software collaboration graphs," *Physical Review E*, vol. 68, no. 4, p. 046116, 2003.

[9] C. Y. Chong and S. P. Lee, "Analyzing maintainability and reliability of object-oriented software using weighted complex network," *Journal of Systems and Software*, vol. 110, pp. 28–53, 2015.

[10] A. Potanin, J. Noble, M. Frean, and R. Biddle, "Scale-free geometry in oo programs," *Communications of the ACM*, vol. 48, no. 5, pp. 99–103, 2005.

[11] G. Concas, M. Marchesi, S. Pinna, and N. Serra, "Power-laws in a large object-oriented software system," *IEEE Transactions on Software Engineering*, vol. 33, no. 10, pp. 687–708, 2007.

[12] M. Wang and W. Pan, "A comparative study of network centrality metrics in identifying key classes in software," *Journal of Computational Information Systems*, vol. 8, no. 24, pp. 10205–10212, 2012.

[13] S. Valverde and R. V. Solé, "Hierarchical small worlds in software architecture," *arXiv preprint cond-mat/0307278*, 2003.

[14] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world'networks," *nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[15] S. Valverde, R. F. Cancho, and R. V. Sole, "Scale-free networks from optimal design," *EPL (Europhysics Letters)*, vol. 60, no. 4, p. 512, 2002.

[16] T. Zimmermann and N. Nagappan, "Predicting defects using network analysis on dependency graphs," in *Proceedings of the 30th international conference on Software engineering*, pp. 531–540, 2008.

[17] P. Meyer, H. Siy, and S. Bhowmick, "Identifying important classes of large software systems through k-core decomposition," *Advances in Complex Systems*, vol. 17, no. 07n08, p. 1550004, 2014.

[18] A. Tosun, B. Turhan, and A. Bener, "Validation of network measures as indicators of defective modules in software systems," in *Proceedings of the 5th international conference on predictor models in software engineering*, p. 5, ACM, 2009.

[19] G. Concas, M. Marchesi, A. Murgia, and R. Tonelli, "An empirical study of social networks metrics in object-oriented software," *Advances in Software Engineering*, vol. 2010, p. 4, 2010.

[20] R. Vasa, J.-G. Schneider, and O. Nierstrasz, "The inevitable stability of software change," in *2007 IEEE International Conference on Software Maintenance*, pp. 4–13, IEEE, 2007.

[21] W. Pan and C. Chai, "Measuring software stability based on complex networks in software," *Cluster Computing*, vol. 22, no. 2, pp. 2589–2598, 2019.

[22] S. Valverde and R. V. Solé, "Network motifs in computational graphs: A case study in software architecture," *Physical Review E*, vol. 72, no. 2, p. 026107, 2005.

[23] P. Bhattacharya, M. Iliofotou, I. Neamtiu, and M. Faloutsos, "Graph-based analysis and prediction for software evolution," in *2012 34th International Conference on Software Engineering (ICSE)*, pp. 419–429, IEEE, 2012.

[24] S. R. Chidamber and C. F. Kemerer, "A metrics suite for object oriented design," *IEEE Transactions on software engineering*, vol. 20, no. 6, pp. 476–493, 1994.

[25] F. B. Abreu and R. Carapuça, "Object-oriented software engineering: Measuring and controlling the development process," in *Proceedings of the 4th international conference on software quality*, vol. 186, 1994.

[26] T. J. McCabe, "A complexity measure," *IEEE Transactions on software Engineering*, no. 4, pp. 308–320, 1976.

[27] M. Hitz and B. Montazeri, *Measuring coupling and cohesion in object-oriented systems*. Citeseer, 1995.

[28] E. K. Piveta, A. Moreira, M. S. Pimenta, J. Araújo, P. Guerreiro, and R. T. Price, "An empirical study of aspect-oriented metrics," *Science of Computer Programming*, vol. 78, no. 1, pp. 117–144, 2012.

[29] J. Martin and C. L. McClure, *Software Maintenance: The Problems and Its Solutions*. Prentice Hall Professional Technical Reference, 1983.

[30] D. P. Darcy, C. F. Kemerer, S. A. Slaughter, and J. E. Tomayko, "The structural complexity of software an experimental test," *IEEE Transactions on software engineering*, vol. 31, no. 11, pp. 982–995, 2005.

[31] L. Šubelj and M. Bajec, "Community structure of complex software systems: Analysis and applications," *Physica A: Statistical Mechanics and its Applications*, vol. 390, no. 16, pp. 2968–2975, 2011.

[32] D. Hyland-Wood, D. Carrington, and S. Kaplan, "Scale-free nature of java software package, class and method collaboration graphs," in *Proceedings of the 5th International Symposium on Empirical Software Engineering*, Citeseer, 2006.

[33] M. Han, D. Li, C. Liu, and H. Li, "Networked characteristics in software and its contribution to software quality," *Computer Engineering and Applications*, vol. 42, no. 20, pp. 29–31, 2006.

[34] L. Jing, H. Keqing, M. Yutao, and P. Rong, "Scale free in software metrics," in *30th Annual International Computer Software and Applications Conference (COMPSAC'06)*, vol. 1, pp. 229–235, IEEE, 2006.

[35] H. Zhang, H. Zhao, W. Cai, J. Liu, and W. Zhou, "Using the k-core decomposition to analyze the static structure of large-scale software systems," *The Journal of Supercomputing*, vol. 53, no. 2, pp. 352–369, 2010.

[36] R. Tonelli, G. Concas, M. Marchesi, and A. Murgia, "An analysis of sna metrics on the java qualitas corpus," in *Proceedings of the 4th India Software Engineering Conference*, pp. 205–213, 2011.

[37] L. Wang, Z. Wang, C. Yang, L. Zhang, and Q. Ye, "Linux kernels as complex networks: A novel method to study evolution," in *2009 IEEE International Conference on Software Maintenance*, pp. 41–50, IEEE, 2009.

[38] J. Petrić and T. G. Grbac, "Software structure evolution and relation to system defectiveness," in *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*, pp. 1–10, 2014.

[39] W. Pan, B. Li, Y. Ma, and J. Liu, "Multi-granularity evolution analysis of software using complex network theory," *Journal of Systems Science and Complexity*, vol. 24, no. 6, pp. 1068–1082, 2011.

[40] W. Pan, B. Li, Y. Ma, J. Liu, and Y. Qin, "Class structure refactoring of object-oriented softwares using community detection in dependency networks," *Frontiers of Computer Science in China*, vol. 3, no. 3, pp. 396–404, 2009.

[41] W.-F. Pan, B. Jiang, and B. Li, "Refactoring software packages via community detection in complex software networks," *International Journal of Automation and Computing*, vol. 10, no. 2, pp. 157–166, 2013.

[42] C. F. Kemerer, "Software complexity and software maintenance: A survey of empirical research," *Annals of Software Engineering*, vol. 1, no. 1, pp. 1–22, 1995.

[43] N. F. Schneidewind, "The state of software maintenance," *IEEE Transactions on Software Engineering*, no. 3, pp. 303–310, 1987.

[44] R. S. Sangwan, P. Vercellone-Smith, and P. A. Laplante, "Structural epochs in the complexity of software over time," *IEEE software*, vol. 25, no. 4, pp. 66–73, 2008.

[45] V. R. Gibson and J. A. Senn, "System structure and software maintenance performance," *Communications of the ACM*, vol. 32, no. 3, pp. 347–358, 1989.

[46] J. Hagemeister, B. Lowther, P. Oman, X. Yu, and W. Zhu, "An annotated bibliography on software maintenance," *ACM SIGSOFT Software Engineering Notes*, vol. 17, no. 2, pp. 79–84, 1992.

[47] R. D. Banker, S. M. Datar, C. F. Kemerer, and D. Zweig, "Software complexity and maintenance costs," *Communications of the ACM*, vol. 36, no. 11, pp. 81–95, 1993.

[48] L. H. Rosenberg and L. E. Hyatt, "Software quality metrics for object-oriented environments," *Crosstalk journal*, vol. 10, no. 4, pp. 1–6, 1997.

[49] H. M. Olague, L. H. Etzkorn, S. Gholston, and S. Quattlebaum, "Empirical validation of three software metrics suites to predict fault-proneness of object-oriented classes developed using highly iterative or agile software development processes," *IEEE Transactions on software Engineering*, vol. 33, no. 6, pp. 402–419, 2007.

[50] S. R. Chidamber, D. P. Darcy, and C. F. Kemerer, "Managerial use of metrics for object-oriented software: An exploratory analysis," *IEEE Transactions on software Engineering*, vol. 24, no. 8, pp. 629–639, 1998.

[51] L. C. Briand, J. Wüst, S. V. Ikonomovski, and H. Lounis, "Investigating quality factors in object-oriented designs: an industrial case study," in *Proceedings of the 21st international conference on Software engineering*, pp. 345–354, 1999.

[52] M.-H. Tang, M.-H. Kao, and M.-H. Chen, "An empirical study on object-oriented metrics," in *Proceedings sixth international software metrics symposium (Cat. No. PR00403)*, pp. 242–249, IEEE, 1999.

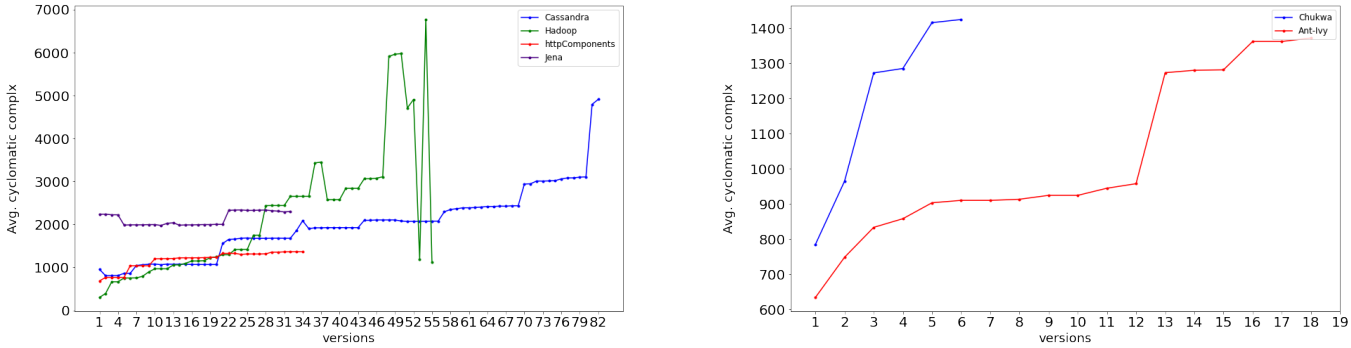
[53] P. R. M. Meirelles, *Monitoramento de métricas de código-fonte em projetos de software livre*. PhD thesis, Universidade de São Paulo, 2013.

[54] A. Terceiro, J. Costa, J. Miranda, P. Meirelles, L. R. Rios, L. Almeida, C. Chavez, and F. Kon, "Analizo: an extensible multi-language source code analysis and visualization toolkit," in *Brazilian conference on software: theory and practice (Tools Session)*, 2010.

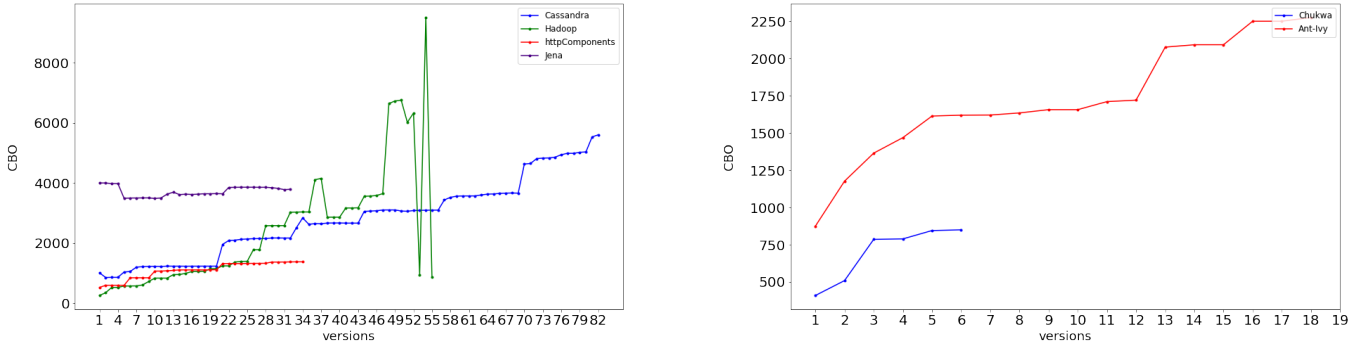
[55] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Reviews of modern physics*, vol. 74, no. 1, p. 47, 2002.

Appendix A Descriptive plots

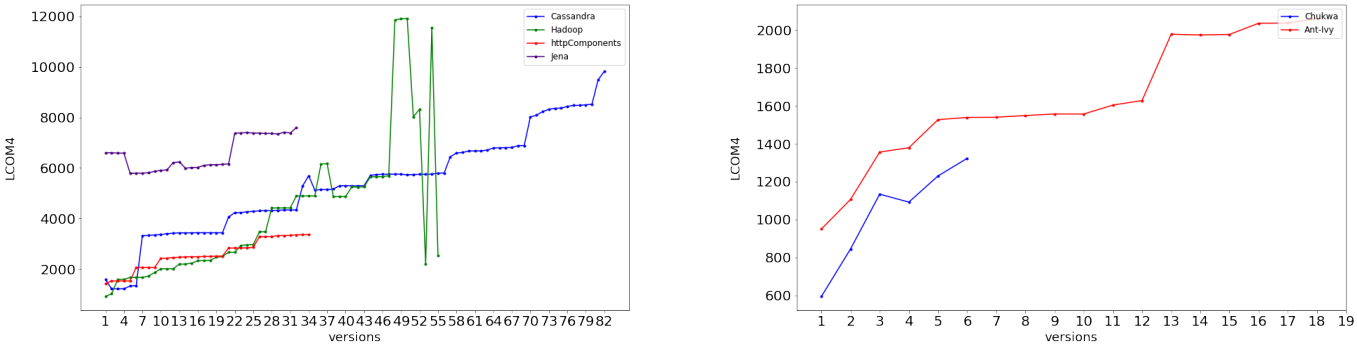
Software metrics changes



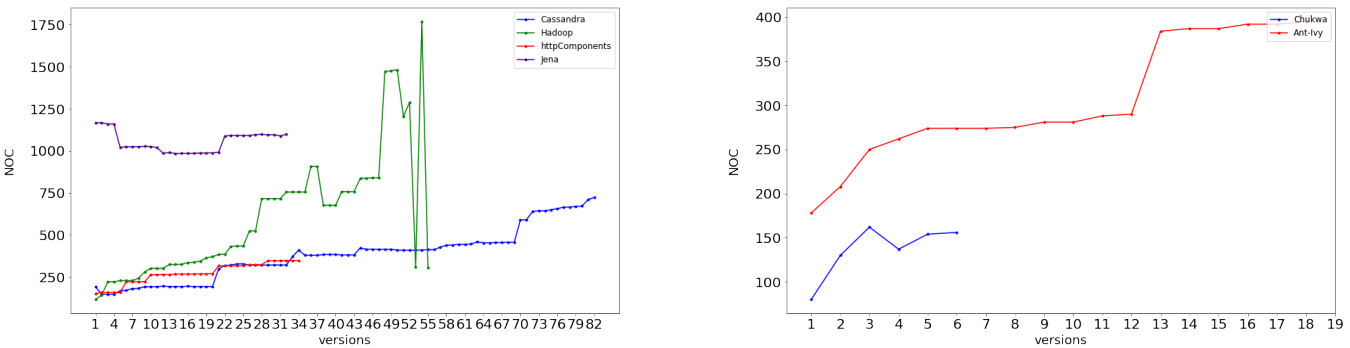
Average cyclomatic complexity



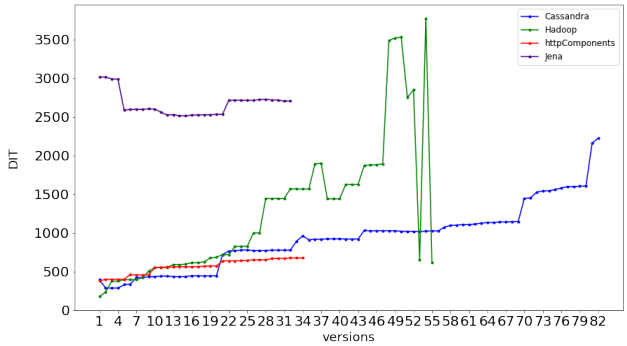
Coupling between objects

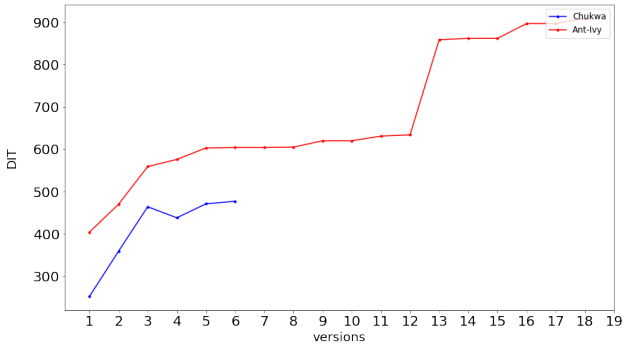


Lack of cohesion

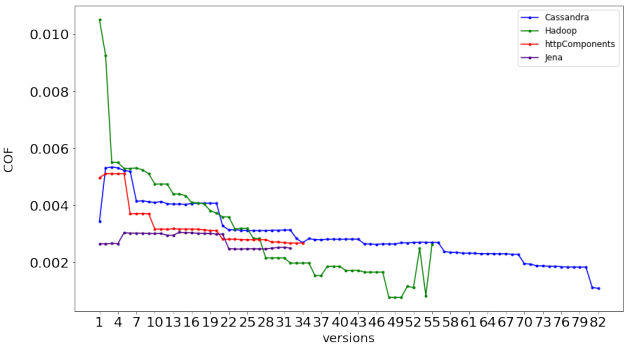


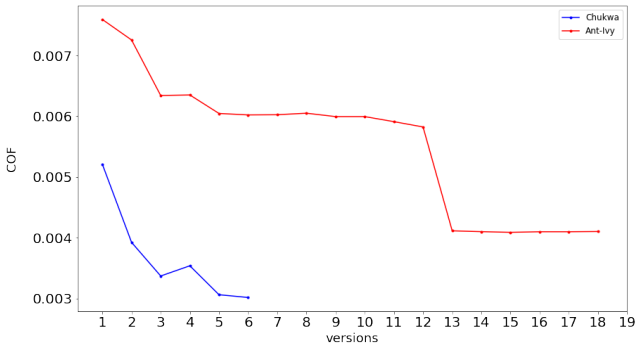
Number of Children



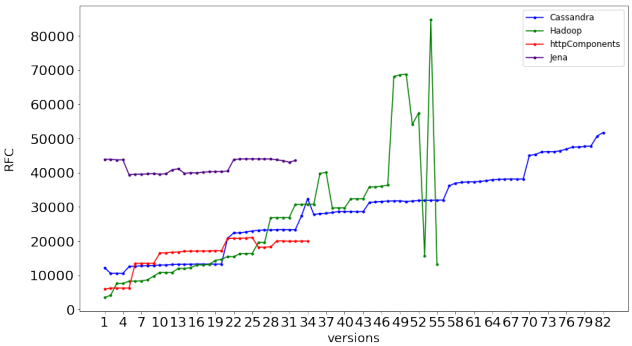


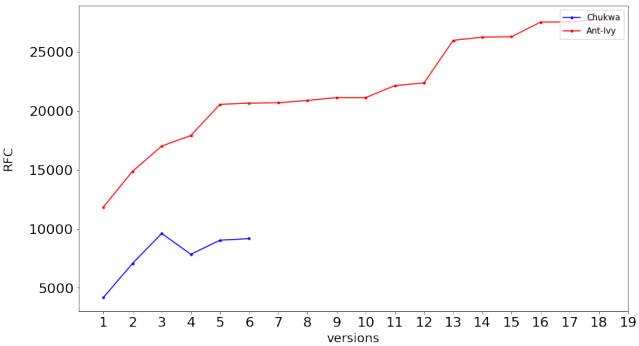
Depth of inheritance



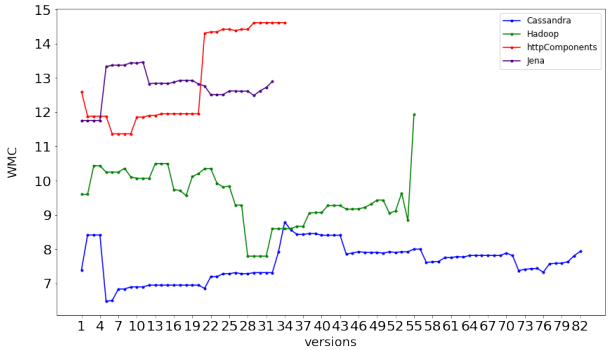


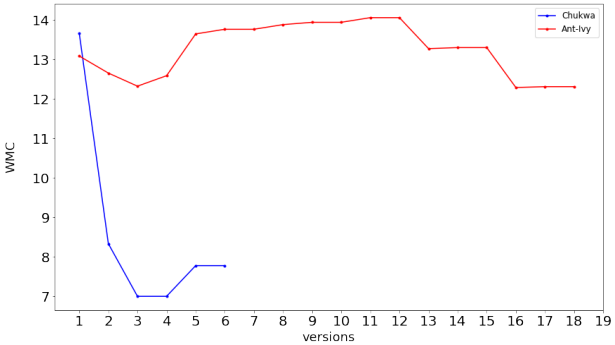
Coupling factor



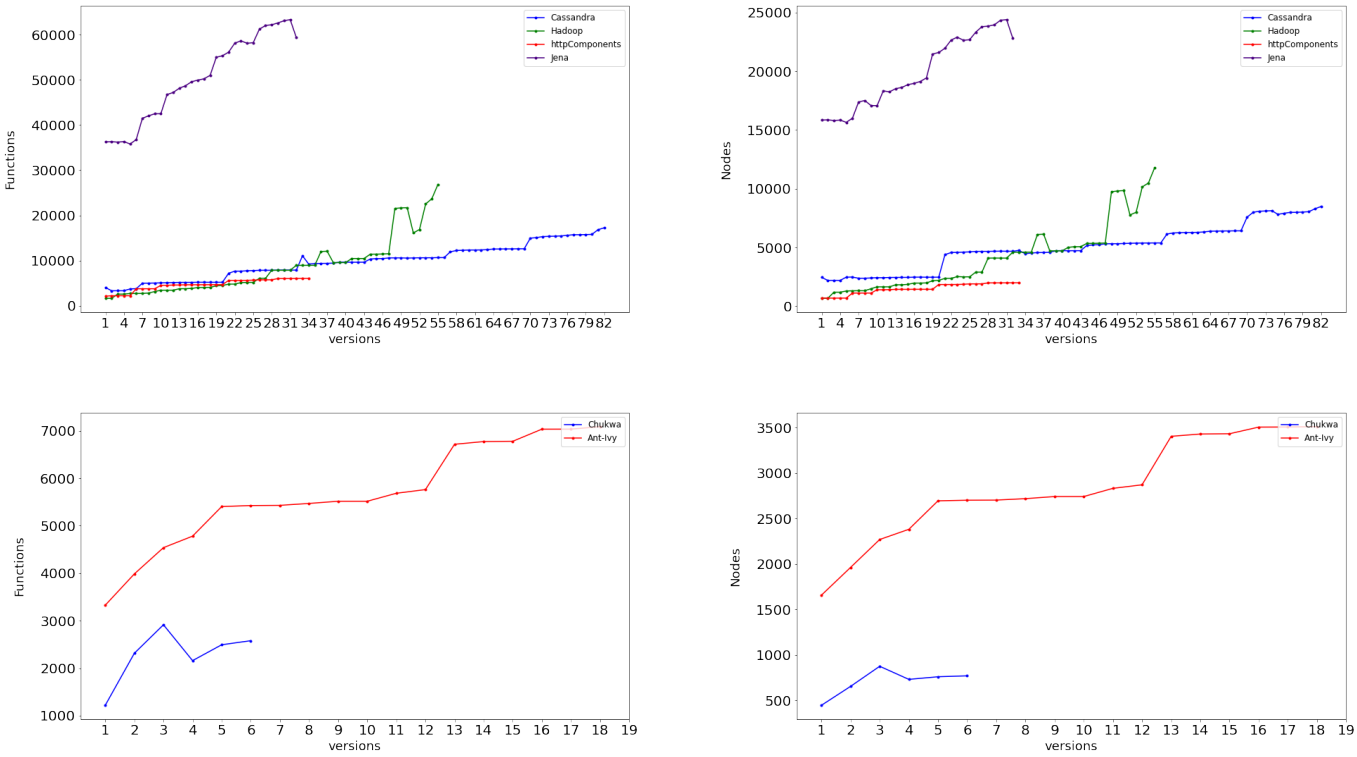


Response for a class

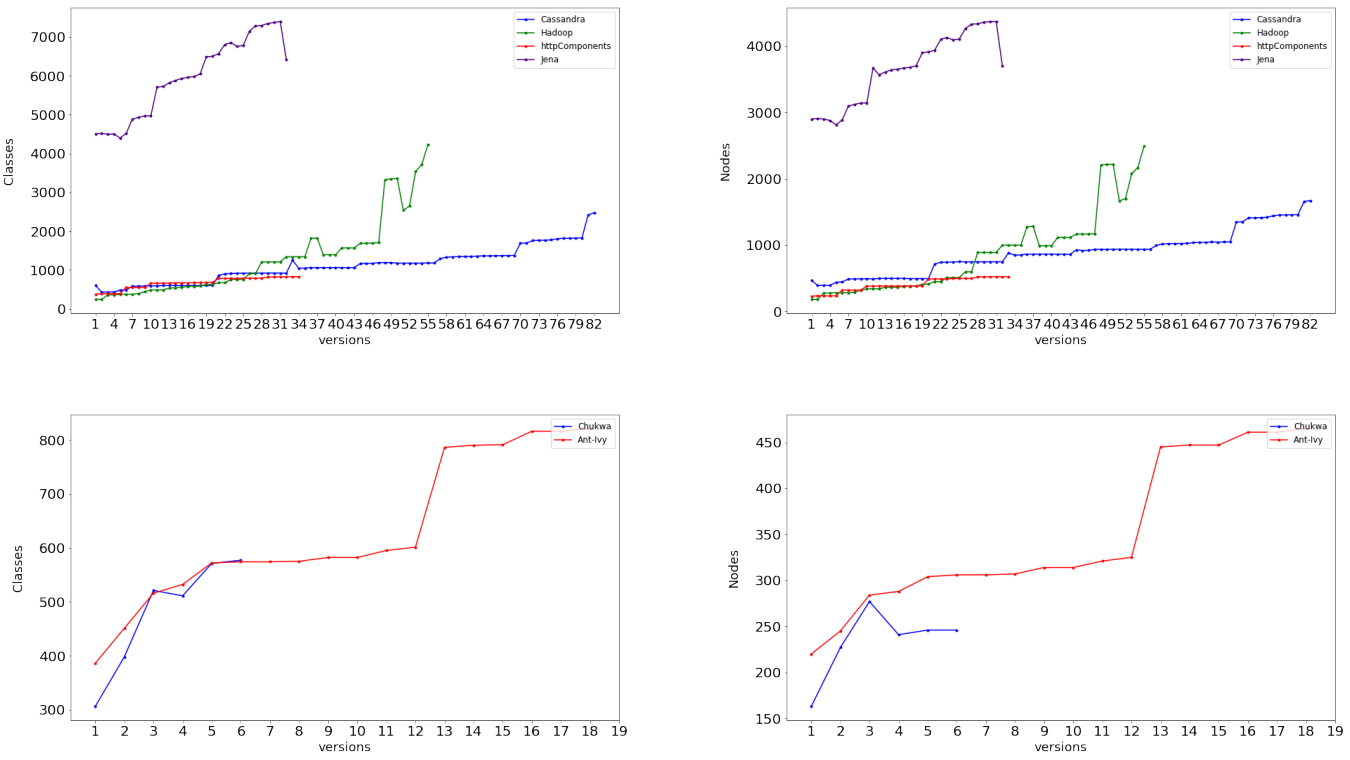




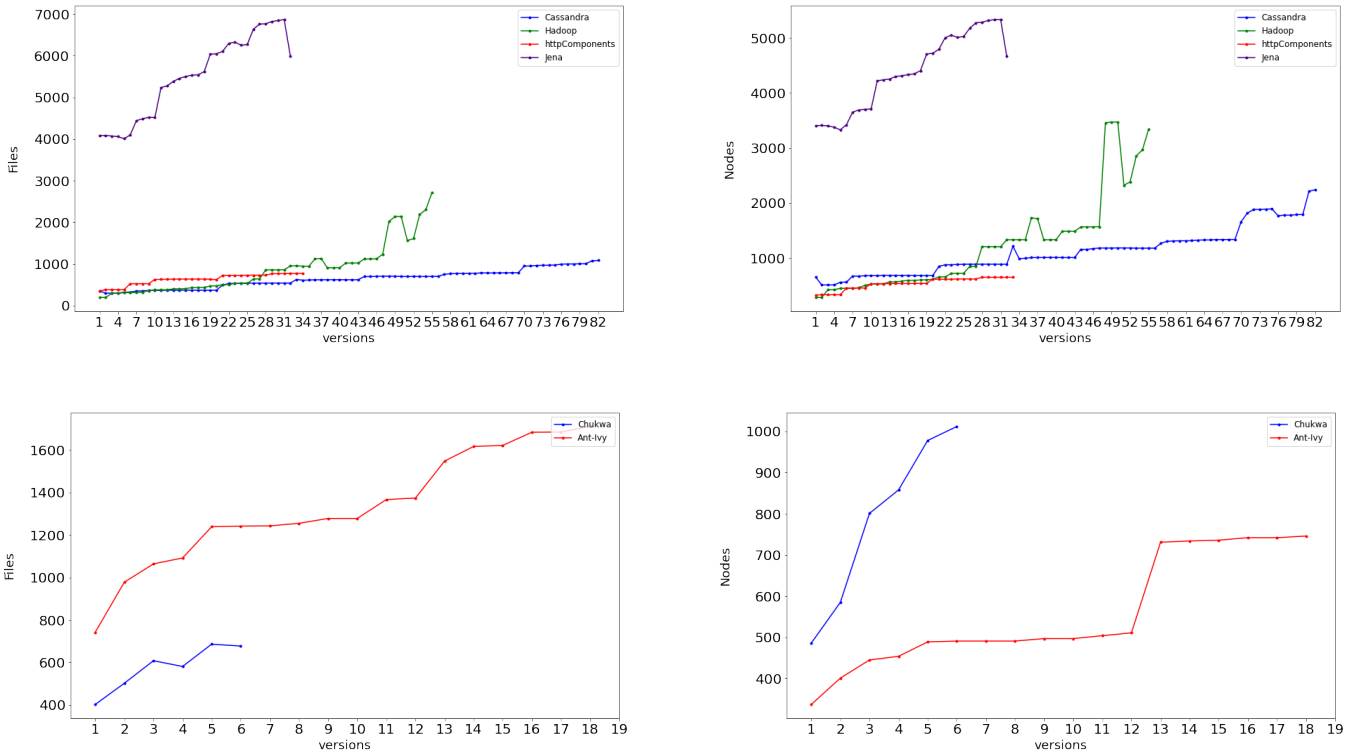
Weighted methods per class



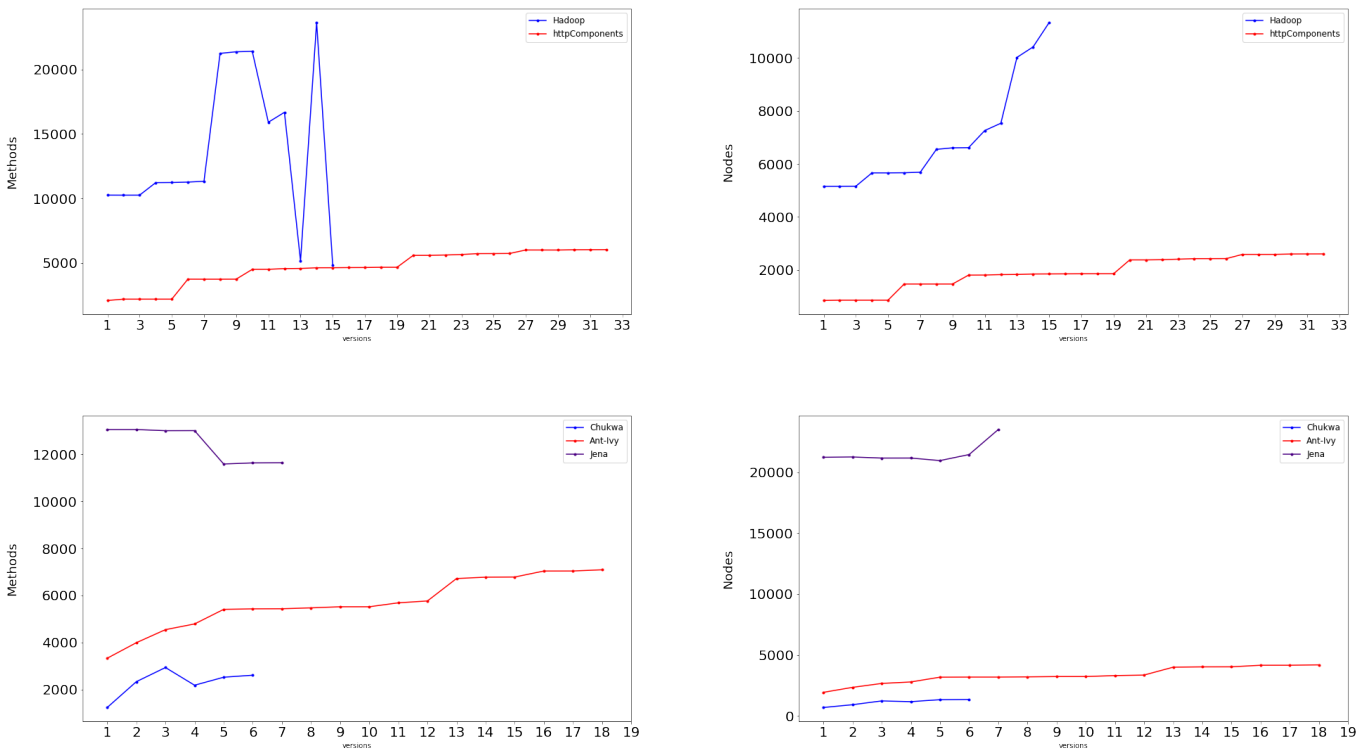
Functions vs. Nodes in call networks



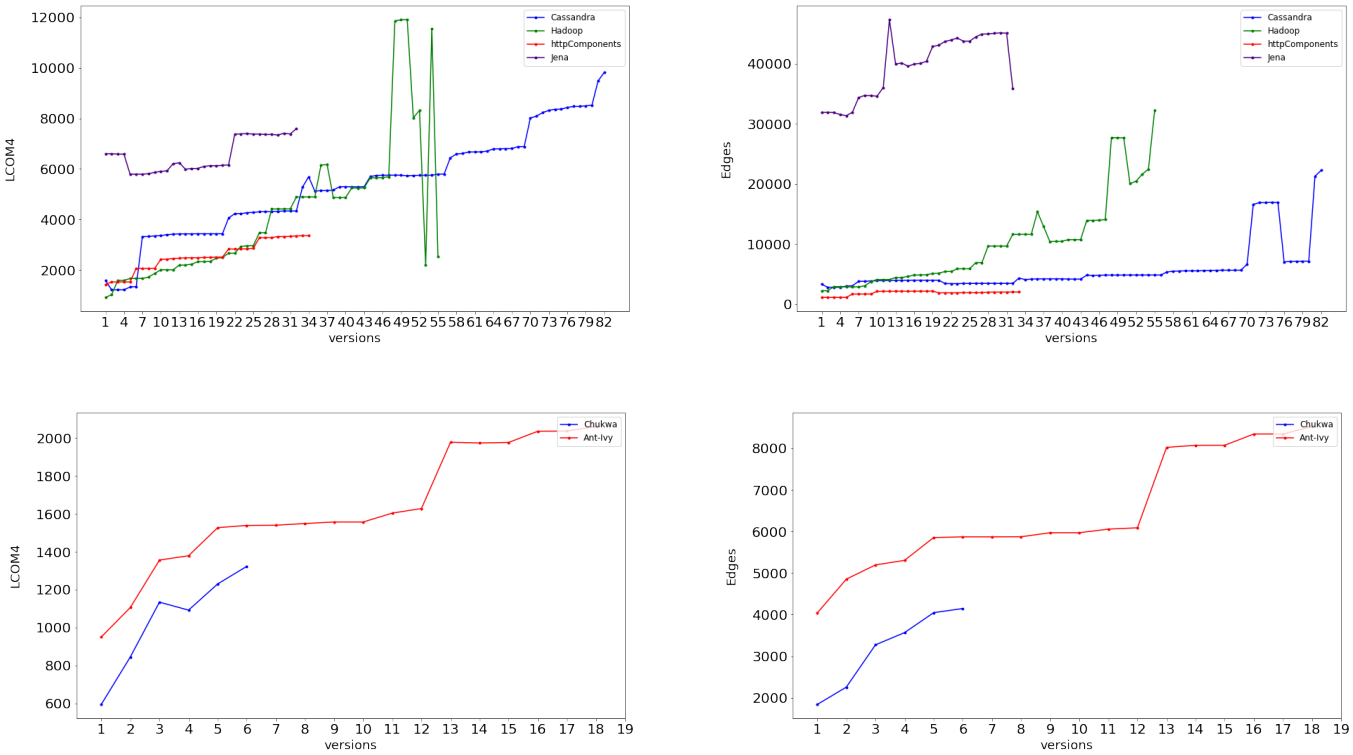
Classes vs. Nodes in inheritance networks



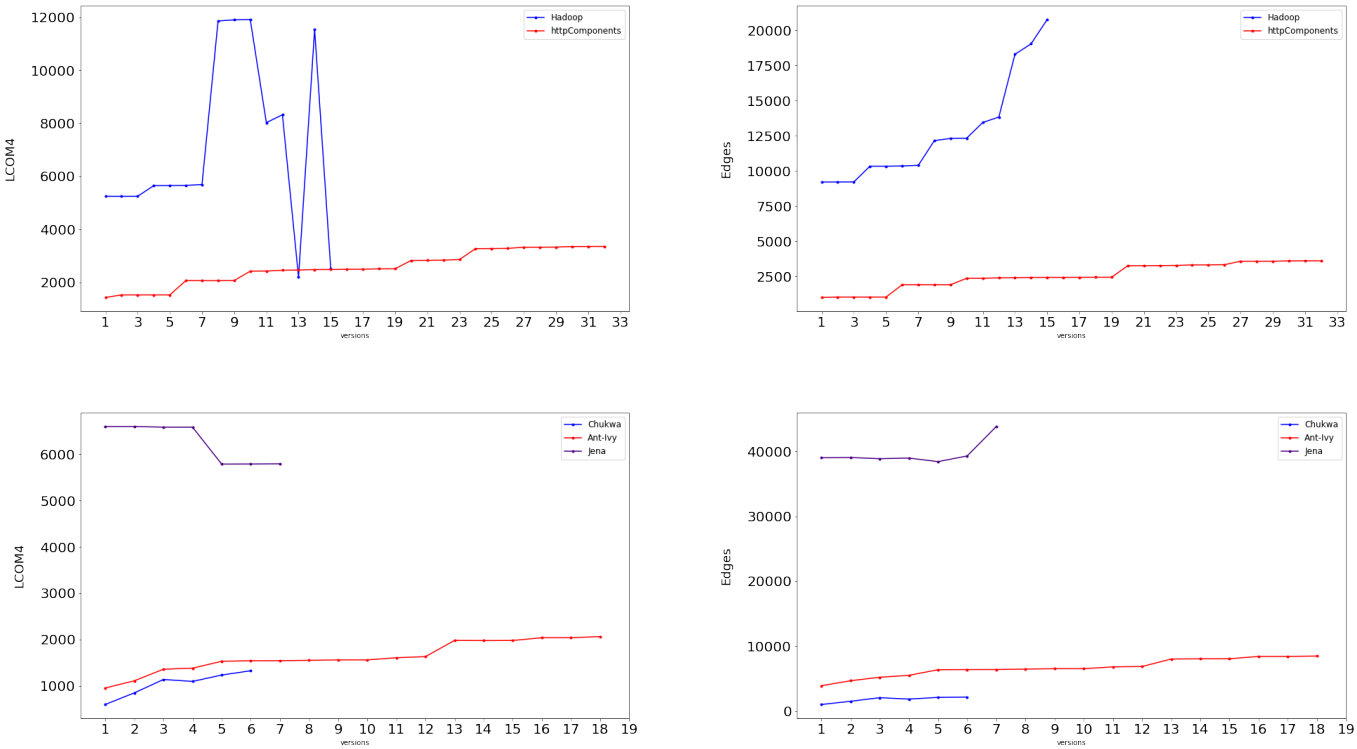
Files vs. Nodes in collaboration networks



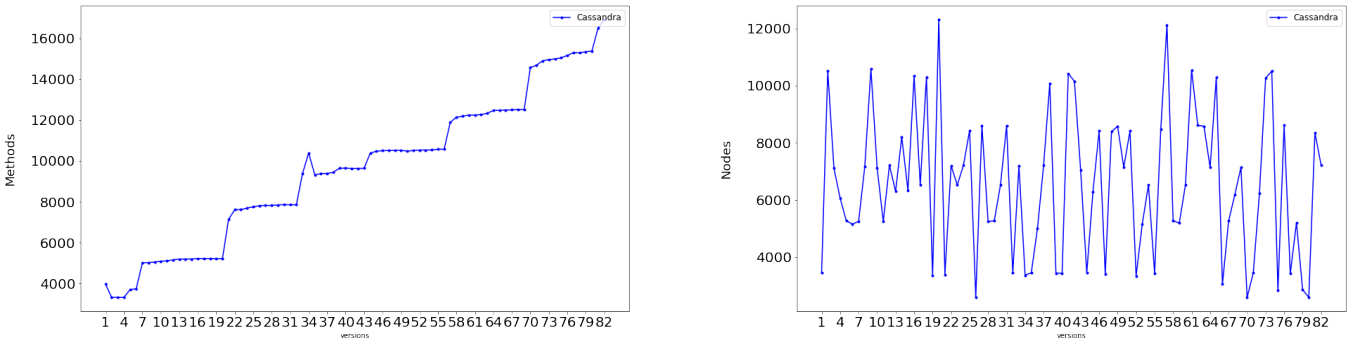
Methods vs. Nodes in methods networks



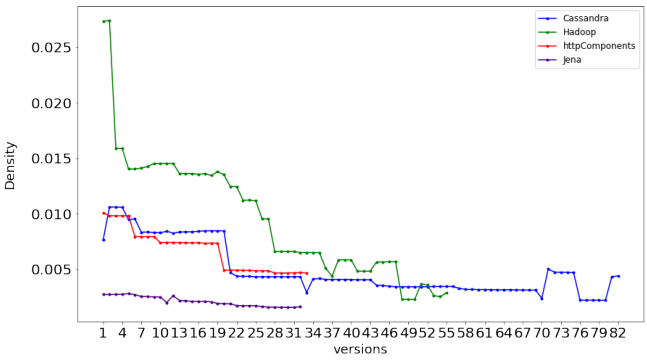
LCOM4 vs. Edges in collaboration networks

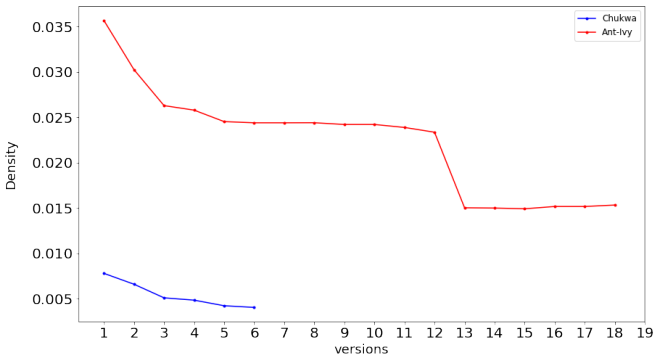


LCOM4 vs. Edges in methods networks

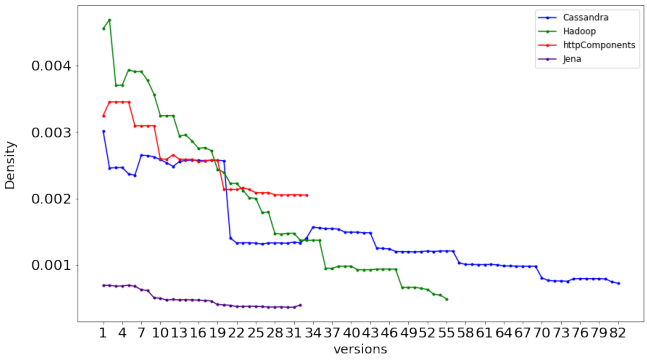


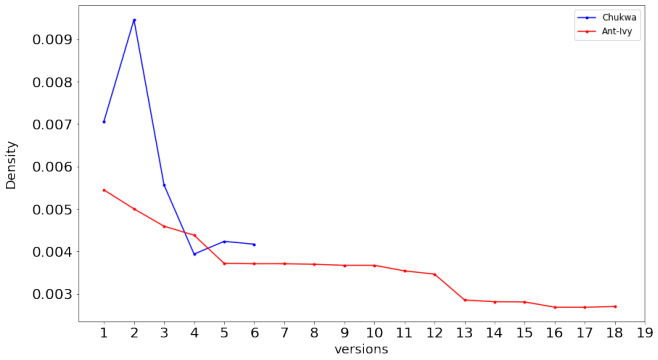
Cassandra's methods networks have been excluded as an outlier



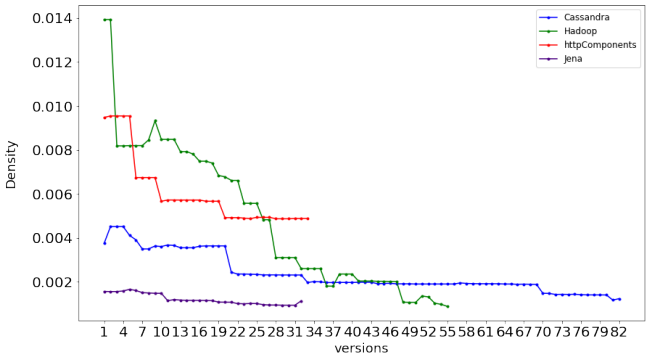


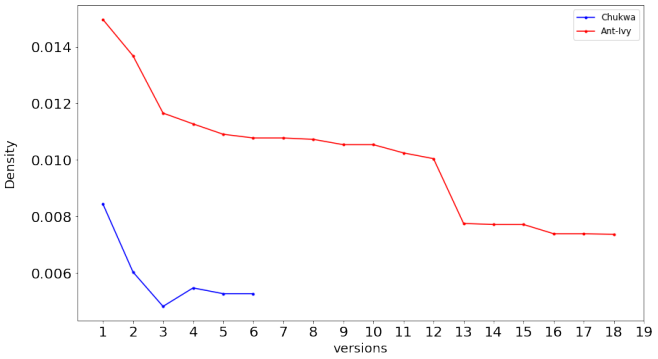
Density in collaboration networks



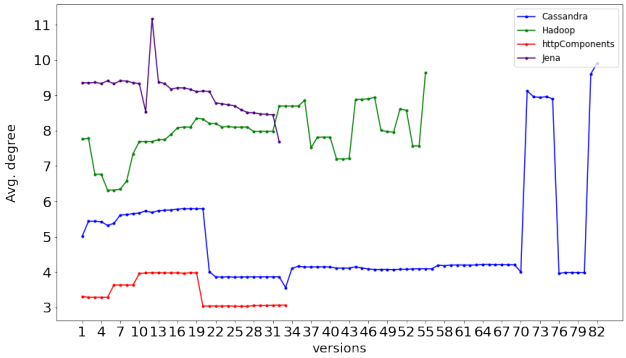


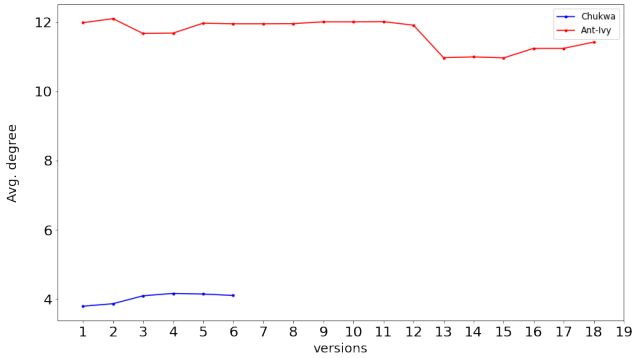
Density in call networks



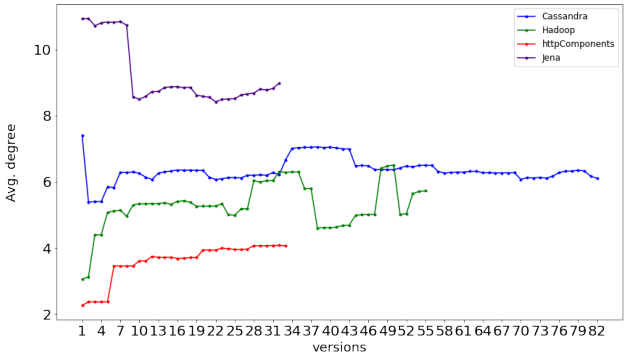


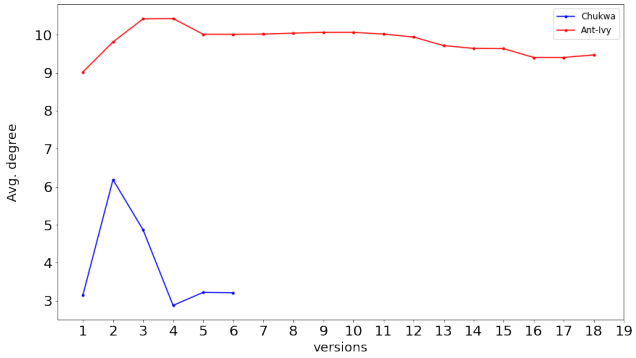
Density in inheritance networks



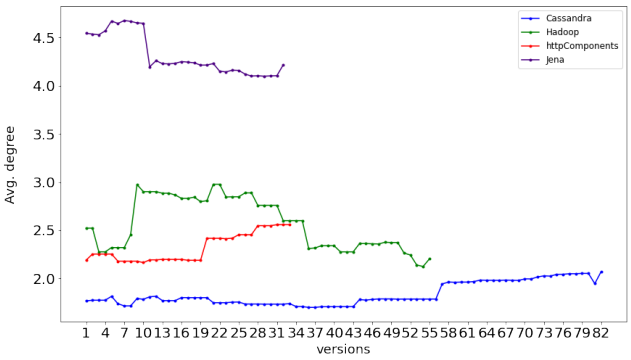


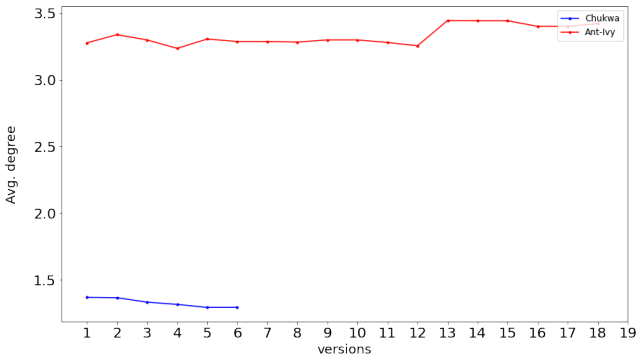
Average degree in collaboration networks



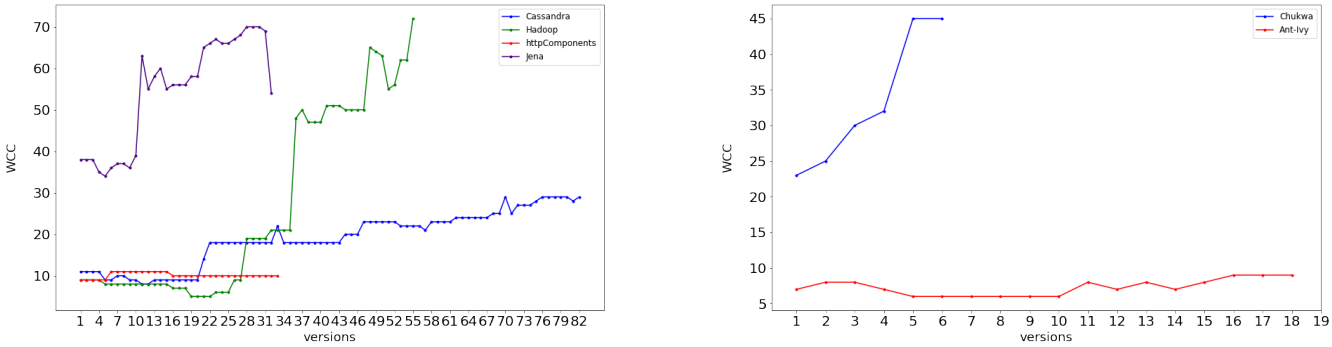


Average degree call networks

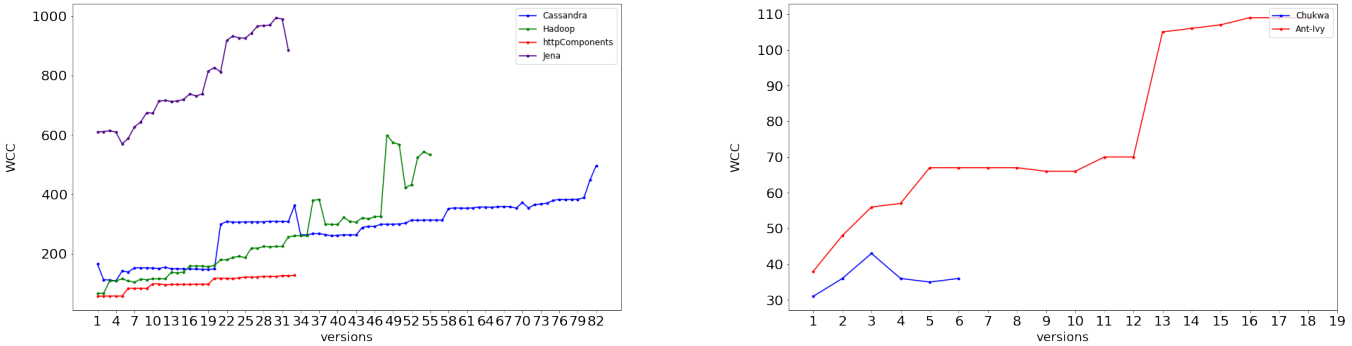




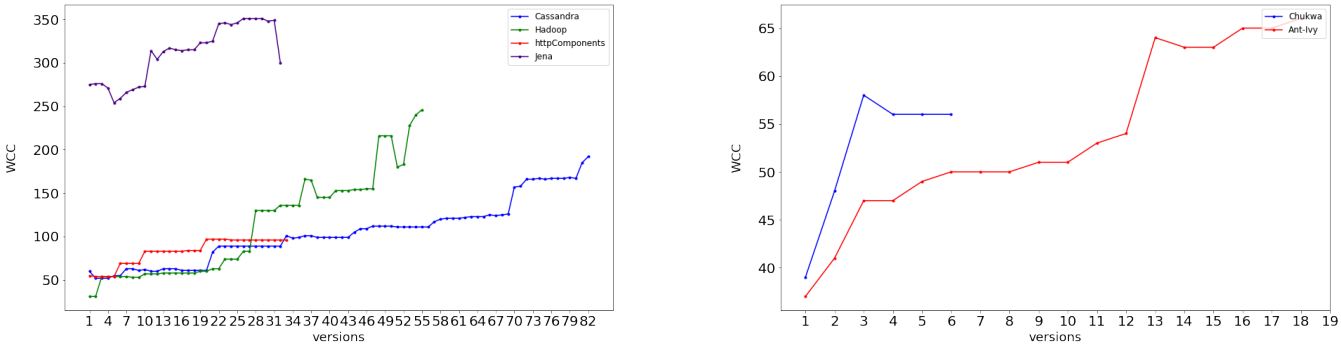
Average degree in inheritance networks



Weakly connected components in collaboration networks



Weakly connected components in call networks



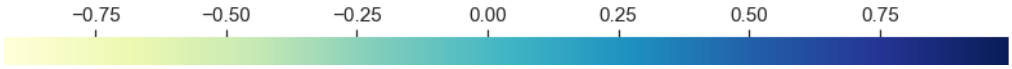
Weakly connected components in inheritance networks

Appendix B Heat maps

Heat maps using Pearson and Spearman rank correlation between software and network metrics for all software systems and all network granularities, the network metrics of importance are the diameter, Communities, Density , Average clustering coefficient and Weakly connected components. The 2-stars, 3-stars and 4-stars were used to show the importance of, but not usedf for depth analysis.

Pearson correlation

Call networks



	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	-0.36	-0.43	-0.39	-0.4	-0.41	-0.5	0.55	-0.56	-0.49
2-stars	0.87	0.91	0.9	0.88	0.92	0.89	-0.8	0.2	0.88
3-stars	0.65	0.69	0.68	0.66	0.71	0.66	-0.56	0.028	0.65
4-stars	0.52	0.57	0.55	0.54	0.58	0.54	-0.44	-0.028	0.52
γ	-0.52	-0.62	-0.57	-0.62	-0.62	-0.66	0.66	-0.69	-0.66
Nodes	0.94	0.99	0.96	0.96	0.98	0.99	-0.93	0.28	0.98
Edges	0.93	0.98	0.96	0.97	0.98	0.99	-0.96	0.36	0.99
Avg. degree	0.0079	0.015	0.033	0.077	0.045	0.051	-0.25	0.5	0.069
Communities	0.91	0.92	0.92	0.92	0.92	0.93	-0.95	0.29	0.94
Avg. shortest path	-0.63	-0.66	-0.67	-0.64	-0.68	-0.7	0.82	-0.55	-0.7
Density	-0.83	-0.88	-0.86	-0.87	-0.88	-0.91	0.89	-0.36	-0.91
Avg. clust. coeff.	0.32	0.31	0.35	0.22	0.38	0.33	-0.28	0.34	0.3
SCC	0.94	0.99	0.96	0.96	0.98	0.99	-0.93	0.28	0.98
WCC	0.89	0.89	0.9	0.89	0.89	0.91	-0.93	0.28	0.92
Lrgst WCC nodes	0.93	0.99	0.96	0.97	0.98	0.99	-0.93	0.31	0.98
Lrgst WCC edges	0.93	0.98	0.96	0.97	0.98	0.99	-0.96	0.37	0.99
γ_in	-0.54	-0.6	-0.56	-0.55	-0.59	-0.6	0.52	-0.11	-0.59
γ_out	-0.6	-0.64	-0.62	-0.61	-0.64	-0.66	0.63	-0.49	-0.65

Cassandra

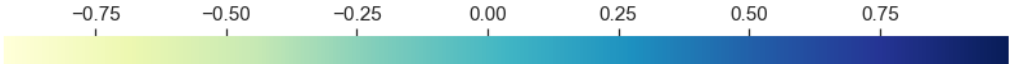
	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	-0.92	-0.93	-0.85	-0.88	-0.7	-0.72	0.81	0.72	-0.95
2-stars	0.94	0.91	0.96	0.97	0.96	0.95	-0.99	-0.86	0.9
3-stars	0.89	0.84	0.79	0.88	0.69	0.7	-0.83	-0.57	0.83
4-stars	0.82	0.76	0.69	0.8	0.57	0.59	-0.73	-0.43	0.75
γ	-0.59	-0.53	-0.43	-0.58	-0.32	-0.35	0.46	0.062	-0.5
Nodes	0.85	0.87	0.94	0.87	0.98	0.98	-0.91	-0.92	0.85
Edges	0.12	0.12	0.34	0.21	0.56	0.52	-0.38	-0.56	0.096
Avg. degree	-0.32	-0.34	-0.11	-0.22	0.13	0.072	0.025	-0.19	-0.36
Communities	0.17	0.19	0.39	0.24	0.58	0.54	-0.39	-0.63	0.18
Avg. shortest path	-0.89	-0.86	-0.82	-0.86	-0.68	-0.69	0.83	0.7	-0.89
Density	-0.8	-0.84	-0.68	-0.73	-0.49	-0.54	0.59	0.41	-0.85
Avg. clust. coeff.	0.75	0.73	0.58	0.71	0.41	0.46	-0.56	-0.23	0.72
SCC	0.85	0.86	0.94	0.87	0.98	0.98	-0.91	-0.92	0.85
WCC	0.47	0.52	0.64	0.52	0.76	0.76	-0.58	-0.73	0.5
Lrgst WCC nodes	0.87	0.88	0.95	0.89	0.98	0.99	-0.92	-0.92	0.87
Lrgst WCC edges	0.11	0.11	0.33	0.2	0.55	0.51	-0.37	-0.55	0.085
γ_in	-0.53	-0.53	-0.33	-0.44	-0.091	-0.14	0.27	0.034	-0.55
γ_out	-0.38	-0.35	-0.19	-0.3	0.035	0.016	0.2	0.027	-0.4

Chukwa

	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.68	0.66	0.69	0.68	0.67	0.68	-0.77	-0.21	0.73
2-stars	0.74	0.68	0.74	0.8	0.7	0.73	-0.47	-0.074	0.85
3-stars	0.63	0.56	0.64	0.71	0.59	0.61	-0.33	-0.03	0.67
4-stars	0.6	0.53	0.61	0.68	0.57	0.58	-0.3	-0.029	0.63
γ	-0.57	-0.56	-0.57	-0.55	-0.56	-0.57	0.62	0.17	-0.61
Nodes	0.8	0.79	0.8	0.8	0.79	0.81	-0.74	-0.27	0.99
Edges	0.8	0.78	0.8	0.81	0.79	0.81	-0.71	-0.27	0.99
Avg. degree	0.45	0.4	0.45	0.46	0.46	0.44	-0.63	-0.32	0.47
Communities	0.85	0.83	0.85	0.86	0.83	0.85	-0.78	-0.3	0.99
Avg. shortest path	0.45	0.45	0.44	0.43	0.43	0.46	-0.38	-0.059	0.48
Density	-0.77	-0.74	-0.77	-0.75	-0.78	-0.77	0.94	0.51	-0.8
Avg. clust. coeff.	0.42	0.37	0.43	0.43	0.43	0.4	-0.61	-0.54	0.25
SCC	0.8	0.79	0.8	0.8	0.79	0.81	-0.74	-0.27	0.99
WCC	0.85	0.82	0.84	0.85	0.83	0.85	-0.77	-0.28	0.99
Lrgst WCC nodes	0.78	0.76	0.77	0.77	0.76	0.78	-0.73	-0.27	0.98
Lrgst WCC edges	0.78	0.76	0.78	0.79	0.77	0.79	-0.69	-0.27	0.98
γ_in	-0.6	-0.59	-0.59	-0.59	-0.59	-0.61	0.57	0.33	-0.67
γ_out	-0.3	-0.29	-0.3	-0.29	-0.29	-0.3	0.32	0.14	-0.46

Hadoop

Call networks



	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.85	0.88	0.86	0.9	0.9	0.86	-0.88	0.75	0.88
2-stars	0.92	0.98	0.99	0.97	0.98	0.91	-0.9	0.86	0.94
3-stars	0.9	0.96	0.98	0.96	0.96	0.88	-0.86	0.88	0.91
4-stars	0.9	0.96	0.98	0.94	0.94	0.88	-0.86	0.82	0.9
γ	-0.81	-0.73	-0.67	-0.67	-0.72	-0.83	0.82	-0.27	-0.79
Nodes	0.96	0.99	0.99	0.98	1	0.95	-0.95	0.79	0.98
Edges	0.96	0.99	0.99	0.98	1	0.95	-0.95	0.79	0.98
Avg. degree	0.98	0.95	0.91	0.91	0.95	0.98	-0.99	0.54	0.98
Communities	0.96	0.99	0.98	0.99	1	0.95	-0.95	0.78	0.98
Avg. shortest path	0.8	0.73	0.69	0.74	0.75	0.79	-0.86	0.26	0.82
Density	-0.95	-0.99	-0.99	-0.97	-0.98	-0.94	0.93	-0.8	-0.96
Avg. clust. coeff.	0.88	0.8	0.74	0.72	0.78	0.9	-0.91	0.25	0.86
SCC	0.96	0.99	0.99	0.98	1	0.95	-0.95	0.79	0.98
WCC	0.97	0.99	0.98	0.99	1	0.96	-0.96	0.77	0.98
Lrgst WCC nodes	0.94	0.98	0.99	0.99	0.99	0.92	-0.92	0.83	0.96
Lrgst WCC edges	0.94	0.98	0.99	0.99	0.99	0.93	-0.92	0.82	0.96
γ_in	0.58	0.7	0.75	0.75	0.71	0.56	-0.55	0.83	0.62
γ_out	-0.48	-0.37	-0.31	-0.29	-0.35	-0.51	0.55	0.093	-0.45

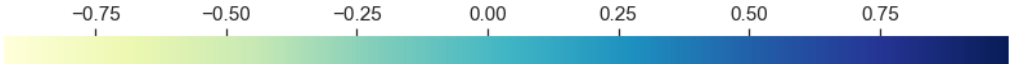
HttpComponents

	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.83	0.89	0.84	0.9	0.86	0.92	-0.86	0.19	0.87
2-stars	0.96	0.99	0.96	0.99	0.97	0.99	-0.96	-0.13	0.99
3-stars	0.97	0.97	0.96	0.96	0.95	0.96	-0.94	-0.25	0.99
4-stars	0.95	0.94	0.94	0.92	0.93	0.93	-0.91	-0.35	0.97
γ	0.43	0.3	0.42	0.3	0.4	0.24	-0.42	-0.77	0.33
Nodes	0.96	0.99	0.97	1	0.98	1	-0.98	-0.11	0.98
Edges	0.92	0.98	0.93	0.98	0.95	0.98	-0.95	-0.033	0.93
Avg. degree	-0.41	-0.26	-0.37	-0.27	-0.33	-0.26	0.34	0.31	-0.42
Communities	0.99	0.98	1	0.98	0.99	0.97	-0.99	-0.3	0.98
Avg. shortest path	0.96	0.94	0.97	0.96	0.98	0.94	-0.98	-0.21	0.94
Density	-0.93	-0.98	-0.94	-0.99	-0.95	-0.99	0.95	0.0083	-0.95
Avg. clust. coeff.	0.36	0.21	0.32	0.2	0.27	0.19	-0.26	-0.47	0.37
SCC	0.97	0.99	0.97	1	0.98	1	-0.98	-0.12	0.98
WCC	1	0.97	1	0.98	0.99	0.96	-0.99	-0.29	0.99
Lrgst WCC nodes	0.96	0.99	0.96	1	0.97	1	-0.97	-0.084	0.97
Lrgst WCC edges	0.92	0.98	0.93	0.98	0.95	0.98	-0.94	-0.031	0.93
γ_in	-0.63	-0.76	-0.64	-0.75	-0.66	-0.79	0.65	-0.3	-0.68
γ_out	-0.58	-0.55	-0.59	-0.61	-0.6	-0.6	0.63	-0.22	-0.6

Ant-Ivy

	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.099	0.18	0.6	-0.073	0.52	0.14	-0.084	-0.25	-0.51
2-stars	0.52	0.24	-0.21	0.71	-0.0019	0.41	-0.54	0.0063	0.98
3-stars	0.46	0.16	-0.15	0.65	0.053	0.35	-0.49	0.098	0.86
4-stars	0.29	0.072	-0.05	0.43	0.09	0.22	-0.32	0.12	0.52
γ	0.56	0.24	-0.038	0.67	0.16	0.46	-0.59	0.03	0.77
Nodes	0.6	0.31	-0.12	0.76	0.089	0.49	-0.62	-0.053	0.98
Edges	0.71	0.47	0.22	0.78	0.4	0.62	-0.72	-0.24	0.78
Avg. degree	-0.12	0.081	0.5	-0.33	0.36	-0.048	0.15	-0.24	-0.72
Communities	0.69	0.41	-0.0038	0.83	0.21	0.59	-0.71	-0.15	0.98
Avg. shortest path	0.49	0.49	0.72	0.3	0.71	0.51	-0.47	-0.46	-0.16
Density	-0.35	-0.089	0.37	-0.56	0.19	-0.26	0.38	-0.13	-0.92
Avg. clust. coeff.	-0.22	-0.035	0.45	-0.43	0.31	-0.16	0.25	-0.13	-0.81
SCC	0.59	0.3	-0.13	0.76	0.076	0.48	-0.61	-0.045	0.98
WCC	0.67	0.38	-0.038	0.82	0.18	0.57	-0.69	-0.11	0.98
Lrgst WCC nodes	0.59	0.3	-0.12	0.76	0.088	0.49	-0.61	-0.053	0.98
Lrgst WCC edges	0.71	0.47	0.25	0.76	0.42	0.62	-0.71	-0.26	0.74
γ_in	-0.00048	0.15	0.3	-0.14	0.22	0.062	0.022	-0.2	-0.39
γ_out	-0.57	-0.29	-0.12	-0.63	-0.29	-0.46	0.58	0.072	-0.68

Jena



	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	-0.041	-0.087	-0.069	-0.22	-0.088	-0.11	0.33	0.13	-0.16
2-stars	0.81	0.9	0.85	0.88	0.88	0.9	-0.83	0.29	0.9
3-stars	0.62	0.74	0.67	0.71	0.72	0.74	-0.65	0.21	0.73
4-stars	0.53	0.65	0.58	0.62	0.65	0.65	-0.54	0.13	0.64
γ	0.17	0.22	0.21	0.19	0.23	0.28	-0.33	0.49	0.3
Nodes	0.97	0.98	0.98	0.97	0.98	0.97	-0.9	0.24	0.96
Edges	0.79	0.73	0.77	0.7	0.73	0.68	-0.61	0.043	0.66
Avg. degree	0.37	0.28	0.34	0.27	0.29	0.22	-0.13	-0.2	0.2
Communities	0.93	0.93	0.94	0.9	0.93	0.89	-0.81	0.17	0.88
Avg. shortest path	0.058	0.018	0.038	-0.042	0.022	-0.046	0.22	-0.32	-0.046
Density	-0.67	-0.73	-0.71	-0.76	-0.73	-0.78	0.86	-0.45	-0.8
Avg. clust. coeff.	0.41	0.52	0.44	0.49	0.48	0.58	-0.6	0.44	0.56
SCC	0.97	0.99	0.98	0.97	0.98	0.97	-0.9	0.24	0.96
WCC	0.9	0.95	0.93	0.93	0.95	0.97	-0.92	0.39	0.97
Lrgst WCC nodes	0.95	0.99	0.97	0.97	0.98	0.97	-0.9	0.23	0.97
Lrgst WCC edges	0.79	0.72	0.77	0.7	0.73	0.68	-0.61	0.041	0.66
γ_in	0.34	0.23	0.29	0.17	0.23	0.18	-0.075	-0.068	0.15
γ_out	-0.3	-0.39	-0.33	-0.4	-0.37	-0.38	0.29	0.22	-0.37

Cassandra

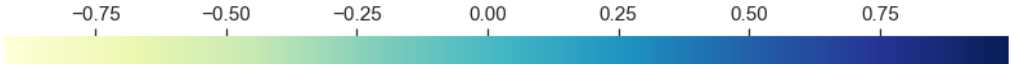
	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.88	0.88	0.86	0.88	0.83	0.86	-0.81	-0.59	0.84
2-stars	0.99	0.99	0.98	0.99	0.92	0.94	-0.96	-0.82	0.98
3-stars	0.98	0.98	0.98	0.98	0.94	0.96	-0.95	-0.83	0.97
4-stars	0.95	0.96	0.97	0.96	0.94	0.96	-0.93	-0.81	0.94
γ	0.16	0.085	-0.044	0.14	-0.19	-0.17	-0.021	0.32	0.083
Nodes	0.99	0.97	0.93	0.98	0.83	0.85	-0.92	-0.73	0.97
Edges	0.99	0.98	0.93	0.97	0.83	0.85	-0.91	-0.73	0.98
Avg. degree	0.95	0.97	0.93	0.91	0.82	0.85	-0.87	-0.82	0.99
Communities	0.94	0.91	0.85	0.92	0.74	0.76	-0.86	-0.61	0.9
Avg. shortest path	0.85	0.83	0.88	0.91	0.85	0.84	-0.9	-0.83	0.84
Density	-1	-0.99	-0.97	-0.99	-0.89	-0.91	0.96	0.82	-0.98
Avg. clust. coeff.	-0.52	-0.49	-0.31	-0.47	-0.12	-0.17	0.29	-0.069	-0.48
SCC	0.99	0.97	0.93	0.98	0.83	0.85	-0.92	-0.73	0.97
WCC	0.88	0.84	0.78	0.88	0.68	0.7	-0.8	-0.51	0.82
Lrgst WCC nodes	1	0.99	0.98	1	0.92	0.94	-0.96	-0.82	0.98
Lrgst WCC edges	1	0.99	0.96	0.98	0.86	0.89	-0.93	-0.77	0.99
γ_in	-0.34	-0.39	-0.17	-0.26	0.035	-0.038	0.039	-0.15	-0.39
γ_out	-0.4	-0.4	-0.34	-0.47	-0.28	-0.31	0.31	0.11	-0.38

Chukwa

	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.68	0.66	0.69	0.68	0.67	0.68	-0.77	-0.21	0.73
2-stars	0.74	0.68	0.74	0.8	0.7	0.73	-0.47	-0.074	0.85
3-stars	0.63	0.56	0.64	0.71	0.59	0.61	-0.33	-0.03	0.67
4-stars	0.6	0.53	0.61	0.68	0.57	0.58	-0.3	-0.029	0.63
γ	-0.57	-0.56	-0.57	-0.55	-0.56	-0.57	0.62	0.17	-0.61
Nodes	0.83	0.8	0.83	0.84	0.81	0.83	-0.71	-0.25	1
Edges	0.8	0.77	0.8	0.81	0.78	0.8	-0.7	-0.22	0.99
Avg. degree	0.3	0.3	0.32	0.29	0.33	0.3	-0.44	-0.24	0.37
Communities	0.83	0.81	0.84	0.84	0.81	0.83	-0.76	-0.32	0.95
Avg. shortest path	0.46	0.43	0.48	0.44	0.48	0.44	-0.75	-0.58	0.28
Density	-0.77	-0.74	-0.77	-0.75	-0.77	-0.76	0.97	0.46	-0.8
Avg. clust. coeff.	0.092	0.082	0.076	0.12	0.069	0.1	0.052	0.34	0.28
SCC	0.83	0.8	0.83	0.84	0.81	0.83	-0.71	-0.25	1
WCC	0.77	0.76	0.77	0.77	0.75	0.77	-0.68	-0.27	0.89
Lrgst WCC nodes	0.83	0.8	0.83	0.84	0.81	0.83	-0.72	-0.27	1
Lrgst WCC edges	0.8	0.77	0.8	0.81	0.78	0.8	-0.7	-0.22	0.99
γ_in	-0.67	-0.67	-0.66	-0.64	-0.67	-0.67	0.78	0.44	-0.57
γ_out	-0.58	-0.56	-0.57	-0.59	-0.55	-0.57	0.44	0.1	-0.58

Hadoop

Collaboration networks



	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	-0.53	-0.66	-0.7	-0.72	-0.69	-0.52	0.47	-0.96	-0.58
2-stars	0.74	0.84	0.86	0.88	0.87	0.73	-0.7	0.98	0.78
3-stars	0.68	0.79	0.82	0.85	0.83	0.67	-0.64	0.98	0.73
4-stars	0.67	0.78	0.81	0.85	0.82	0.66	-0.63	0.98	0.73
γ	-0.036	0.14	0.2	0.2	0.16	-0.025	0.098	0.71	0.013
Nodes	0.99	1	0.99	0.98	1	0.97	-0.97	0.76	0.99
Edges	0.88	0.79	0.75	0.71	0.75	0.86	-0.89	0.23	0.84
Avg. degree	-0.15	-0.31	-0.35	-0.4	-0.36	-0.16	0.098	-0.82	-0.22
Communities	0.91	0.91	0.93	0.89	0.9	0.86	-0.87	0.65	0.9
Avg. shortest path	-0.81	-0.9	-0.91	-0.92	-0.91	-0.81	0.78	-0.95	-0.85
Density	-0.92	-0.97	-0.97	-0.97	-0.98	-0.92	0.9	-0.87	-0.95
Avg. clust. coeff.	0.75	0.85	0.86	0.87	0.86	0.76	-0.73	0.95	0.79
SCC	0.98	1	0.99	0.98	1	0.97	-0.97	0.76	0.99
WCC	0.41	0.27	0.19	0.2	0.26	0.43	-0.49	-0.28	0.38
Lrgst WCC nodes	0.98	1	0.99	0.98	1	0.97	-0.97	0.77	0.99
Lrgst WCC edges	0.88	0.79	0.75	0.71	0.76	0.86	-0.89	0.23	0.84
γ_in	-0.16	-0.0035	0.042	0.13	0.059	-0.17	0.2	0.57	-0.081
γ_out	0.72	0.77	0.78	0.69	0.71	0.73	-0.7	0.66	0.69

HttpComponents

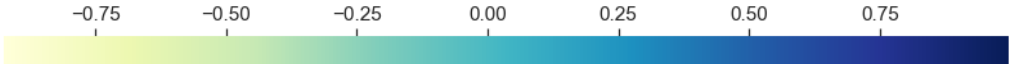
	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.73	0.55	0.71	0.57	0.68	0.51	-0.69	-0.75	0.65
2-stars	0.99	0.93	0.99	0.94	0.98	0.92	-0.98	-0.34	0.97
3-stars	0.97	0.89	0.97	0.91	0.96	0.88	-0.96	-0.4	0.95
4-stars	0.96	0.87	0.95	0.89	0.94	0.86	-0.95	-0.43	0.93
γ	0.72	0.55	0.71	0.58	0.69	0.53	-0.7	-0.54	0.66
Nodes	1	0.96	1	0.97	0.99	0.95	-0.99	-0.32	0.98
Edges	1	0.98	1	0.98	0.99	0.97	-0.99	-0.29	0.99
Avg. degree	-0.85	-0.75	-0.86	-0.78	-0.86	-0.73	0.87	0.46	-0.79
Communities	0.94	0.97	0.95	0.98	0.96	0.97	-0.95	-0.16	0.94
Avg. shortest path	-0.72	-0.69	-0.73	-0.73	-0.74	-0.73	0.75	-0.15	-0.76
Density	-0.98	-0.98	-0.99	-0.99	-0.99	-0.98	0.99	0.22	-0.97
Avg. clust. coeff.	0.89	0.77	0.88	0.78	0.86	0.74	-0.87	-0.5	0.84
SCC	1	0.96	1	0.97	0.99	0.95	-0.99	-0.33	0.98
WCC	0.6	0.48	0.57	0.46	0.53	0.44	-0.53	-0.76	0.57
Lrgst WCC nodes	1	0.97	1	0.98	1	0.97	-1	-0.28	0.99
Lrgst WCC edges	1	0.98	1	0.98	0.99	0.97	-0.99	-0.28	0.99
γ_in	-0.17	-0.046	-0.15	-0.061	-0.13	-0.046	0.14	0.27	-0.15
γ_out	-0.72	-0.77	-0.72	-0.75	-0.74	-0.76	0.73	0.21	-0.69

Ant-Ivy

	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	-0.66	-0.41	-0.066	-0.78	-0.25	-0.58	0.68	0.17	-0.87
2-stars	0.63	0.25	0.085	0.7	0.31	0.49	-0.65	0.038	0.81
3-stars	0.49	0.12	0.21	0.48	0.39	0.36	-0.51	0.13	0.48
4-stars	0.37	0.036	0.21	0.34	0.37	0.25	-0.4	0.18	0.32
γ	0.54	0.3	0.014	0.66	0.2	0.45	-0.56	-0.063	0.81
Nodes	0.54	0.27	-0.19	0.71	0.0039	0.44	-0.56	-0.041	0.97
Edges	0.35	0.16	-0.33	0.52	-0.19	0.29	-0.37	-0.013	0.83
Avg. degree	-0.57	-0.31	-0.17	-0.65	-0.35	-0.47	0.58	0.057	-0.68
Communities	0.46	0.26	-0.23	0.63	-0.055	0.39	-0.48	-0.064	0.89
Avg. shortest path	0.099	0.098	-0.45	0.3	-0.38	0.095	-0.11	-0.056	0.62
Density	-0.52	-0.25	0.19	-0.7	-0.016	-0.42	0.54	-0.0032	-0.97
Avg. clust. coeff.	-0.35	-0.26	0.31	-0.55	0.18	-0.32	0.36	0.12	-0.84
SCC	0.54	0.27	-0.19	0.71	0.0039	0.44	-0.56	-0.041	0.97
WCC	0.45	0.26	-0.23	0.63	-0.063	0.39	-0.47	-0.068	0.9
Lrgst WCC nodes	0.55	0.28	-0.18	0.72	0.018	0.45	-0.57	-0.042	0.97
Lrgst WCC edges	0.35	0.16	-0.33	0.52	-0.19	0.28	-0.36	-0.013	0.83
γ_in	-0.053	0.091	0.025	-0.052	-0.027	0.0086	0.054	-0.19	-0.061
γ_out	-0.38	-0.37	-0.22	-0.41	-0.27	-0.39	0.39	0.28	-0.34

Jena

Inheritance networks



	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.63	0.67	0.68	0.69	0.72	0.69	-0.78	0.33	0.71
2-stars	0.98	0.98	0.99	0.98	0.98	0.98	-0.92	0.31	0.97
3-stars	0.97	0.98	0.98	0.96	0.97	0.97	-0.9	0.28	0.96
4-stars	0.96	0.97	0.96	0.94	0.96	0.95	-0.86	0.24	0.94
γ	0.67	0.67	0.69	0.56	0.7	0.67	-0.64	0.39	0.63
Nodes	0.97	0.99	0.99	0.97	1	0.98	-0.93	0.3	0.98
Edges	0.96	0.99	0.98	0.96	0.99	0.97	-0.89	0.23	0.96
Avg. degree	0.73	0.79	0.73	0.76	0.75	0.77	-0.63	-0.063	0.76
Communities	0.96	1	0.98	0.98	0.99	0.99	-0.92	0.28	0.98
Avg. shortest path	0.62	0.73	0.69	0.73	0.74	0.76	-0.83	0.49	0.78
Density	-0.85	-0.88	-0.88	-0.9	-0.89	-0.91	0.96	-0.45	-0.92
Avg. clust. coeff.	0.39	0.5	0.41	0.56	0.45	0.56	-0.61	0.24	0.59
SCC	0.97	0.99	0.99	0.97	1	0.98	-0.93	0.3	0.98
WCC	0.96	1	0.99	0.97	1	0.98	-0.92	0.28	0.98
Lrgst WCC nodes	0.85	0.94	0.89	0.88	0.94	0.92	-0.81	0.22	0.91
Lrgst WCC edges	0.86	0.95	0.9	0.92	0.95	0.94	-0.83	0.21	0.93
γ_in	-0.16	-0.16	-0.16	-0.17	-0.16	-0.16	0.2	-0.16	-0.12
γ_out	-0.48	-0.51	-0.51	-0.51	-0.49	-0.52	0.52	-0.081	-0.51

Cassandra

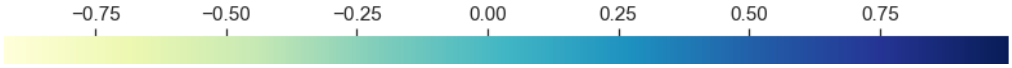
	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.86	0.91	0.86	0.82	0.76	0.8	-0.76	-0.75	0.92
2-stars	0.36	0.38	0.56	0.41	0.72	0.69	-0.55	-0.77	0.37
3-stars	0.062	0.097	0.29	0.11	0.45	0.41	-0.28	-0.61	0.1
4-stars	0.0045	0.055	0.2	0.027	0.31	0.27	-0.18	-0.56	0.085
γ	-0.81	-0.82	-0.67	-0.75	-0.48	-0.54	0.59	0.35	-0.82
Nodes	0.81	0.83	0.92	0.84	0.97	0.97	-0.9	-0.95	0.81
Edges	0.72	0.74	0.86	0.76	0.94	0.93	-0.84	-0.93	0.73
Avg. degree	-0.96	-0.94	-0.86	-0.93	-0.73	-0.76	0.84	0.62	-0.94
Communities	0.94	0.96	0.98	0.94	0.94	0.94	-0.94	-0.93	0.96
Avg. shortest path	0.79	0.76	0.66	0.74	0.48	0.49	-0.66	-0.49	0.79
Density	-0.87	-0.87	-0.96	-0.89	-0.99	-0.98	0.96	0.98	-0.87
Avg. clust. coeff.									
SCC	0.81	0.83	0.92	0.84	0.97	0.97	-0.9	-0.95	0.81
WCC	0.93	0.95	0.98	0.93	0.96	0.97	-0.95	-0.94	0.95
Lrgst WCC nodes	0.39	0.43	0.4	0.33	0.28	0.27	-0.35	-0.57	0.5
Lrgst WCC edges	-0.13	-0.055	-0.065	-0.18	-0.12	-0.13	0.13	-0.22	0.011
γ_in	0.68	0.61	0.55	0.69	0.47	0.49	-0.61	-0.25	0.58
γ_out	-0.74	-0.68	-0.78	-0.8	-0.83	-0.82	0.85	0.65	-0.64

Chukwa

	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.68	0.66	0.69	0.68	0.67	0.68	-0.77	-0.21	0.73
2-stars	0.74	0.68	0.74	0.8	0.7	0.73	-0.47	-0.074	0.85
3-stars	0.63	0.56	0.64	0.71	0.59	0.61	-0.33	-0.03	0.67
4-stars	0.6	0.53	0.61	0.68	0.57	0.58	-0.3	-0.029	0.63
γ	-0.57	-0.56	-0.57	-0.55	-0.56	-0.57	0.62	0.17	-0.61
Nodes	0.82	0.8	0.82	0.82	0.8	0.82	-0.73	-0.28	0.99
Edges	0.83	0.8	0.83	0.84	0.82	0.83	-0.77	-0.33	0.99
Avg. degree	-0.47	-0.48	-0.46	-0.45	-0.44	-0.48	0.29	0.11	-0.57
Communities	0.82	0.8	0.82	0.81	0.81	0.82	-0.81	-0.41	0.95
Avg. shortest path	0.4	0.38	0.4	0.39	0.4	0.4	-0.65	-0.076	0.43
Density	-0.77	-0.75	-0.77	-0.75	-0.78	-0.77	0.96	0.51	-0.8
Avg. clust. coeff.	0.69	0.62	0.7	0.77	0.65	0.68	-0.42	-0.063	0.74
SCC	0.82	0.8	0.82	0.82	0.8	0.82	-0.73	-0.28	0.99
WCC	0.82	0.8	0.82	0.81	0.81	0.82	-0.8	-0.4	0.95
Lrgst WCC nodes	0.7	0.7	0.7	0.69	0.69	0.71	-0.67	-0.17	0.93
Lrgst WCC edges	0.75	0.74	0.75	0.73	0.74	0.75	-0.73	-0.25	0.93
γ_in	0.28	0.3	0.29	0.26	0.3	0.29	-0.26	-0.35	0.27
γ_out	-0.65	-0.63	-0.65	-0.64	-0.65	-0.65	0.76	0.5	-0.68

Hadoop

Inheritance networks



	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.69	0.8	0.82	0.84	0.83	0.7	-0.66	0.97	0.73
2-stars	0.84	0.9	0.92	0.95	0.94	0.82	-0.81	0.88	0.88
3-stars	0.69	0.75	0.78	0.83	0.81	0.67	-0.66	0.77	0.75
4-stars	0.53	0.58	0.61	0.67	0.65	0.5	-0.49	0.63	0.59
γ	-0.78	-0.86	-0.86	-0.91	-0.9	-0.78	0.78	-0.89	-0.83
Nodes	0.94	0.99	0.99	0.98	0.99	0.94	-0.92	0.83	0.96
Edges	0.89	0.95	0.96	0.97	0.97	0.88	-0.86	0.89	0.91
Avg. degree	0.52	0.66	0.7	0.74	0.7	0.5	-0.47	0.95	0.58
Communities	0.96	1	0.99	0.97	0.99	0.96	-0.95	0.79	0.97
Avg. shortest path	0.7	0.81	0.83	0.84	0.83	0.71	-0.67	0.97	0.74
Density	-0.99	-0.96	-0.93	-0.92	-0.95	-0.99	1	-0.56	-0.99
Avg. clust. coeff.	0.71	0.61	0.53	0.51	0.59	0.75	-0.75	0.11	0.68
SCC	0.94	0.99	0.99	0.98	0.99	0.94	-0.92	0.83	0.96
WCC	0.98	1	0.99	0.96	0.99	0.97	-0.97	0.73	0.98
Lrgst WCC nodes	0.72	0.82	0.84	0.87	0.85	0.71	-0.68	0.96	0.76
Lrgst WCC edges	0.67	0.79	0.81	0.84	0.82	0.66	-0.63	0.98	0.72
γ_in	-0.29	-0.2	-0.16	-0.13	-0.15	-0.25	0.23	0.24	-0.24
γ_out	0.13	0.018	-0.067	-0.034	0.011	0.18	-0.21	-0.35	0.12

HttpComponents

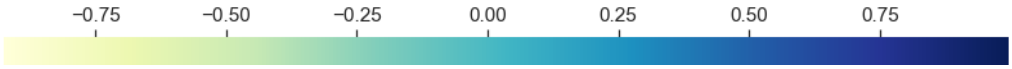
	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.22	0.13	0.2	0.13	0.18	0.15	-0.17	0.037	0.27
2-stars	0.93	0.84	0.93	0.86	0.92	0.83	-0.93	-0.39	0.89
3-stars	0.88	0.77	0.88	0.81	0.87	0.76	-0.89	-0.36	0.83
4-stars	0.84	0.74	0.85	0.78	0.85	0.73	-0.86	-0.31	0.8
γ	0.53	0.34	0.51	0.37	0.48	0.33	-0.49	-0.42	0.49
Nodes	1	0.96	1	0.97	0.99	0.95	-0.99	-0.33	0.99
Edges	1	0.95	1	0.96	0.99	0.94	-0.99	-0.34	0.98
Avg. degree	0.85	0.74	0.84	0.76	0.83	0.73	-0.84	-0.42	0.8
Communities	0.99	0.96	0.99	0.97	0.99	0.96	-0.98	-0.27	0.99
Avg. shortest path	-0.83	-0.79	-0.85	-0.8	-0.85	-0.75	0.86	0.54	-0.77
Density	-0.97	-0.99	-0.98	-1	-0.99	-0.99	0.98	0.17	-0.97
Avg. clust. coeff.	-0.35	-0.16	-0.31	-0.17	-0.27	-0.15	0.28	0.42	-0.33
SCC	1	0.96	1	0.97	0.99	0.95	-0.99	-0.33	0.99
WCC	0.99	0.99	0.99	0.99	0.99	0.98	-0.99	-0.23	0.99
Lrgst WCC nodes	0.88	0.76	0.87	0.79	0.86	0.75	-0.87	-0.36	0.84
Lrgst WCC edges	0.54	0.36	0.52	0.38	0.49	0.36	-0.5	-0.39	0.51
γ_in	-0.36	-0.44	-0.35	-0.44	-0.36	-0.5	0.36	-0.52	-0.43
γ_out	-0.49	-0.48	-0.49	-0.45	-0.47	-0.43	0.46	0.43	-0.45

Ant-Ivy

	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	-0.32	-0.5	-0.75	-0.13	-0.67	-0.4	0.29	0.58	0.4
2-stars	0.6	0.32	-0.11	0.77	0.096	0.49	-0.62	-0.067	0.99
3-stars	0.65	0.37	-0.015	0.81	0.19	0.55	-0.67	-0.12	0.98
4-stars	0.67	0.39	0.028	0.82	0.23	0.57	-0.69	-0.14	0.96
γ	0.59	0.71	0.48	0.56	0.5	0.65	-0.57	-0.68	0.36
Nodes	0.49	0.25	-0.22	0.67	-0.035	0.4	-0.52	-0.025	0.95
Edges	0.48	0.19	-0.25	0.65	-0.053	0.37	-0.5	0.047	0.95
Avg. degree	-0.45	-0.33	0.19	-0.63	0.048	-0.41	0.46	0.17	-0.87
Communities	0.52	0.32	-0.17	0.68	0.0046	0.45	-0.54	-0.12	0.92
Avg. shortest path	-0.29	-0.51	-0.73	-0.096	-0.64	-0.38	0.26	0.61	0.42
Density	-0.42	-0.22	0.29	-0.61	0.11	-0.35	0.44	0.02	-0.92
Avg. clust. coeff.	-0.042	-0.18	0.3	-0.2	0.29	-0.11	0.051	0.22	-0.46
SCC	0.49	0.25	-0.22	0.67	-0.035	0.4	-0.52	-0.025	0.95
WCC	0.54	0.37	-0.13	0.7	0.04	0.48	-0.56	-0.17	0.91
Lrgst WCC nodes	0.4	0.1	-0.33	0.61	-0.12	0.29	-0.43	0.14	0.96
Lrgst WCC edges	0.23	-0.082	-0.48	0.44	-0.29	0.11	-0.25	0.3	0.86
γ_in	-0.019	-0.046	0.088	-0.11	0.075	-0.033	0.031	0.018	-0.28
γ_out	-0.0021	-0.012	0.11	0.0057	0.11	0.022	-0.015	0.015	-0.032

Jena

Methods networks



	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.01	0.023	0.0047	0.045	0.0071	0.022	-0.024	-0.017	0.035
2-stars	-0.082	-0.12	-0.11	-0.081	-0.13	-0.12	0.11	-0.058	-0.1
3-stars	-0.079	-0.1	-0.1	-0.068	-0.12	-0.11	0.11	-0.048	-0.09
4-stars	-0.082	-0.097	-0.099	-0.062	-0.11	-0.1	0.1	-0.044	-0.086
γ	-0.025	-0.036	-0.025	-0.02	-0.034	-0.02	-0.05	0.13	-0.023
Nodes	-0.084	-0.12	-0.11	-0.085	-0.14	-0.12	0.11	-0.048	-0.1
Edges	-0.084	-0.12	-0.11	-0.085	-0.14	-0.12	0.12	-0.052	-0.1
Avg. degree	-0.059	-0.083	-0.081	-0.072	-0.1	-0.092	0.13	-0.16	-0.076
Communities	-0.11	-0.15	-0.14	-0.11	-0.17	-0.15	0.14	-0.08	-0.13
Avg. shortest path	0.12	0.14	0.14	0.12	0.17	0.14	-0.15	0.18	0.12
Density	0.15	0.2	0.18	0.17	0.22	0.2	-0.18	0.064	0.18
Avg. clust. coeff.	0.061	0.11	0.085	0.11	0.13	0.099	-0.063	0.037	0.097
SCC	-0.085	-0.12	-0.11	-0.085	-0.14	-0.12	0.11	-0.048	-0.1
WCC	-0.12	-0.15	-0.15	-0.11	-0.18	-0.15	0.14	-0.091	-0.13
Lrgst WCC nodes	-0.074	-0.11	-0.098	-0.076	-0.12	-0.11	0.1	-0.038	-0.094
Lrgst WCC edges	-0.08	-0.12	-0.11	-0.082	-0.13	-0.12	0.11	-0.049	-0.1
γ_in	0.017	0.053	0.017	0.038	0.032	0.061	0.015	-0.095	0.058
γ_out	0.05	0.083	0.068	0.04	0.096	0.074	-0.082	-0.07	0.06

Cassandra - Excluded from analysis as an outlier

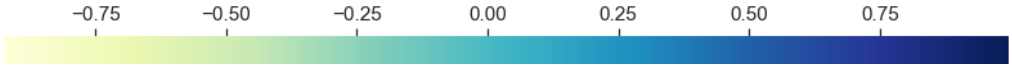
	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.22	0.22	0.36	0.29	0.53	0.54	-0.34	-0.3	0.16
2-stars	0.98	0.95	0.96	0.99	0.91	0.92	-0.98	-0.81	0.94
3-stars	0.86	0.81	0.75	0.86	0.64	0.66	-0.79	-0.5	0.8
4-stars	0.79	0.73	0.66	0.78	0.54	0.56	-0.7	-0.38	0.72
γ	-0.34	-0.35	-0.55	-0.4	-0.71	-0.67	0.55	0.77	-0.34
Nodes	0.99	0.98	0.99	1	0.93	0.95	-0.97	-0.83	0.97
Edges	0.97	0.97	1	0.99	0.97	0.98	-0.98	-0.87	0.95
Avg. degree	0.52	0.52	0.7	0.58	0.84	0.81	-0.71	-0.86	0.51
Communities	0.94	0.91	0.84	0.93	0.72	0.75	-0.84	-0.58	0.9
Avg. shortest path	-0.59	-0.58	-0.39	-0.51	-0.16	-0.2	0.36	0.13	-0.61
Density	-1	-0.99	-0.98	-0.99	-0.92	-0.93	0.97	0.85	-0.99
Avg. clust. coeff.	0.49	0.57	0.44	0.4	0.28	0.31	-0.31	-0.4	0.61
SCC	0.99	0.98	0.99	1	0.93	0.95	-0.97	-0.83	0.97
WCC	0.88	0.84	0.77	0.88	0.66	0.69	-0.79	-0.48	0.83
Lrgst WCC nodes	0.99	0.98	1	0.99	0.95	0.97	-0.98	-0.87	0.97
Lrgst WCC edges	0.97	0.97	1	0.98	0.97	0.98	-0.98	-0.89	0.95
γ_in	0.57	0.65	0.53	0.47	0.41	0.46	-0.38	-0.37	0.65
γ_out	0.77	0.74	0.6	0.73	0.42	0.46	-0.59	-0.28	0.74

Chukwa

	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	-0.34	-0.27	-0.38	-0.38	-0.34	-0.31	0.48	0.73	0.3
2-stars	0.036	0.093	-0.0046	0.0064	0.021	0.085	0.17	0.52	0.71
3-stars	0.056	0.12	0.017	0.02	0.043	0.1	0.14	0.49	0.7
4-stars	0.081	0.14	0.045	0.037	0.068	0.12	0.11	0.47	0.69
γ	-0.56	-0.44	-0.61	-0.65	-0.53	-0.51	0.61	0.051	-0.44
Nodes	-0.04	0.022	-0.083	-0.069	-0.053	0.012	0.25	0.56	0.67
Edges	-0.016	0.043	-0.058	-0.043	-0.03	0.035	0.22	0.56	0.69
Avg. degree	0.58	0.51	0.6	0.63	0.55	0.57	-0.53	0.053	0.73
Communities	-0.071	-0.006	-0.11	-0.099	-0.082	-0.016	0.28	0.56	0.64
Avg. shortest path	-0.16	-0.07	-0.22	-0.22	-0.16	-0.1	0.36	0.47	0.48
Density	-0.063	-0.12	-0.024	-0.031	-0.048	-0.11	-0.14	-0.48	-0.72
Avg. clust. coeff.	0.64	0.54	0.69	0.69	0.63	0.6	-0.76	-0.61	0.1
SCC	-0.041	0.022	-0.084	-0.069	-0.053	0.012	0.25	0.56	0.66
WCC	-0.083	-0.019	-0.13	-0.11	-0.095	-0.03	0.29	0.58	0.64
Lrgst WCC nodes	-0.043	0.02	-0.086	-0.072	-0.056	0.0092	0.25	0.56	0.66
Lrgst WCC edges	-0.018	0.04	-0.06	-0.045	-0.032	0.033	0.23	0.56	0.69
γ_in	-0.1	-0.11	-0.07	-0.1	-0.11	-0.1	0.076	-0.2	-0.017
γ_out	0.21	0.15	0.23	0.28	0.21	0.17	-0.27	0.26	-0.0087

Hadoop

Methods networks



	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.66	0.59	0.54	0.44	0.53	0.72	-0.69	0.18	0.6
2-stars	0.9	0.96	0.97	0.97	0.97	0.89	-0.87	0.91	0.93
3-stars	0.79	0.88	0.9	0.92	0.91	0.77	-0.74	0.97	0.83
4-stars	0.71	0.82	0.85	0.87	0.85	0.7	-0.66	0.97	0.76
γ	-0.96	-0.91	-0.87	-0.86	-0.9	-0.97	0.98	-0.5	-0.96
Nodes	0.96	0.99	0.99	0.99	1	0.95	-0.95	0.82	0.98
Edges	0.96	0.99	0.99	0.99	1	0.95	-0.94	0.83	0.98
Avg. degree	0.97	0.97	0.95	0.96	0.98	0.97	-0.97	0.73	0.99
Communities	0.88	0.95	0.96	0.97	0.96	0.87	-0.85	0.93	0.91
Avg. shortest path	0.1	-0.069	-0.13	-0.16	-0.11	0.11	-0.18	-0.65	0.046
Density	-0.99	-0.96	-0.94	-0.92	-0.95	-0.99	1	-0.61	-0.99
Avg. clust. coeff.	0.36	0.21	0.12	0.082	0.16	0.41	-0.43	-0.36	0.3
SCC	0.96	0.99	0.99	0.99	1	0.95	-0.95	0.82	0.98
WCC	0.85	0.93	0.94	0.95	0.95	0.84	-0.82	0.94	0.89
Lrgst WCC nodes	0.97	1	0.99	0.98	1	0.96	-0.96	0.79	0.99
Lrgst WCC edges	0.96	0.99	0.99	0.99	1	0.95	-0.94	0.83	0.98
γ_in	-0.45	-0.43	-0.43	-0.36	-0.38	-0.42	0.42	-0.16	-0.4
γ_out	0.51	0.58	0.65	0.63	0.58	0.43	-0.42	0.61	0.52

HttpComponents

	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.92	0.95	0.93	0.96	0.94	0.96	-0.94	-0.045	0.93
2-stars	0.96	0.99	0.97	0.99	0.97	1	-0.97	-0.093	0.98
3-stars	0.95	0.97	0.95	0.98	0.97	0.98	-0.96	-0.035	0.97
4-stars	0.94	0.96	0.94	0.97	0.96	0.97	-0.95	-0.0036	0.96
γ	0.45	0.31	0.46	0.33	0.44	0.25	-0.46	-0.74	0.34
Nodes	0.97	1	0.97	1	0.98	1	-0.98	-0.12	0.98
Edges	0.96	0.99	0.97	0.99	0.98	1	-0.97	-0.1	0.98
Avg. degree	0.22	0.3	0.21	0.29	0.21	0.36	-0.19	0.47	0.34
Communities	0.99	0.96	0.99	0.97	0.99	0.95	-0.99	-0.3	0.98
Avg. shortest path	0.98	0.97	0.99	0.98	0.99	0.96	-0.99	-0.24	0.97
Density	-0.88	-0.96	-0.9	-0.96	-0.92	-0.96	0.91	0.012	-0.89
Avg. clust. coeff.	0.82	0.76	0.81	0.76	0.81	0.76	-0.8	-0.24	0.84
SCC	0.97	1	0.97	1	0.98	1	-0.98	-0.12	0.98
WCC	0.99	0.97	0.99	0.98	1	0.97	-1	-0.25	0.98
Lrgst WCC nodes	0.96	1	0.97	1	0.98	1	-0.97	-0.11	0.97
Lrgst WCC edges	0.96	0.99	0.96	0.99	0.97	1	-0.96	-0.091	0.98
γ_in	-0.2	-0.27	-0.2	-0.27	-0.22	-0.3	0.22	-0.24	-0.22
γ_out	-0.5	-0.63	-0.53	-0.64	-0.57	-0.65	0.57	-0.25	-0.5

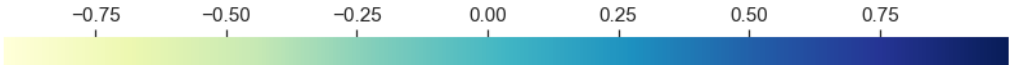
Ant-Ivy

	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter									
2-stars	-0.44	-0.44	-0.44	-0.44	-0.44	-0.43	0.44	0.46	0.95
3-stars	-0.45	-0.45	-0.45	-0.46	-0.45	-0.45	0.45	0.47	0.94
4-stars	-0.46	-0.46	-0.46	-0.46	-0.46	-0.46	0.46	0.48	0.94
γ	-0.38	-0.39	-0.38	-0.39	-0.38	-0.39	0.39	0.4	-0.39
Nodes	-0.45	-0.46	-0.46	-0.46	-0.46	-0.45	0.45	0.48	0.94
Edges	-0.43	-0.43	-0.43	-0.44	-0.43	-0.42	0.43	0.45	0.95
Avg. degree	-0.23	-0.23	-0.23	-0.23	-0.23	-0.23	0.23	0.24	0.98
Communities	-0.28	-0.28	-0.28	-0.28	-0.28	-0.27	0.27	0.3	0.62
Avg. shortest path	0.74	0.74	0.75	0.75	0.74	0.74	-0.74	-0.76	-0.64
Density	0.49	0.49	0.49	0.5	0.49	0.48	-0.49	-0.51	-0.92
Avg. clust. coeff.	0.93	0.93	0.93	0.93	0.93	0.93	-0.93	-0.94	-0.38
SCC	-0.46	-0.46	-0.46	-0.47	-0.46	-0.45	0.46	0.48	0.94
WCC	-0.31	-0.31	-0.31	-0.32	-0.31	-0.31	0.31	0.34	0.96
Lrgst WCC nodes	-0.49	-0.5	-0.49	-0.5	-0.5	-0.49	0.49	0.51	0.93
Lrgst WCC edges	-0.45	-0.45	-0.45	-0.46	-0.45	-0.44	0.45	0.47	0.95
γ_in	-0.62	-0.62	-0.62	-0.61	-0.62	-0.61	0.61	0.61	-0.31
γ_out	0.29	0.29	0.29	0.3	0.29	0.29	-0.29	-0.31	-0.7

Jena

Spearman rank correlation

Call networks



	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	-0.42	-0.43	-0.41	-0.43	-0.41	-0.42	0.43	-0.5	-0.42
2-stars	0.98	0.98	0.98	0.98	0.98	0.97	-0.98	0.33	0.97
3-stars	0.97	0.97	0.98	0.97	0.98	0.97	-0.98	0.34	0.96
4-stars	0.97	0.97	0.97	0.97	0.97	0.97	-0.97	0.34	0.97
γ	-0.56	-0.56	-0.56	-0.55	-0.56	-0.56	0.56	-0.69	-0.53
Nodes	0.96	0.97	0.97	0.98	0.96	0.97	-0.95	0.25	0.98
Edges	0.98	0.99	0.99	0.99	0.98	0.99	-0.98	0.34	0.99
Avg. degree	0.034	0.033	0.04	0.044	0.041	0.046	-0.043	0.56	0.044
Communities	0.89	0.9	0.89	0.91	0.89	0.9	-0.89	0.094	0.92
Avg. shortest path	-0.57	-0.58	-0.57	-0.58	-0.58	-0.58	0.59	-0.61	-0.57
Density	-0.93	-0.94	-0.93	-0.94	-0.93	-0.93	0.93	-0.16	-0.93
Avg. clust. coeff.	-0.041	-0.043	-0.045	-0.047	-0.044	-0.041	0.043	0.074	-0.05
SCC	0.96	0.97	0.97	0.98	0.96	0.97	-0.95	0.25	0.98
WCC	0.85	0.87	0.86	0.88	0.85	0.87	-0.85	0.071	0.9
Lrgst WCC nodes	0.98	0.99	0.98	0.99	0.98	0.98	-0.97	0.33	0.98
Lrgst WCC edges	0.98	0.99	0.99	0.99	0.98	0.99	-0.98	0.34	0.99
γ_in	-0.58	-0.57	-0.58	-0.57	-0.58	-0.57	0.59	-0.12	-0.58
γ_out	-0.56	-0.57	-0.56	-0.57	-0.56	-0.57	0.58	-0.34	-0.56

Cassandra

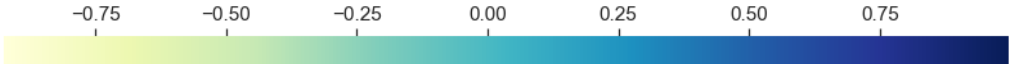
	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	-0.81	-0.81	-0.64	-0.64	-0.46	-0.46	0.64	0.75	-0.81
2-stars	0.94	0.94	1	1	0.83	0.83	-1	-0.5	0.94
3-stars	1	1	0.94	0.94	0.66	0.66	-0.94	-0.5	1
4-stars	0.94	0.94	0.89	0.89	0.6	0.6	-0.89	-0.5	0.94
γ	-0.6	-0.6	-0.54	-0.54	-0.14	-0.14	0.54	-0.088	-0.6
Nodes	0.66	0.66	0.83	0.83	1	1	-0.83	-0.74	0.66
Edges	0.14	0.14	0.37	0.37	0.66	0.66	-0.37	-0.35	0.14
Avg. degree	-0.14	-0.14	0.086	0.086	0.26	0.26	-0.086	0.12	-0.14
Communities	-0.029	-0.029	0.086	0.086	0.49	0.49	-0.086	-0.59	-0.029
Avg. shortest path	-0.94	-0.94	-0.89	-0.89	-0.6	-0.6	0.89	0.5	-0.94
Density	-0.77	-0.77	-0.6	-0.6	-0.43	-0.43	0.6	0.68	-0.77
Avg. clust. coeff.	0.77	0.77	0.66	0.66	0.26	0.26	-0.66	-0.18	0.77
SCC	0.66	0.66	0.83	0.83	1	1	-0.83	-0.74	0.66
WCC	0.21	0.21	0.33	0.33	0.7	0.7	-0.33	-0.72	0.21
Lrgst WCC nodes	0.66	0.66	0.83	0.83	1	1	-0.83	-0.74	0.66
Lrgst WCC edges	0.14	0.14	0.37	0.37	0.66	0.66	-0.37	-0.35	0.14
γ_in	-0.66	-0.66	-0.43	-0.43	-0.14	-0.14	0.43	0.41	-0.66
γ_out	-0.54	-0.54	-0.26	-0.26	0.14	0.14	0.26	0.21	-0.54

Chukwa

	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.71	0.7	0.68	0.72	0.67	0.72	-0.77	-0.3	0.8
2-stars	0.91	0.89	0.91	0.92	0.89	0.93	-0.96	-0.6	1
3-stars	0.91	0.89	0.92	0.91	0.89	0.93	-0.95	-0.61	0.99
4-stars	0.91	0.89	0.92	0.91	0.89	0.93	-0.95	-0.61	0.99
γ	-0.68	-0.67	-0.67	-0.69	-0.66	-0.69	0.72	0.21	-0.77
Nodes	0.9	0.87	0.89	0.9	0.86	0.91	-0.95	-0.59	1
Edges	0.88	0.86	0.89	0.89	0.86	0.9	-0.92	-0.64	0.96
Avg. degree	0.29	0.28	0.32	0.29	0.32	0.29	-0.28	-0.27	0.32
Communities	0.91	0.89	0.91	0.92	0.88	0.93	-0.96	-0.59	1
Avg. shortest path	0.45	0.44	0.42	0.44	0.41	0.46	-0.51	-0.22	0.53
Density	-0.89	-0.86	-0.88	-0.89	-0.85	-0.9	0.94	0.57	-0.98
Avg. clust. coeff.	0.55	0.55	0.55	0.55	0.56	0.54	-0.51	-0.61	0.4
SCC	0.9	0.87	0.89	0.9	0.86	0.91	-0.95	-0.59	1
WCC	0.91	0.89	0.91	0.91	0.88	0.93	-0.96	-0.59	1
Lrgst WCC nodes	0.9	0.87	0.9	0.9	0.87	0.91	-0.94	-0.58	0.99
Lrgst WCC edges	0.87	0.85	0.88	0.87	0.85	0.89	-0.91	-0.62	0.95
γ_in	-0.62	-0.59	-0.6	-0.62	-0.57	-0.63	0.68	0.46	-0.72
γ_out	-0.4	-0.37	-0.38	-0.4	-0.36	-0.41	0.43	0.31	-0.5

Hadoop

Call networks



	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.76	0.79	0.79	0.78	0.79	0.75	-0.8	0.57	0.78
2-stars	0.99	0.99	0.99	0.99	0.99	0.94	-0.98	0.84	0.99
3-stars	0.98	0.99	0.99	0.99	0.99	0.93	-0.97	0.83	0.99
4-stars	0.96	0.95	0.95	0.95	0.95	0.91	-0.93	0.87	0.95
γ	-0.46	-0.42	-0.42	-0.42	-0.43	-0.57	0.42	-0.15	-0.43
Nodes	0.97	0.99	0.99	0.99	0.99	0.93	-0.98	0.86	0.98
Edges	0.96	0.98	0.98	0.98	0.98	0.92	-0.97	0.84	0.99
Avg. degree	0.94	0.96	0.95	0.95	0.95	0.92	-0.96	0.83	0.96
Communities	0.97	0.99	0.99	0.99	0.99	0.93	-0.98	0.83	0.98
Avg. shortest path	0.63	0.71	0.7	0.7	0.7	0.51	-0.7	0.52	0.69
Density	-0.97	-0.98	-0.99	-0.98	-0.99	-0.92	0.98	-0.85	-0.97
Avg. clust. coeff.	0.54	0.52	0.53	0.53	0.53	0.68	-0.53	0.31	0.51
SCC	0.97	0.99	0.99	0.99	0.99	0.93	-0.98	0.86	0.98
WCC	0.94	0.95	0.96	0.96	0.96	0.89	-0.94	0.77	0.95
Lrgst WCC nodes	0.95	0.97	0.97	0.97	0.97	0.91	-0.96	0.86	0.97
Lrgst WCC edges	0.95	0.98	0.97	0.97	0.97	0.91	-0.96	0.84	0.99
γ_in	0.72	0.78	0.79	0.79	0.78	0.71	-0.81	0.86	0.76
γ_out	-0.094	-0.05	-0.019	-0.019	-0.034	-0.12	0.0099	0.26	-0.058

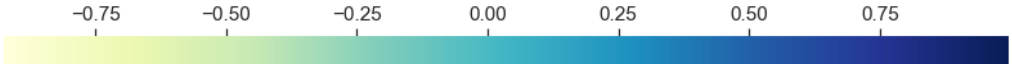
HttpComponents

	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.91	0.91	0.91	0.91	0.91	0.91	-0.93	-0.14	0.91
2-stars	1	1	1	0.99	1	1	-0.96	-0.14	1
3-stars	0.99	1	1	0.99	1	1	-0.95	-0.14	1
4-stars	0.99	0.99	0.99	0.99	1	0.99	-0.95	-0.14	0.99
γ	0.14	0.14	0.14	0.13	0.14	0.14	-0.21	-0.66	0.14
Nodes	1	1	1	0.99	1	1	-0.96	-0.13	1
Edges	0.96	0.96	0.96	0.97	0.96	0.96	-0.93	-0.049	0.96
Avg. degree	-0.47	-0.46	-0.46	-0.45	-0.46	-0.46	0.48	0.47	-0.46
Communities	0.93	0.92	0.93	0.92	0.92	0.93	-0.91	-0.2	0.93
Avg. shortest path	0.83	0.83	0.83	0.83	0.84	0.83	-0.87	-0.11	0.83
Density	-0.99	-0.99	-0.99	-0.99	-0.99	-0.99	0.97	0.14	-0.99
Avg. clust. coeff.	0.44	0.45	0.45	0.45	0.44	0.45	-0.39	-0.43	0.45
SCC	1	1	1	0.99	1	1	-0.96	-0.13	1
WCC	0.94	0.94	0.94	0.94	0.94	0.94	-0.92	-0.2	0.94
Lrgst WCC nodes	1	1	1	0.99	1	1	-0.96	-0.13	1
Lrgst WCC edges	0.96	0.96	0.96	0.97	0.96	0.96	-0.93	-0.049	0.96
γ_in	-0.72	-0.73	-0.72	-0.7	-0.72	-0.73	0.62	-0.15	-0.73
γ_out	-0.65	-0.65	-0.65	-0.64	-0.64	-0.65	0.74	-0.0041	-0.65

Ant-Ivy

	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.011	0.1	0.49	-0.11	0.54	-0.083	-0.007	-0.072	-0.49
2-stars	0.6	0.28	0.086	0.69	0.1	0.5	-0.57	-0.23	0.99
3-stars	0.56	0.23	0.069	0.66	0.089	0.46	-0.55	-0.17	0.98
4-stars	0.54	0.18	0.15	0.6	0.17	0.4	-0.55	-0.1	0.93
γ	0.68	0.31	0.25	0.62	0.21	0.58	-0.72	-0.24	0.76
Nodes	0.6	0.29	0.1	0.68	0.098	0.52	-0.58	-0.25	0.99
Edges	0.72	0.5	0.53	0.7	0.53	0.6	-0.66	-0.47	0.72
Avg. degree	-0.38	-0.048	0.054	-0.31	0.12	-0.29	0.44	0.016	-0.53
Communities	0.66	0.35	0.13	0.72	0.12	0.58	-0.63	-0.31	0.98
Avg. shortest path	0.4	0.5	0.74	0.22	0.7	0.38	-0.38	-0.52	-0.27
Density	-0.61	-0.29	-0.11	-0.67	-0.097	-0.52	0.6	0.25	-0.97
Avg. clust. coeff.	-0.52	-0.22	0.081	-0.5	0.14	-0.47	0.49	0.23	-0.78
SCC	0.6	0.3	0.1	0.68	0.099	0.52	-0.58	-0.25	0.99
WCC	0.64	0.34	0.13	0.71	0.12	0.56	-0.62	-0.3	0.98
Lrgst WCC nodes	0.6	0.29	0.098	0.68	0.1	0.51	-0.58	-0.24	0.99
Lrgst WCC edges	0.69	0.47	0.52	0.67	0.53	0.56	-0.64	-0.44	0.69
γ_in	-0.2	0.11	0.13	-0.3	0.13	-0.056	0.25	-0.091	-0.55
γ_out	-0.58	-0.3	-0.4	-0.56	-0.41	-0.46	0.59	0.26	-0.68

Jena



	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.2	0.18	0.21	0.18	0.21	0.17	-0.21	0.15	0.17
2-stars	0.96	0.97	0.97	0.97	0.96	0.97	-0.96	0.35	0.97
3-stars	0.85	0.86	0.85	0.86	0.85	0.85	-0.85	0.44	0.85
4-stars	0.83	0.84	0.84	0.84	0.84	0.84	-0.83	0.45	0.84
γ	0.25	0.26	0.25	0.26	0.25	0.25	-0.26	0.31	0.27
Nodes	0.98	0.98	0.98	0.98	0.97	0.98	-0.98	0.33	0.99
Edges	0.95	0.96	0.95	0.96	0.95	0.95	-0.94	0.3	0.96
Avg. degree	0.047	0.045	0.051	0.042	0.051	0.049	-0.044	-0.084	0.029
Communities	0.97	0.97	0.97	0.97	0.97	0.97	-0.97	0.34	0.98
Avg. shortest path	0.053	0.05	0.059	0.055	0.056	0.028	-0.06	-0.29	0.075
Density	-0.7	-0.7	-0.7	-0.7	-0.7	-0.69	0.7	-0.42	-0.72
Avg. clust. coeff.	0.6	0.59	0.6	0.6	0.6	0.59	-0.61	0.41	0.57
SCC	0.98	0.98	0.98	0.98	0.97	0.98	-0.98	0.33	0.99
WCC	0.96	0.96	0.96	0.97	0.96	0.96	-0.96	0.3	0.97
Lrgst WCC nodes	0.98	0.98	0.98	0.98	0.98	0.98	-0.98	0.33	0.99
Lrgst WCC edges	0.95	0.96	0.95	0.96	0.95	0.95	-0.94	0.3	0.96
γ_in	0.12	0.1	0.12	0.095	0.12	0.09	-0.12	-0.02	0.086
γ_out	-0.37	-0.37	-0.37	-0.37	-0.36	-0.36	0.36	0.26	-0.35

Cassandra

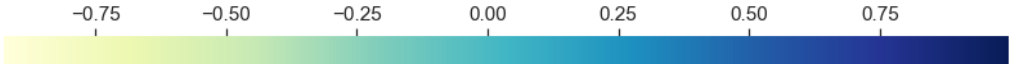
	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.8	0.8	0.93	0.93	0.93	0.93	-0.93	-0.64	0.8
2-stars	0.94	0.94	1	1	0.83	0.83	-1	-0.5	0.94
3-stars	0.94	0.94	1	1	0.83	0.83	-1	-0.5	0.94
4-stars	0.66	0.66	0.83	0.83	1	1	-0.83	-0.74	0.66
γ	0.6	0.6	0.49	0.49	-0.029	-0.029	-0.49	0.29	0.6
Nodes	1	1	0.94	0.94	0.66	0.66	-0.94	-0.5	1
Edges	1	1	0.94	0.94	0.66	0.66	-0.94	-0.5	1
Avg. degree	0.77	0.77	0.6	0.6	0.43	0.43	-0.6	-0.74	0.77
Communities	0.94	0.94	0.89	0.89	0.6	0.6	-0.89	-0.5	0.94
Avg. shortest path	0.83	0.83	0.77	0.77	0.71	0.71	-0.77	-0.74	0.83
Density	-1	-1	-0.94	-0.94	-0.66	-0.66	0.94	0.5	-1
Avg. clust. coeff.	-0.66	-0.66	-0.6	-0.6	-0.2	-0.2	0.6	-0.088	-0.66
SCC	1	1	0.94	0.94	0.66	0.66	-0.94	-0.5	1
WCC	0.99	0.99	0.93	0.93	0.64	0.64	-0.93	-0.51	0.99
Lrgst WCC nodes	0.94	0.94	1	1	0.83	0.83	-1	-0.5	0.94
Lrgst WCC edges	1	1	0.94	0.94	0.66	0.66	-0.94	-0.5	1
γ_in	-0.54	-0.54	-0.37	-0.37	-0.086	-0.086	0.37	0.15	-0.54
γ_out	-0.31	-0.31	-0.43	-0.43	-0.54	-0.54	0.43	0.26	-0.31

Chukwa

	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.71	0.7	0.68	0.72	0.67	0.72	-0.77	-0.3	0.8
2-stars	0.91	0.89	0.91	0.92	0.89	0.93	-0.96	-0.6	1
3-stars	0.91	0.89	0.92	0.91	0.89	0.93	-0.95	-0.61	0.99
4-stars	0.91	0.89	0.92	0.91	0.89	0.93	-0.95	-0.61	0.99
γ	-0.68	-0.67	-0.67	-0.69	-0.66	-0.69	0.72	0.21	-0.77
Nodes	0.91	0.89	0.91	0.92	0.89	0.93	-0.96	-0.59	1
Edges	0.9	0.88	0.9	0.91	0.87	0.92	-0.94	-0.6	0.99
Avg. degree	0.46	0.46	0.46	0.47	0.46	0.45	-0.43	-0.34	0.44
Communities	0.89	0.86	0.88	0.89	0.85	0.9	-0.94	-0.57	0.98
Avg. shortest path	0.65	0.66	0.64	0.65	0.65	0.64	-0.63	-0.65	0.5
Density	-0.9	-0.88	-0.9	-0.9	-0.87	-0.92	0.95	0.58	-0.99
Avg. clust. coeff.	-0.13	-0.13	-0.11	-0.12	-0.12	-0.13	0.11	0.4	0.0032
SCC	0.91	0.89	0.91	0.92	0.89	0.93	-0.96	-0.6	1
WCC	0.7	0.67	0.69	0.7	0.66	0.72	-0.77	-0.54	0.81
Lrgst WCC nodes	0.91	0.89	0.91	0.92	0.89	0.93	-0.96	-0.6	1
Lrgst WCC edges	0.9	0.88	0.9	0.91	0.87	0.92	-0.94	-0.6	0.99
γ_in	-0.73	-0.73	-0.71	-0.72	-0.71	-0.73	0.74	0.55	-0.69
γ_out	-0.49	-0.48	-0.45	-0.49	-0.44	-0.5	0.54	0.24	-0.56

Hadoop

Collaboration networks



	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	-0.7	-0.7	-0.7	-0.7	-0.7	-0.7	0.68	-0.78	-0.7
2-stars	0.97	0.99	0.99	0.99	0.99	0.92	-0.98	0.88	0.98
3-stars	0.97	0.99	0.99	0.99	0.99	0.92	-0.98	0.88	0.98
4-stars	0.97	0.99	0.99	0.99	0.99	0.92	-0.98	0.88	0.98
γ	0.19	0.18	0.18	0.18	0.18	0.32	-0.18	0.39	0.19
Nodes	0.98	0.99	1	1	0.99	0.91	-0.99	0.89	0.99
Edges	0.43	0.44	0.45	0.45	0.44	0.37	-0.44	0.24	0.44
Avg. degree	-0.49	-0.5	-0.5	-0.5	-0.5	-0.54	0.49	-0.61	-0.5
Communities	0.85	0.86	0.86	0.85	0.86	0.71	-0.84	0.76	0.87
Avg. shortest path	-0.94	-0.95	-0.95	-0.95	-0.95	-0.87	0.94	-0.85	-0.95
Density	-0.97	-0.98	-0.98	-0.98	-0.98	-0.9	0.97	-0.88	-0.98
Avg. clust. coeff.	0.8	0.8	0.8	0.8	0.8	0.88	-0.8	0.73	0.8
SCC	0.98	0.99	1	1	0.99	0.91	-0.99	0.89	0.99
WCC	-0.027	-0.022	-0.027	-0.027	-0.027	-0.027	0.022	-0.36	-0.0082
Lrgst WCC nodes	0.98	0.99	1	1	1	0.91	-0.99	0.89	0.99
Lrgst WCC edges	0.43	0.44	0.45	0.45	0.44	0.37	-0.44	0.24	0.44
γ_in	0.055	0.088	0.087	0.087	0.087	-0.0063	-0.061	0.3	0.084
γ_out	0.47	0.47	0.46	0.46	0.46	0.63	-0.46	0.42	0.46

HttpComponents

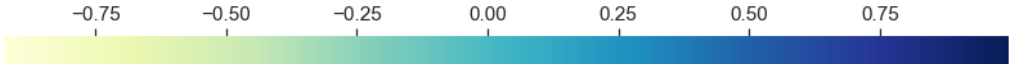
	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.51	0.51	0.51	0.51	0.51	0.51	-0.51	-0.8	0.51
2-stars	1	1	1	0.99	1	1	-0.96	-0.14	1
3-stars	0.98	0.98	0.98	0.96	0.97	0.98	-0.97	-0.076	0.98
4-stars	0.96	0.96	0.96	0.95	0.96	0.96	-0.95	-0.05	0.96
γ	0.4	0.4	0.4	0.4	0.41	0.4	-0.46	-0.71	0.4
Nodes	1	1	1	0.99	1	1	-0.96	-0.14	1
Edges	1	1	1	0.99	1	1	-0.96	-0.13	1
Avg. degree	-0.61	-0.61	-0.61	-0.61	-0.6	-0.61	0.65	0.5	-0.61
Communities	0.9	0.9	0.9	0.91	0.9	0.9	-0.89	-0.27	0.9
Avg. shortest path	-0.81	-0.81	-0.81	-0.81	-0.81	-0.81	0.82	-0.018	-0.81
Density	-0.93	-0.93	-0.93	-0.92	-0.93	-0.93	0.97	0.013	-0.93
Avg. clust. coeff.	0.45	0.45	0.45	0.44	0.45	0.45	-0.45	-0.6	0.45
SCC	1	1	1	0.99	1	1	-0.96	-0.14	1
WCC	0.5	0.5	0.5	0.51	0.5	0.5	-0.47	-0.69	0.5
Lrgst WCC nodes	1	1	1	0.99	1	1	-0.96	-0.14	1
Lrgst WCC edges	1	1	1	0.99	1	1	-0.96	-0.13	1
γ_in	-0.22	-0.22	-0.22	-0.23	-0.21	-0.22	0.31	0.3	-0.22
γ_out	-0.59	-0.59	-0.59	-0.6	-0.59	-0.59	0.64	0.56	-0.59

Ant-Ivy

	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	-0.73	-0.45	-0.32	-0.73	-0.32	-0.62	0.71	0.4	-0.86
2-stars	0.58	0.25	0.4	0.56	0.38	0.43	-0.63	-0.16	0.78
3-stars	0.41	0.099	0.49	0.32	0.47	0.23	-0.49	-0.013	0.49
4-stars	0.39	0.12	0.59	0.3	0.57	0.22	-0.48	-0.04	0.37
γ	0.5	0.3	0.19	0.63	0.2	0.49	-0.54	-0.25	0.88
Nodes	0.61	0.3	0.094	0.66	0.071	0.53	-0.58	-0.27	0.97
Edges	0.55	0.23	-0.021	0.58	-0.073	0.47	-0.52	-0.26	0.85
Avg. degree	-0.53	-0.33	-0.24	-0.66	-0.27	-0.49	0.58	0.23	-0.88
Communities	0.59	0.35	0.12	0.65	0.074	0.57	-0.6	-0.34	0.86
Avg. shortest path	0.0022	0.017	-0.54	0.16	-0.55	0.11	0.091	-0.099	0.44
Density	-0.58	-0.26	-0.13	-0.66	-0.14	-0.49	0.58	0.19	-0.98
Avg. clust. coeff.	-0.67	-0.49	-0.18	-0.77	-0.18	-0.67	0.65	0.44	-0.91
SCC	0.61	0.3	0.094	0.66	0.071	0.53	-0.58	-0.27	0.97
WCC	0.6	0.36	0.15	0.66	0.097	0.57	-0.6	-0.37	0.88
Lrgst WCC nodes	0.61	0.3	0.093	0.66	0.071	0.53	-0.58	-0.27	0.97
Lrgst WCC edges	0.55	0.23	-0.021	0.58	-0.073	0.47	-0.52	-0.26	0.85
γ_in	-0.48	-0.28	-0.52	-0.37	-0.53	-0.33	0.47	0.18	-0.3
γ_out	-0.48	-0.45	-0.27	-0.49	-0.29	-0.51	0.46	0.42	-0.44

Jena

Inheritance networks



	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.48	0.48	0.48	0.48	0.48	0.49	-0.48	0.13	0.49
2-stars	0.98	0.99	0.99	0.99	0.98	0.98	-0.98	0.33	0.99
3-stars	0.98	0.99	0.99	0.99	0.98	0.99	-0.98	0.33	0.98
4-stars	0.98	0.99	0.98	0.99	0.98	0.98	-0.98	0.33	0.98
γ	0.49	0.48	0.48	0.49	0.48	0.48	-0.51	0.37	0.49
Nodes	0.98	0.99	0.99	0.99	0.98	0.99	-0.98	0.33	0.99
Edges	0.98	0.99	0.99	0.99	0.99	0.99	-0.98	0.33	0.99
Avg. degree	0.7	0.71	0.71	0.71	0.71	0.7	-0.7	-0.16	0.71
Communities	0.98	0.99	0.99	0.99	0.98	0.98	-0.99	0.33	0.99
Avg. shortest path	0.72	0.72	0.72	0.72	0.72	0.72	-0.72	0.49	0.72
Density	-0.96	-0.97	-0.96	-0.97	-0.95	-0.96	0.96	-0.37	-0.97
Avg. clust. coeff.	0.65	0.64	0.64	0.63	0.64	0.64	-0.64	0.2	0.63
SCC	0.98	0.99	0.99	0.99	0.98	0.99	-0.98	0.33	0.99
WCC	0.98	0.99	0.99	0.99	0.98	0.98	-0.99	0.32	0.99
Lrgst WCC nodes	0.96	0.97	0.97	0.97	0.97	0.96	-0.97	0.36	0.97
Lrgst WCC edges	0.97	0.98	0.98	0.98	0.98	0.98	-0.98	0.35	0.98
γ_in	0.3	0.27	0.32	0.28	0.32	0.26	-0.31	-0.31	0.3
γ_out	-0.53	-0.53	-0.53	-0.53	-0.53	-0.51	0.53	-0.21	-0.52

Cassandra

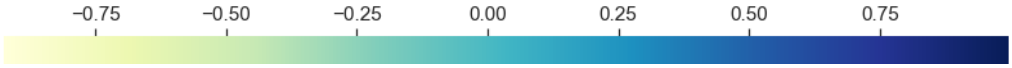
	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.48	0.48	0.48	0.48	0.72	0.72	-0.48	-0.98	0.48
2-stars	-0.058	-0.058	0.058	0.058	0.46	0.46	-0.058	-0.6	-0.058
3-stars	-0.058	-0.058	0.058	0.058	0.46	0.46	-0.058	-0.6	-0.058
4-stars	0.058	0.058	0.12	0.12	0.52	0.52	-0.12	-0.81	0.058
γ	-0.93	-0.93	-0.87	-0.87	-0.58	-0.58	0.87	0.45	-0.93
Nodes	0.64	0.64	0.81	0.81	0.99	0.99	-0.81	-0.75	0.64
Edges	0.64	0.64	0.81	0.81	0.99	0.99	-0.81	-0.75	0.64
Avg. degree	-0.99	-0.99	-0.93	-0.93	-0.64	-0.64	0.93	0.51	-0.99
Communities	0.5	0.5	0.5	0.5	0.74	0.74	-0.5	-1	0.5
Avg. shortest path	0.79	0.79	0.59	0.59	0.35	0.35	-0.59	-0.65	0.79
Density	-0.64	-0.64	-0.81	-0.81	-0.99	-0.99	0.81	0.75	-0.64
Avg. clust. coeff.									
SCC	0.64	0.64	0.81	0.81	0.99	0.99	-0.81	-0.75	0.64
WCC	0.58	0.58	0.7	0.7	0.94	0.94	-0.7	-0.91	0.58
Lrgst WCC nodes	0.51	0.51	0.34	0.34	0.34	0.34	-0.34	-0.78	0.51
Lrgst WCC edges	-0.44	-0.44	-0.53	-0.53	-0.18	-0.18	0.53	-0.41	-0.44
γ_in	0.46	0.46	0.41	0.41	-0.12	-0.12	-0.41	0.51	0.46
γ_out	-0.81	-0.81	-0.93	-0.93	-0.75	-0.75	0.93	0.3	-0.81

Chukwa

	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.71	0.7	0.68	0.72	0.67	0.72	-0.77	-0.3	0.8
2-stars	0.91	0.89	0.91	0.92	0.89	0.93	-0.96	-0.6	1
3-stars	0.91	0.89	0.92	0.91	0.89	0.93	-0.95	-0.61	0.99
4-stars	0.91	0.89	0.92	0.91	0.89	0.93	-0.95	-0.61	0.99
γ	-0.68	-0.67	-0.67	-0.69	-0.66	-0.69	0.72	0.21	-0.77
Nodes	0.91	0.88	0.91	0.91	0.88	0.92	-0.95	-0.59	1
Edges	0.9	0.88	0.91	0.91	0.88	0.92	-0.94	-0.61	0.98
Avg. degree	-0.36	-0.33	-0.35	-0.36	-0.32	-0.38	0.43	0.27	-0.49
Communities	0.89	0.87	0.89	0.9	0.86	0.91	-0.95	-0.58	1
Avg. shortest path	0.25	0.26	0.23	0.27	0.23	0.25	-0.3	0.082	0.33
Density	-0.89	-0.86	-0.88	-0.89	-0.85	-0.9	0.94	0.57	-0.99
Avg. clust. coeff.	0.68	0.66	0.65	0.7	0.63	0.7	-0.77	-0.31	0.81
SCC	0.91	0.88	0.91	0.91	0.88	0.92	-0.95	-0.59	1
WCC	0.89	0.87	0.89	0.9	0.86	0.91	-0.94	-0.58	0.99
Lrgst WCC nodes	0.9	0.87	0.9	0.9	0.87	0.91	-0.94	-0.59	0.99
Lrgst WCC edges	0.89	0.87	0.89	0.9	0.87	0.91	-0.93	-0.61	0.98
γ_in	0.46	0.44	0.45	0.46	0.43	0.48	-0.47	-0.44	0.5
γ_out	-0.8	-0.78	-0.78	-0.8	-0.76	-0.81	0.85	0.61	-0.86

Hadoop

Inheritance networks



	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.93	0.94	0.93	0.93	0.93	0.9	-0.93	0.75	0.93
2-stars	0.98	0.99	0.99	0.99	0.99	0.93	-0.97	0.84	0.99
3-stars	0.98	0.99	0.99	0.99	0.99	0.93	-0.97	0.84	0.99
4-stars	0.99	0.98	0.97	0.97	0.98	0.94	-0.96	0.83	0.98
γ	-0.76	-0.78	-0.78	-0.78	-0.78	-0.72	0.79	-0.57	-0.78
Nodes	0.98	0.99	1	1	1	0.94	-0.98	0.84	0.99
Edges	0.98	0.99	1	1	1	0.94	-0.98	0.84	0.99
Avg. degree	0.64	0.67	0.66	0.66	0.66	0.6	-0.63	0.82	0.66
Communities	0.97	0.96	0.97	0.96	0.96	0.97	-0.96	0.81	0.96
Avg. shortest path	0.9	0.92	0.93	0.92	0.92	0.94	-0.92	0.78	0.92
Density	-0.95	-0.94	-0.94	-0.94	-0.94	-0.95	0.95	-0.82	-0.94
Avg. clust. coeff.	0.069	-0.0081	-0.02	-0.02	-0.012	0.12	0.011	-0.38	0.00099
SCC	0.98	0.99	1	1	1	0.94	-0.98	0.84	0.99
WCC	0.95	0.92	0.93	0.93	0.93	0.97	-0.93	0.8	0.92
Lrgst WCC nodes	0.96	0.98	0.98	0.98	0.98	0.92	-0.97	0.83	0.98
Lrgst WCC edges	0.97	0.98	0.98	0.98	0.98	0.92	-0.97	0.83	0.98
γ_in	0.34	0.37	0.36	0.36	0.36	0.3	-0.39	0.5	0.37
γ_out	-0.021	0.027	0.032	0.031	0.023	0.074	-0.038	-0.23	0.021

HttpComponents

	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.38	0.38	0.38	0.38	0.38	0.38	-0.38	-0.035	0.38
2-stars	0.94	0.94	0.94	0.93	0.94	0.94	-0.95	-0.0067	0.94
3-stars	0.88	0.88	0.88	0.87	0.88	0.88	-0.9	-0.042	0.88
4-stars	0.69	0.69	0.69	0.68	0.69	0.69	-0.7	-0.26	0.69
γ	0.5	0.5	0.5	0.5	0.5	0.5	-0.52	-0.24	0.5
Nodes	1	1	1	0.99	1	1	-0.96	-0.13	1
Edges	1	1	1	0.99	1	1	-0.96	-0.14	1
Avg. degree	0.59	0.59	0.59	0.6	0.6	0.59	-0.64	-0.41	0.59
Communities	0.99	0.99	0.99	1	0.99	0.99	-0.95	-0.14	0.99
Avg. shortest path	-0.61	-0.61	-0.61	-0.61	-0.61	-0.61	0.65	0.5	-0.61
Density	-1	-1	-1	-0.99	-1	-1	0.96	0.13	-1
Avg. clust. coeff.	-0.55	-0.55	-0.55	-0.53	-0.54	-0.55	0.5	0.37	-0.55
SCC	1	1	1	0.99	1	1	-0.96	-0.13	1
WCC	0.99	0.99	0.99	1	0.99	0.99	-0.95	-0.14	0.99
Lrgst WCC nodes	0.74	0.74	0.74	0.74	0.74	0.74	-0.75	-0.19	0.74
Lrgst WCC edges	0.49	0.49	0.49	0.49	0.49	0.49	-0.52	-0.2	0.49
γ_in	-0.24	-0.24	-0.24	-0.23	-0.24	-0.24	0.2	-0.54	-0.24
γ_out	-0.46	-0.47	-0.47	-0.47	-0.46	-0.46	0.33	0.43	-0.46

Ant-Ivy

	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	-0.49	-0.61	-0.74	-0.44	-0.74	-0.5	0.49	0.65	0.028
2-stars	0.62	0.32	0.12	0.71	0.12	0.54	-0.6	-0.27	0.99
3-stars	0.62	0.32	0.11	0.72	0.12	0.54	-0.6	-0.26	0.99
4-stars	0.62	0.32	0.11	0.72	0.12	0.53	-0.59	-0.25	0.99
γ	0.47	0.65	0.38	0.52	0.36	0.6	-0.42	-0.7	0.34
Nodes	0.61	0.32	0.12	0.67	0.095	0.54	-0.6	-0.28	0.97
Edges	0.59	0.28	0.087	0.64	0.061	0.51	-0.57	-0.25	0.96
Avg. degree	-0.65	-0.46	-0.23	-0.74	-0.2	-0.64	0.66	0.43	-0.89
Communities	0.62	0.37	0.12	0.63	0.064	0.59	-0.59	-0.35	0.91
Avg. shortest path	-0.46	-0.66	-0.62	-0.39	-0.59	-0.55	0.43	0.73	0.04
Density	-0.61	-0.32	-0.13	-0.68	-0.11	-0.54	0.61	0.28	-0.97
Avg. clust. coeff.	-0.016	-0.064	0.49	-0.16	0.5	-0.15	-0.026	0.097	-0.42
SCC	0.61	0.32	0.12	0.67	0.095	0.54	-0.6	-0.28	0.97
WCC	0.63	0.45	0.16	0.68	0.11	0.65	-0.61	-0.45	0.89
Lrgst WCC nodes	0.6	0.29	0.11	0.71	0.12	0.51	-0.59	-0.24	0.99
Lrgst WCC edges	0.59	0.25	0.085	0.69	0.097	0.48	-0.58	-0.2	0.99
γ_in	-0.29	-0.18	-0.1	-0.37	-0.13	-0.18	0.29	0.14	-0.49
γ_out	0.02	0.05	0.08	0.08	0.11	0.019	-0.13	-0.044	-0.037

Jena

Methods networks

	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.01	0.023	0.0047	0.045	0.0071	0.022	-0.024	-0.017	0.035
2-stars	-0.082	-0.12	-0.11	-0.081	-0.13	-0.12	0.11	-0.058	-0.1
3-stars	-0.079	-0.1	-0.1	-0.068	-0.12	-0.11	0.11	-0.048	-0.09
4-stars	-0.082	-0.097	-0.099	-0.062	-0.11	-0.1	0.1	-0.044	-0.086
γ	-0.025	-0.036	-0.025	-0.02	-0.034	-0.02	-0.05	0.13	-0.023
Nodes	-0.084	-0.12	-0.11	-0.085	-0.14	-0.12	0.11	-0.048	-0.1
Edges	-0.084	-0.12	-0.11	-0.085	-0.14	-0.12	0.12	-0.052	-0.1
Avg. degree	-0.059	-0.083	-0.081	-0.072	-0.1	-0.092	0.13	-0.16	-0.076
Communities	-0.11	-0.15	-0.14	-0.11	-0.17	-0.15	0.14	-0.08	-0.13
Avg. shortest path	0.12	0.14	0.14	0.12	0.17	0.14	-0.15	0.18	0.12
Density	0.15	0.2	0.18	0.17	0.22	0.2	-0.18	0.064	0.18
Avg. clust. coeff.	0.061	0.11	0.085	0.11	0.13	0.099	-0.063	0.037	0.097
SCC	-0.085	-0.12	-0.11	-0.085	-0.14	-0.12	0.11	-0.048	-0.1
WCC	-0.12	-0.15	-0.15	-0.11	-0.18	-0.15	0.14	-0.091	-0.13
Lrgst WCC nodes	-0.074	-0.11	-0.098	-0.076	-0.12	-0.11	0.1	-0.038	-0.094
Lrgst WCC edges	-0.08	-0.12	-0.11	-0.082	-0.13	-0.12	0.11	-0.049	-0.1
γ_in	0.017	0.053	0.017	0.038	0.032	0.061	0.015	-0.095	0.058
γ_out	0.05	0.083	0.068	0.04	0.096	0.074	-0.082	-0.07	0.06

Cassandra - Excluded from analysis as an outlier

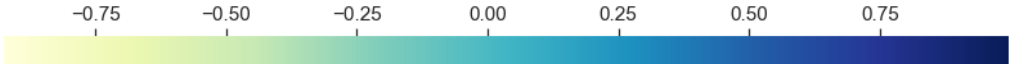
	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.22	0.22	0.36	0.29	0.53	0.54	-0.34	-0.3	0.16
2-stars	0.98	0.95	0.96	0.99	0.91	0.92	-0.98	-0.81	0.94
3-stars	0.86	0.81	0.75	0.86	0.64	0.66	-0.79	-0.5	0.8
4-stars	0.79	0.73	0.66	0.78	0.54	0.56	-0.7	-0.38	0.72
γ	-0.34	-0.35	-0.55	-0.4	-0.71	-0.67	0.55	0.77	-0.34
Nodes	0.99	0.98	0.99	1	0.93	0.95	-0.97	-0.83	0.97
Edges	0.97	0.97	1	0.99	0.97	0.98	-0.98	-0.87	0.95
Avg. degree	0.52	0.52	0.7	0.58	0.84	0.81	-0.71	-0.86	0.51
Communities	0.94	0.91	0.84	0.93	0.72	0.75	-0.84	-0.58	0.9
Avg. shortest path	-0.59	-0.58	-0.39	-0.51	-0.16	-0.2	0.36	0.13	-0.61
Density	-1	-0.99	-0.98	-0.99	-0.92	-0.93	0.97	0.85	-0.99
Avg. clust. coeff.	0.49	0.57	0.44	0.4	0.28	0.31	-0.31	-0.4	0.61
SCC	0.99	0.98	0.99	1	0.93	0.95	-0.97	-0.83	0.97
WCC	0.88	0.84	0.77	0.88	0.66	0.69	-0.79	-0.48	0.83
Lrgst WCC nodes	0.99	0.98	1	0.99	0.95	0.97	-0.98	-0.87	0.97
Lrgst WCC edges	0.97	0.97	1	0.98	0.97	0.98	-0.98	-0.89	0.95
γ_in	0.57	0.65	0.53	0.47	0.41	0.46	-0.38	-0.37	0.65
γ_out	0.77	0.74	0.6	0.73	0.42	0.46	-0.59	-0.28	0.74

Chukwa

	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	-0.34	-0.27	-0.38	-0.38	-0.34	-0.31	0.48	0.73	0.3
2-stars	0.036	0.093	-0.0046	0.0064	0.021	0.085	0.17	0.52	0.71
3-stars	0.056	0.12	0.017	0.02	0.043	0.1	0.14	0.49	0.7
4-stars	0.081	0.14	0.045	0.037	0.068	0.12	0.11	0.47	0.69
γ	-0.56	-0.44	-0.61	-0.65	-0.53	-0.51	0.61	0.051	-0.44
Nodes	-0.04	0.022	-0.083	-0.069	-0.053	0.012	0.25	0.56	0.67
Edges	-0.016	0.043	-0.058	-0.043	-0.03	0.035	0.22	0.56	0.69
Avg. degree	0.58	0.51	0.6	0.63	0.55	0.57	-0.53	0.053	0.73
Communities	-0.071	-0.006	-0.11	-0.099	-0.082	-0.016	0.28	0.56	0.64
Avg. shortest path	-0.16	-0.07	-0.22	-0.22	-0.16	-0.1	0.36	0.47	0.48
Density	-0.063	-0.12	-0.024	-0.031	-0.048	-0.11	-0.14	-0.48	-0.72
Avg. clust. coeff.	0.64	0.54	0.69	0.69	0.63	0.6	-0.76	-0.61	0.1
SCC	-0.041	0.022	-0.084	-0.069	-0.053	0.012	0.25	0.56	0.66
WCC	-0.083	-0.019	-0.13	-0.11	-0.095	-0.03	0.29	0.58	0.64
Lrgst WCC nodes	-0.043	0.02	-0.086	-0.072	-0.056	0.0092	0.25	0.56	0.66
Lrgst WCC edges	-0.018	0.04	-0.06	-0.045	-0.032	0.033	0.23	0.56	0.69
γ_in	-0.1	-0.11	-0.07	-0.1	-0.11	-0.1	0.076	-0.2	-0.017
γ_out	0.21	0.15	0.23	0.28	0.21	0.17	-0.27	0.26	-0.0087

Hadoop

Methods networks



	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.66	0.59	0.54	0.44	0.53	0.72	-0.69	0.18	0.6
2-stars	0.9	0.96	0.97	0.97	0.97	0.89	-0.87	0.91	0.93
3-stars	0.79	0.88	0.9	0.92	0.91	0.77	-0.74	0.97	0.83
4-stars	0.71	0.82	0.85	0.87	0.85	0.7	-0.66	0.97	0.76
γ	-0.96	-0.91	-0.87	-0.86	-0.9	-0.97	0.98	-0.5	-0.96
Nodes	0.96	0.99	0.99	0.99	1	0.95	-0.95	0.82	0.98
Edges	0.96	0.99	0.99	0.99	1	0.95	-0.94	0.83	0.98
Avg. degree	0.97	0.97	0.95	0.96	0.98	0.97	-0.97	0.73	0.99
Communities	0.88	0.95	0.96	0.97	0.96	0.87	-0.85	0.93	0.91
Avg. shortest path	0.1	-0.069	-0.13	-0.16	-0.11	0.11	-0.18	-0.65	0.046
Density	-0.99	-0.96	-0.94	-0.92	-0.95	-0.99	1	-0.61	-0.99
Avg. clust. coeff.	0.36	0.21	0.12	0.082	0.16	0.41	-0.43	-0.36	0.3
SCC	0.96	0.99	0.99	0.99	1	0.95	-0.95	0.82	0.98
WCC	0.85	0.93	0.94	0.95	0.95	0.84	-0.82	0.94	0.89
Lrgst WCC nodes	0.97	1	0.99	0.98	1	0.96	-0.96	0.79	0.99
Lrgst WCC edges	0.96	0.99	0.99	0.99	1	0.95	-0.94	0.83	0.98
γ_in	-0.45	-0.43	-0.43	-0.36	-0.38	-0.42	0.42	-0.16	-0.4
γ_out	0.51	0.58	0.65	0.63	0.58	0.43	-0.42	0.61	0.52

HttpComponents

	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter	0.92	0.95	0.93	0.96	0.94	0.96	-0.94	-0.045	0.93
2-stars	0.96	0.99	0.97	0.99	0.97	1	-0.97	-0.093	0.98
3-stars	0.95	0.97	0.95	0.98	0.97	0.98	-0.96	-0.035	0.97
4-stars	0.94	0.96	0.94	0.97	0.96	0.97	-0.95	-0.0036	0.96
γ	0.45	0.31	0.46	0.33	0.44	0.25	-0.46	-0.74	0.34
Nodes	0.97	1	0.97	1	0.98	1	-0.98	-0.12	0.98
Edges	0.96	0.99	0.97	0.99	0.98	1	-0.97	-0.1	0.98
Avg. degree	0.22	0.3	0.21	0.29	0.21	0.36	-0.19	0.47	0.34
Communities	0.99	0.96	0.99	0.97	0.99	0.95	-0.99	-0.3	0.98
Avg. shortest path	0.98	0.97	0.99	0.98	0.99	0.96	-0.99	-0.24	0.97
Density	-0.88	-0.96	-0.9	-0.96	-0.92	-0.96	0.91	0.012	-0.89
Avg. clust. coeff.	0.82	0.76	0.81	0.76	0.81	0.76	-0.8	-0.24	0.84
SCC	0.97	1	0.97	1	0.98	1	-0.98	-0.12	0.98
WCC	0.99	0.97	0.99	0.98	1	0.97	-1	-0.25	0.98
Lrgst WCC nodes	0.96	1	0.97	1	0.98	1	-0.97	-0.11	0.97
Lrgst WCC edges	0.96	0.99	0.96	0.99	0.97	1	-0.96	-0.091	0.98
γ_in	-0.2	-0.27	-0.2	-0.27	-0.22	-0.3	0.22	-0.24	-0.22
γ_out	-0.5	-0.63	-0.53	-0.64	-0.57	-0.65	0.57	-0.25	-0.5

Ant-Ivy

	Avg. cc	CBO	DIT	LCOM4	NOC	RFC	COF	WMC	Cog. c
Diameter									
2-stars	-0.44	-0.44	-0.44	-0.44	-0.44	-0.43	0.44	0.46	0.95
3-stars	-0.45	-0.45	-0.45	-0.46	-0.45	-0.45	0.45	0.47	0.94
4-stars	-0.46	-0.46	-0.46	-0.46	-0.46	-0.46	0.46	0.48	0.94
γ	-0.38	-0.39	-0.38	-0.39	-0.38	-0.39	0.39	0.4	-0.39
Nodes	-0.45	-0.46	-0.46	-0.46	-0.46	-0.45	0.45	0.48	0.94
Edges	-0.43	-0.43	-0.43	-0.44	-0.43	-0.42	0.43	0.45	0.95
Avg. degree	-0.23	-0.23	-0.23	-0.23	-0.23	-0.23	0.23	0.24	0.98
Communities	-0.28	-0.28	-0.28	-0.28	-0.28	-0.27	0.27	0.3	0.62
Avg. shortest path	0.74	0.74	0.75	0.75	0.74	0.74	-0.74	-0.76	-0.64
Density	0.49	0.49	0.49	0.5	0.49	0.48	-0.49	-0.51	-0.92
Avg. clust. coeff.	0.93	0.93	0.93	0.93	0.93	0.93	-0.93	-0.94	-0.38
SCC	-0.46	-0.46	-0.46	-0.47	-0.46	-0.45	0.46	0.48	0.94
WCC	-0.31	-0.31	-0.31	-0.32	-0.31	-0.31	0.31	0.34	0.96
Lrgst WCC nodes	-0.49	-0.5	-0.49	-0.5	-0.5	-0.49	0.49	0.51	0.93
Lrgst WCC edges	-0.45	-0.45	-0.45	-0.46	-0.45	-0.44	0.45	0.47	0.95
γ_in	-0.62	-0.62	-0.62	-0.61	-0.62	-0.61	0.61	0.61	-0.31
γ_out	0.29	0.29	0.29	0.3	0.29	0.29	-0.29	-0.31	-0.7

Jena

Appendix C Correlation tables

Correlation tables between each network metric and software metrics for all software systems for all network regularities. Avg. cc indicates Average cyclomatic complexity. Call Networks

Correlation between communities and software metrics

Communities	Cassandra	Chukwa	Hadoop	Http-Components	IVY	Jena
Avg. cc.	0.9103	0.1692	0.8503	0.9589	0.9930	0.6905
CBO	0.9213	0.1947	0.8252	0.9875	0.9765	0.4057
DIT	0.9230	0.3924	0.8477	0.9828	0.9951	-0.0038
LCOM4	0.9197	0.2374	0.8552	0.9880	0.9820	0.8305
NOC	0.9159	0.5782	0.8342	0.9973	0.9950	0.2059
RFC	0.9344	0.5448	0.8537	0.9478	0.9667	0.5914
COF	-0.9480	-0.3874	-0.7819	-0.9445	-0.9946	-0.7093
WMC	0.2908	-0.6250	-0.3010	0.7870	-0.2965	-0.1451
Cognitive c.	0.9444	0.1787	0.9896	0.9783	0.9792	0.9762
Struct. ccomplexity	0.9126	0.1009	0.8388	0.9574	0.9531	0.8378

Correlation between Density and software metrics

Density	Cassandra	Chukwa	Hadoop	Http-Components	IVY	Jena
Avg. cc	0.9103	0.1692	0.8503	0.9589	0.9930	0.6905
CBO	0.9213	0.1947	0.8252	0.9875	0.9765	0.4057
DIT	0.9230	0.3924	0.8477	0.9828	0.9951	-0.0038
LCOM4	0.9197	0.2374	0.8552	0.9880	0.9820	0.8305
NOC	0.9159	0.5782	0.8342	0.9973	0.9950	0.2059
RFC	0.9344	0.5448	0.8537	0.9478	0.9667	0.5914
COF	-0.9480	-0.3874	-0.7819	-0.9445	-0.9946	-0.7093
WMC	0.2908	-0.6250	-0.3010	0.7870	-0.2965	-0.1451
Cognitive c.	0.9444	0.1787	0.9896	0.9783	0.9792	0.9762
Struct. complexity	0.9126	0.1009	0.8388	0.9574	0.9531	0.8378

Correlation between Diameter and software metrics

Diameter	Cassandra	Chukwa	Hadoop	Http-Components	IVY	Jena
Avg. cc	-0.3615	-0.9214	0.6835	0.8538	0.8334	0.0986
CBO	-0.4296	-0.9276	0.6558	0.8846	0.8899	0.1849
DIT	-0.3884	-0.8499	0.6883	0.8668	0.8418	0.5995
LCOM4	-0.4011	-0.8752	0.6808	0.8979	0.9021	-0.0735
NOC	-0.4050	-0.6992	0.6714	0.9050	0.8600	0.5187
RFC	-0.4968	-0.7200	0.6794	0.8659	0.9200	0.1410
COF	0.5515	0.8146	-0.7680	-0.8791	-0.8595	-0.0842
WMC	-0.5563	0.7151	-0.2138	0.7497	0.1927	-0.2472
Cognitive c.	-0.4887	-0.9505	0.7325	0.8812	0.8716	-0.5120
Struct. complexity	-0.3761	-0.9641	0.6701	0.8525	0.9259	-0.1551

Correlation between Weakly connected components and software metrics

WCC	Cassandra	Chukwa	Hadoop	Http-Components	IVY	Jena
Avg. cc	0.8931	0.4737	0.8484	0.9662	0.9958	0.6714
CBO	0.8945	0.5215	0.8244	0.9908	0.9694	0.3781
DIT	0.9025	0.6440	0.8448	0.9842	0.9962	-0.0375
LCOM4	0.8934	0.5172	0.8543	0.9849	0.9770	0.8198
NOC	0.8897	0.7583	0.8297	0.9967	0.9944	0.1752
RFC	0.9087	0.7566	0.8524	0.9557	0.9622	0.5689
COF	-0.9322	-0.5824	-0.7670	-0.9541	-0.9944	-0.6916
WMC	0.2756	-0.7272	-0.2791	0.7729	-0.2875	-0.1137
Cognitive c.	0.9208	0.4990	0.9883	0.9817	0.9860	0.9834
Struct. complexity	0.8844	0.4329	0.8389	0.9478	0.9489	0.8270

caption*Correlation between average clustering coeffient and software

Avg. clust. coeff.	Cassandra	Chukwa	Hadoop	Http-Components	IVY	Jena
Avg. cc	0.3199	0.7545	0.4184	0.8699	0.3586	-0.2239
CBO	0.3140	0.7329	0.3705	0.7838	0.2096	-0.0352
DIT	0.3519	0.5841	0.4265	0.7231	0.3183	0.4534
LCOM4	0.2171	0.7080	0.4260	0.6950	0.1977	-0.4339
NOC	0.3769	0.4097	0.4345	0.7515	0.2682	0.3077
RFC	0.3262	0.4570	0.3955	0.8839	0.1907	-0.1557
COF	-0.2778	-0.5603	-0.6051	-0.9034	-0.2636	0.2496
WMC	0.3448	-0.2345	-0.5361	0.2365	-0.4662	-0.1288
Cognitive c.	0.2961	0.7238	0.2527	0.8409	0.3731	-0.8074
Struct. complexity	0.2729	0.7716	0.4015	0.5888	0.1706	-0.4918

Correlation between average degree and software

Avg. degree	Cassandra	Chukwa	Hadoop	Http-Components	IVY	Jena
Avg. cc	0.3199	0.7545	0.4184	0.8699	0.3586	-0.2239
CBO	0.3140	0.7329	0.3705	0.7838	0.2096	-0.0352
DIT	0.3519	0.5841	0.4265	0.7231	0.3183	0.4534
LCOM4	0.2171	0.7080	0.4260	0.6950	0.1977	-0.4339
NOC	0.3769	0.4097	0.4345	0.7515	0.2682	0.3077
RFC	0.3262	0.4570	0.3955	0.8839	0.1907	-0.1557
COF	-0.2778	-0.5603	-0.6051	-0.9034	-0.2636	0.2496
WMC	0.3448	-0.2345	-0.5361	0.2365	-0.4662	-0.1288
Cognitive c.	0.2961	0.7238	0.2527	0.8409	0.3731	-0.8074
Struct. complexity	0.2729	0.7716	0.4015	0.5888	0.1706	-0.4918

Correlation between average shortest path and software

Avg. shortest path	Cassandra	Chukwa	Hadoop	Http-Components	IVY	Jena
Avg. cc	-0.6334	-0.8901	0.4486	0.8098	0.9580	0.4891
CBO	-0.6630	-0.8625	0.4534	0.7384	0.9420	0.4909
DIT	-0.6685	-0.8154	0.4446	0.6962	0.9685	0.7161
LCOM4	-0.6429	-0.8613	0.4335	0.7440	0.9613	0.2978
NOC	-0.6819	-0.6819	0.4281	0.7575	0.9781	0.7111
RFC	-0.7028	-0.6861	0.4568	0.7979	0.9391	0.5054
COF	0.8152	0.8336	-0.3773	-0.8625	-0.9839	-0.4737
WMC	-0.5521	0.7048	-0.0593	0.2749	-0.2069	-0.4635
Cognitive c	-0.6961	-0.8865	0.4817	0.8249	0.9366	-0.1572
Struct. complexity	-0.6320	-0.9066	0.4579	0.7151	0.9300	0.2252

Correlation between in degree distribution γ_{in} and software

γ_{in}	Cassandra	Chukwa	Hadoop	Http-Components	IVY	Jena
Avg. cc	-0.5413	-0.5284	-0.5982	0.5923	-0.6263	-0.0005
CBO	-0.6004	-0.5349	-0.5920	0.7063	-0.7610	0.1520
DIT	-0.5582	-0.3259	-0.5947	0.7577	-0.6418	0.3023
LCOM4	-0.5539	-0.4448	-0.5909	0.7628	-0.7524	-0.1355
NOC	-0.5865	-0.0908	-0.5896	0.7188	-0.6639	0.2232
RFC	-0.6042	-0.1417	-0.6054	0.5634	-0.7880	0.0624
COF	0.5200	0.2703	0.5692	-0.5595	0.6509	0.0220
WMC	-0.1095	0.0335	0.3329	0.8448	-0.2973	-0.2034
Cognitive c.	-0.5938	-0.5542	-0.6663	0.6299	-0.6805	-0.3926
Struct. complexity	-0.5695	-0.6144	-0.6046	0.7818	-0.8021	-0.1407

Collaboration Networks

Correlation between communities and software metrics

Communities	Cassandra	Chukwa	Hadoop	Http-Components	IVY	Jena
Avg. cc	0.9301	0.9419	0.8338	0.9065	0.9356	0.4590
CBO	0.9266	0.9079	0.8093	0.9139	0.9743	0.2608
DIT	0.9389	0.8515	0.8356	0.9283	0.9485	-0.2269
LCOM4	0.9002	0.9240	0.8356	0.8880	0.9764	0.6316
NOC	0.9327	0.7389	0.8149	0.9002	0.9617	-0.0552
RFC	0.8940	0.7611	0.8336	0.8642	0.9713	0.3923
COF	-0.8112	-0.8606	-0.7579	-0.8676	-0.9550	-0.4785
WMC	0.1678	-0.6129	-0.3160	0.6452	-0.1594	-0.0637
Cognitive c	0.8838	0.9011	0.9545	0.8999	0.9407	0.8922
Struct. complexity	0.9241	0.9221	0.8240	0.8659	0.9712	0.6842

Correlation between Density and software metrics

Density	Cassandra	Chukwa	Hadoop	Http-Components	IVY	Jena
Avg. cc	-0.6651	-0.9979	-0.7681	-0.9232	-0.9769	-0.5215
CBO	-0.7320	-0.9867	-0.7382	-0.9685	-0.9823	-0.2490
DIT	-0.7070	-0.9716	-0.7672	-0.9655	-0.9852	0.1893
LCOM4	-0.7567	-0.9911	-0.7516	-0.9699	-0.9889	-0.7027
NOC	-0.7290	-0.8920	-0.7733	-0.9783	-0.9926	-0.0160
RFC	-0.7757	-0.9072	-0.7632	-0.9204	-0.9759	-0.4194
COF	0.8623	0.9589	0.9681	0.9045	0.9911	0.5441
WMC	-0.4512	0.8193	0.4633	-0.8674	0.2244	-0.0032
Cognitive c	-0.7966	-0.9849	-0.8004	-0.9465	-0.9656	-0.9668
Struct. complexity	-0.7302	-0.9848	-0.7394	-0.9371	-0.9664	-0.7157

Correlation between Diameter and software metrics

Diameter	Cassandra	Chukwa	Hadoop	Http-Components	IVY	Jena
Avg. cc	-0.0407	0.8759	0.6835	-0.5269	0.7269	-0.6625
CBO	-0.0865	0.8766	0.6558	-0.6642	0.5545	-0.4095
DIT	-0.0689	0.8557	0.6883	-0.7026	0.7114	-0.0657
LCOM4	-0.2157	0.8804	0.6808	-0.7192	0.5676	-0.7798
NOC	-0.0881	0.8259	0.6714	-0.6906	0.6796	-0.2523
RFC	-0.1056	0.8642	0.6794	-0.5197	0.5100	-0.5793
COF	0.3289	-0.8105	-0.7680	0.4656	-0.6901	0.6824
WMC	0.1306	-0.5893	-0.2138	-0.9628	-0.7537	0.1731
Cognitive c	-0.1605	0.8360	0.7325	-0.5752	0.6543	-0.8727
Struct. complexity	-0.1670	0.8273	0.6701	-0.7383	0.4733	-0.7735

Correlation between Weakly connected components and software metrics

WCC	Cassandra	Chukwa	Hadoop	Http-Components	IVY	Jena
Avg. cc	0.8951	0.8850	0.7720	0.4120	0.6010	0.4519
CBO	0.9549	0.8379	0.7581	0.2683	0.4767	0.2592
DIT	0.9280	0.7812	0.7701	0.1870	0.5731	-0.2334
LCOM4	0.9305	0.8766	0.7654	0.2047	0.4600	0.6285
NOC	0.9463	0.6776	0.7493	0.2568	0.5334	-0.0630
RFC	0.9664	0.7008	0.7745	0.4316	0.4355	0.3852
COF	-0.9239	-0.8048	-0.6844	-0.4888	-0.5261	-0.4717
WMC	0.3864	-0.5074	-0.2711	-0.2755	-0.7553	-0.0684
Cognitive c.	0.9655	0.8224	0.8946	0.3842	0.5726	0.8950
Struct. complexity	0.9347	0.8503	0.7715	0.1304	0.3916	0.6811

caption*Correlation between average clustering coeffient and software

Avg. clust. coeff.	Cassandra	Chukwa	Hadoop	Http-Components	IVY	Jena
Avg. cc	0.4084	-0.5170	0.0923	0.7520	0.8863	-0.3515
CBO	0.5224	-0.4920	0.0818	0.8471	0.7665	-0.2553
DIT	0.4419	-0.3087	0.0761	0.8572	0.8755	0.3063
LCOM4	0.4923	-0.4687	0.1237	0.8652	0.7798	-0.5539
NOC	0.4832	-0.1176	0.0687	0.8613	0.8582	0.1800
RFC	0.5812	-0.1699	0.1014	0.7634	0.7432	-0.3185
COF	-0.6014	0.2872	0.0516	-0.7271	-0.8650	0.3648
WMC	0.4416	-0.0687	0.3448	0.9541	-0.5003	0.1185
Cognitive c.	0.5628	-0.4833	0.2845	0.7878	0.8387	-0.8383
Struct. complexity	0.4716	-0.5502	0.0726	0.8401	0.7141	-0.6520

Correlation between average degree and software

Avg. degree	Cassandra	Chukwa	Hadoop	Http-Components	IVY	Jena
Avg. cc	0.3746	0.9518	0.3042	-0.1479	-0.8533	-0.5681
CBO	0.2845	0.9743	0.2980	-0.3096	-0.7473	-0.3129
DIT	0.3362	0.9296	0.3221	-0.3539	-0.8598	-0.1675
LCOM4	0.2667	0.9102	0.2908	-0.4025	-0.7819	-0.6454
NOC	0.2913	0.8242	0.3256	-0.3570	-0.8562	-0.3519
RFC	0.2227	0.8465	0.3034	-0.1554	-0.7267	-0.4664
COF	-0.1331	-0.8737	-0.4351	0.0984	0.8743	0.5818
WMC	-0.2005	-0.8165	-0.2376	-0.8168	0.4599	0.0567
Cognitive c.	0.2017	0.9855	0.3687	-0.2154	-0.7923	-0.6767
Struct. complexity	0.2928	0.9751	0.2901	-0.4519	-0.7075	-0.5745

Correlation between average shortest path and software

Avg. shortest path	Cassandra	Chukwa	Hadoop	Http-Components	IVY	Jena
Avg. cc	0.0583	0.8545	0.4601	-0.8109	-0.7203	0.0990
CBO	0.0176	0.8308	0.4256	-0.8953	-0.6891	0.0979
DIT	0.0377	0.8789	0.4770	-0.9074	-0.7252	-0.4477
LCOM4	-0.0422	0.9062	0.4429	-0.9204	-0.7304	0.2955
NOC	0.0222	0.8486	0.4833	-0.9144	-0.7363	-0.3848
RFC	-0.0456	0.8414	0.4408	-0.8089	-0.7293	0.0951
COF	0.2152	-0.8988	-0.7489	0.7799	0.7452	-0.1054
WMC	-0.3205	-0.8310	-0.5812	-0.9525	-0.1495	-0.0562
Cognitive c.	-0.0459	0.8361	0.2818	-0.8474	-0.7583	0.6210
Struct.complexity	-0.0002	0.8432	0.4350	-0.9062	-0.7357	0.4282

Correlation between in degree distribution γ_{in} and software

γ_{in}	Cassandra	Chukwa	Hadoop	Http-Components	IVY	Jena
Avg. cc	0.3381	-0.3361	-0.6699	-0.1573	-0.1695	-0.0528
CBO	0.2278	-0.3905	-0.6656	-0.0035	-0.0463	0.0913
DIT	0.2908	-0.1677	-0.6581	0.0418	-0.1489	0.0251
LCOM4	0.1729	-0.2624	-0.6443	0.1347	-0.0605	-0.0519
NOC	0.2313	0.0350	-0.6736	0.0594	-0.1268	-0.0274
RFC	0.1841	-0.0378	-0.6696	-0.1673	-0.0465	0.0086
COF	-0.0748	0.0390	0.7804	0.2022	0.1360	0.0544
WMC	-0.0678	-0.1486	0.4412	0.5734	0.2745	-0.1918
Cognitive c.	0.1525	-0.3928	-0.5662	-0.0810	-0.1465	-0.0611
Struct.complexity	0.2052	-0.4433	-0.6562	0.2181	0.0053	-0.0040

Inheritance Networks

Correlation between communities and software metrics

Communities	Cassandra	Chukwa	Hadoop	Http-Components	IVY	Jena
Avg. cc	0.9578	0.9438	0.8185	0.9659	0.9913	0.5199
CBO	0.9969	0.9593	0.8004	0.9956	0.9630	0.3234
DIT	0.9822	0.9802	0.8174	0.9927	0.9895	-0.1673
LCOM4	0.9757	0.9362	0.8074	0.9715	0.9694	0.6800
NOC	0.9941	0.9362	0.8117	0.9912	0.9857	0.0046
RFC	0.9868	0.9438	0.8212	0.9607	0.9590	0.4506
COF	-0.9218	-0.9441	-0.8136	-0.9473	-0.9832	-0.5369
WMC	0.2779	-0.9333	-0.4083	0.8066	-0.2689	-0.1214
Cognitive c.	0.9786	0.9644	0.9492	0.9749	0.9921	0.9160
0.6842						
Struct.complexity	0.9893	0.9397	0.8054	0.9250	0.9470	0.7269

Correlation between Density and software metrics

Density	Cassandra	Chukwa	Hadoop	Http-Components	IVY	Jena
Avg. cc	-0.8454	-0.8689	-0.7711	-0.9895	-0.9679	-0.4208
CBO	-0.8840	-0.8692	-0.7453	-0.9600	-0.9925	-0.2184
DIT	-0.8803	-0.9591	-0.7723	-0.9301	-0.9779	0.2862
LCOM4	-0.9022	-0.8926	-0.7514	-0.9103	-0.9969	-0.6147
NOC	-0.8909	-0.9852	-0.7791	-0.9460	-0.9871	0.1140
RFC	-0.9061	-0.9758	-0.7669	-0.9891	-0.9899	-0.3486
COF	0.9588	0.9553	0.9594	0.9990	0.9828	0.4422
WMC	-0.4471	0.9773	0.5139	-0.5822	0.1740	0.0196
Cognitive c.	-0.9174	-0.8652	-0.7966	-0.9825	-0.9702	-0.9232
Struct.complexity	-0.8884	-0.8254	-0.7469	-0.8405	-0.9869	-0.6686

Correlation between Diameter and software metrics

Diameter	Cassandra	Chukwa	Hadoop	Http-Components	IVY	Jena
Avg. cc	0.6319	0.8586	0.6835	0.7119	0.2225	-0.3168
CBO	0.6734	0.9132	0.6558	0.8192	0.1277	-0.4956
DIT	0.6812	0.8579	0.6883	0.8422	0.1984	-0.7540
LCOM4	0.6928	0.8172	0.6808	0.8669	0.1290	-0.1253
NOC	0.7190	0.7638	0.6714	0.8518	0.1801	-0.6743
RFC	0.6937	0.7979	0.6794	0.7107	0.1507	-0.3967
COF	-0.7777	-0.7555	-0.7680	-0.6752	-0.1675	0.2931
WMC	0.3300	-0.7461	-0.2138	0.9778	0.0372	0.5780
Cognitive c.	0.7134	0.9195	0.7325	0.7610	0.2748	0.4031
Struct.complexity	0.7008	0.9020	0.6701	0.8731	0.1436	-0.0890

Correlation between Weakly connected components and software metrics

WCC	Cassandra	Chukwa	Hadoop	Http-Components	IVY	Jena
Avg. cc	0.9650	0.9338	0.8191	0.9799	0.9877	0.5446
CBO	0.9955	0.9490	0.8019	0.9954	0.9880	0.3665
DIT	0.9873	0.9841	0.8176	0.9872	0.9922	-0.1280
LCOM4	0.9745	0.9339	0.8080	0.9584	0.9935	0.6974
NOC	0.9951	0.9593	0.8110	0.9834	0.9949	0.0395
RFC	0.9834	0.9662	0.8221	0.9754	0.9844	0.4835
COF	-0.9213	-0.9488	-0.7991	-0.9677	-0.9914	-0.5600
WMC	0.2827	-0.9389	-0.4041	0.7522	-0.2299	-0.1686
Cognitive c.	0.9755	0.9483	0.9509	0.9809	0.9897	0.9059
Struct.complexity	0.9884	0.9186	0.8074	0.9010	0.9767	0.7470

caption*Correlation between average clustering coeffient and software

Avg. clust. coeff.	Cassandra	Chukwa	Hadoop	Http-Components	IVY	Jena
Avg. cc	0.3867		0.6922	0.6591	-0.3492	-0.0423
CBO	0.4955		0.6234	0.5528	-0.1561	-0.1847
DIT	0.4085		0.7012	0.4733	-0.3142	0.3003
LCOM4	0.5632		0.7652	0.4350	-0.1715	-0.2045
NOC	0.4509		0.6531	0.5202	-0.2728	0.2934
RFC	0.5562		0.6773	0.7095	-0.1471	-0.1080
COF	-0.6053		-0.4215	-0.7090	0.2825	0.0506
WMC	0.2426		-0.0631	0.0710	0.4212	0.2156
Cognitive c.	0.5901		0.7381	0.6200	-0.3329	-0.4594
Struct.complexity	0.5185		0.6802	0.3186	-0.1168	-0.3464

Correlation between average degree and software

Avg. degree	Cassandra	Chukwa	Hadoop	Http-Components	IVY	Jena
Avg. cc	0.7261	-0.9553	-0.4677	0.5695	0.8512	-0.4467
CBO	0.7924	-0.9385	-0.4834	0.6968	0.7405	-0.3306
DIT	0.7315	-0.8607	-0.4601	0.7411	0.8432	0.1926
LCOM4	0.7640	-0.9261	-0.4515	0.7846	0.7599	-0.6277
NOC	0.7517	-0.7286	-0.4398	0.7459	0.8278	0.0476
RFC	0.7745	-0.7579	-0.4771	0.5443	0.7256	-0.4116
COF	-0.6277	0.8438	0.2934	-0.5100	-0.8393	0.4639
WMC	-0.0634	0.6182	0.1084	0.9583	-0.4160	0.1657
Cognitive c.	0.7620	-0.9363	-0.5713	0.6315	0.8029	-0.8651
Struct.complexity	0.7799	-0.9575	-0.4907	0.8355	0.6927	-0.7013

Correlation between average shortest path and software

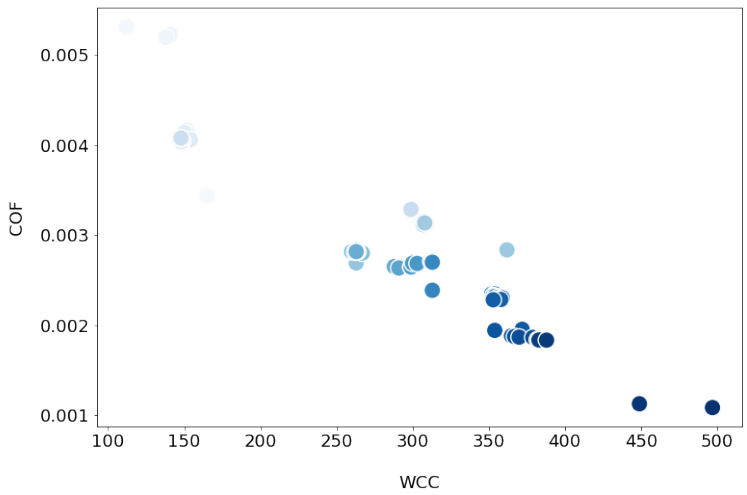
Avg. shortest path	Cassandra	Chukwa	Hadoop	Http-Components	IVY	Jena
Avg. cc	0.6192	0.7882	0.3996	0.7205	-0.8343	-0.2891
CBO	0.7264	0.7634	0.3795	0.8261	-0.7878	-0.5071
DIT	0.6877	0.6578	0.3990	0.8458	-0.8483	-0.7304
LCOM4	0.7317	0.7381	0.3882	0.8640	-0.8038	-0.0958
NOC	0.7430	0.4754	0.4008	0.8509	-0.8460	-0.6354
RFC	0.7642	0.4912	0.4001	0.7214	-0.7503	-0.3836
COF	-0.8272	-0.6645	-0.6489	-0.6835	0.8559	0.2635
WMC	0.4878	-0.4929	-0.0764	0.9712	0.5377	0.6081
Cognitive c.	0.7760	0.7924	0.4333	0.7639	-0.7718	0.4203
Struct.complexity	0.7264	0.8319	0.3751	0.8587	-0.7403	-0.0713

Correlation between in degree distribution γ_{in} and software

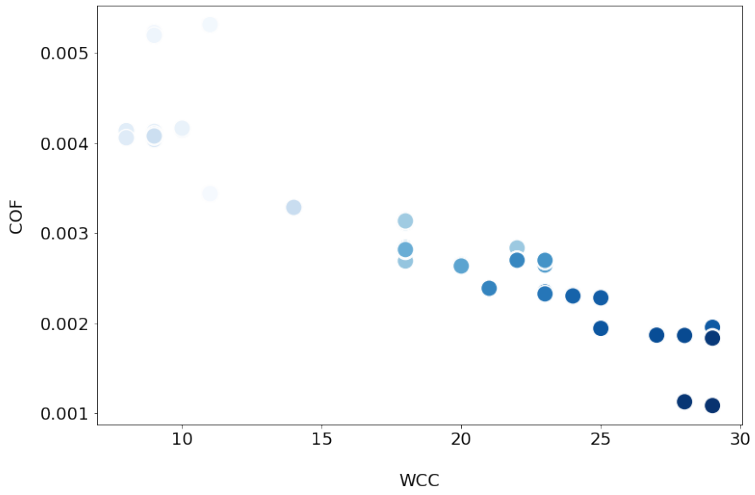
γ_{in}	Cassandra	Chukwa	Hadoop	Http-Components	IVY	Jena
Avg. cc	-0.1609	0.6808	0.2844	-0.2551	-0.3550	-0.0193
CBO	-0.1611	0.6082	0.3020	-0.1584	-0.4431	-0.0462
DIT	-0.1566	0.5539	0.2944	-0.1211	-0.3485	0.0878
LCOM4	-0.1660	0.6867	0.2576	-0.0804	-0.4393	-0.1142
NOC	-0.1569	0.4711	0.3040	-0.1063	-0.3649	0.0751
RFC	-0.1650	0.4916	0.2877	-0.2168	-0.4956	-0.0327
COF	0.1977	-0.6117	-0.2644	0.2009	0.3565	0.0309
WMC	-0.1554	-0.2457	-0.3513	0.2644	-0.5226	0.0184
Cognitive c.	-0.1248	0.5807	0.2743	-0.1988	-0.4261	-0.2753
Struct.complexity	-0.1738	0.6210	0.2835	-0.0339	-0.5042	-0.1403

Appendix D Scatter plots

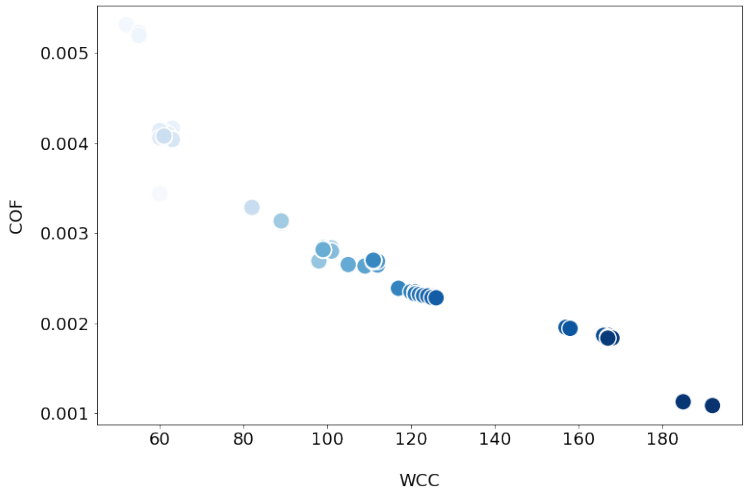
Cassandra system



Call network

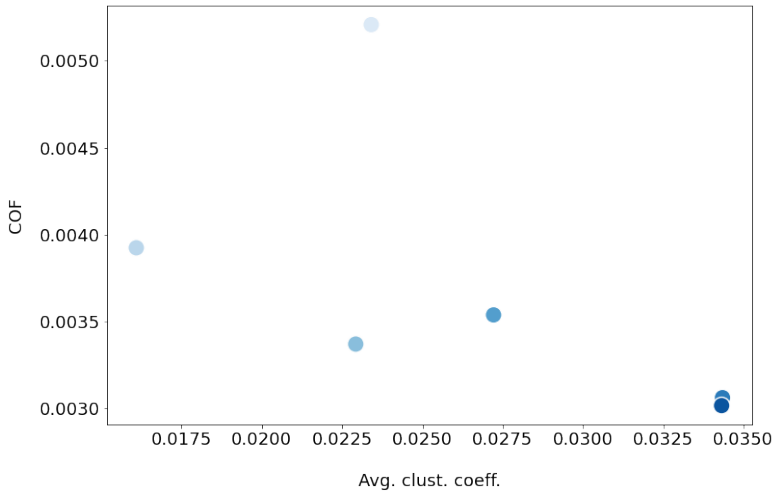


Collaboration network

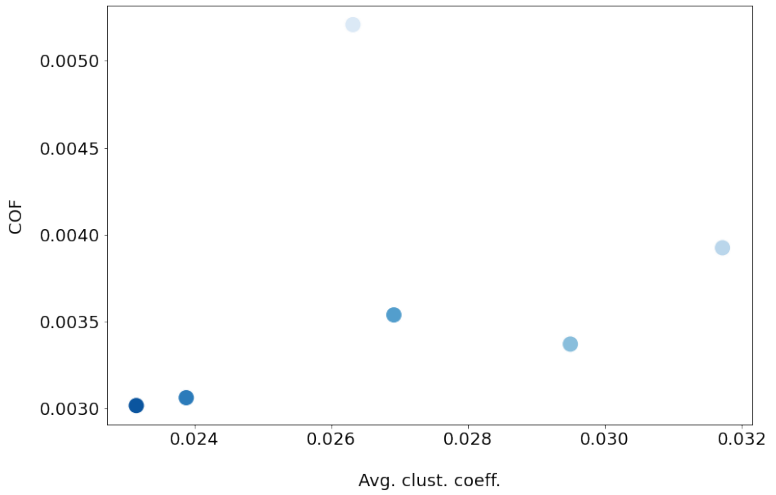


Inheritance network

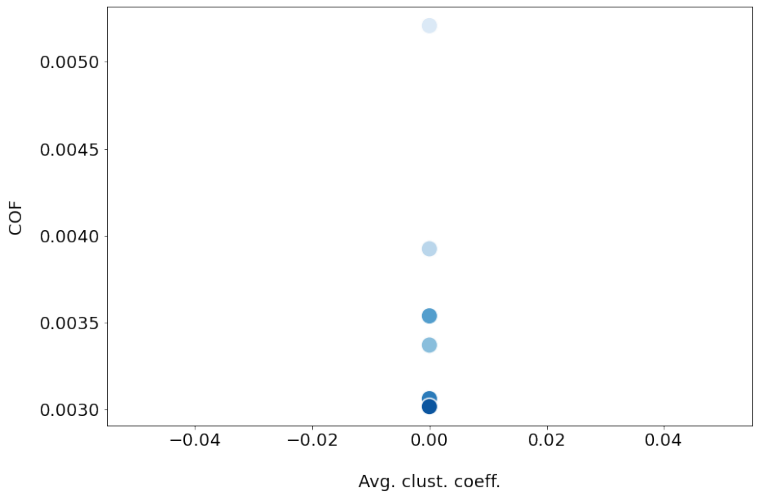
Chukwa system



Call network

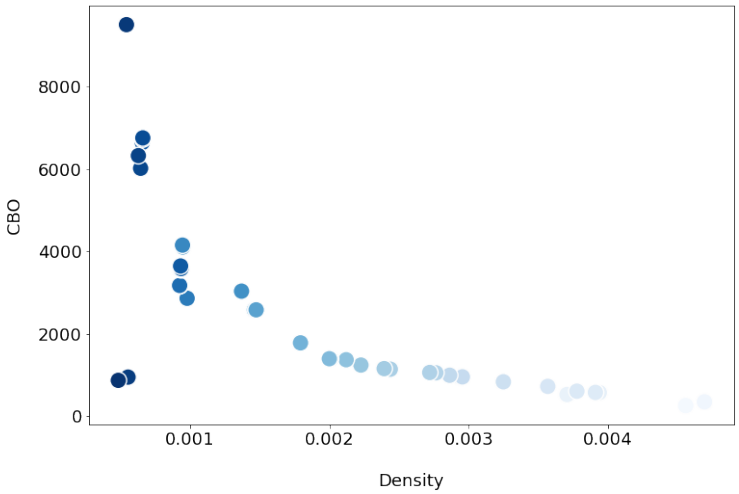


Collaboration network

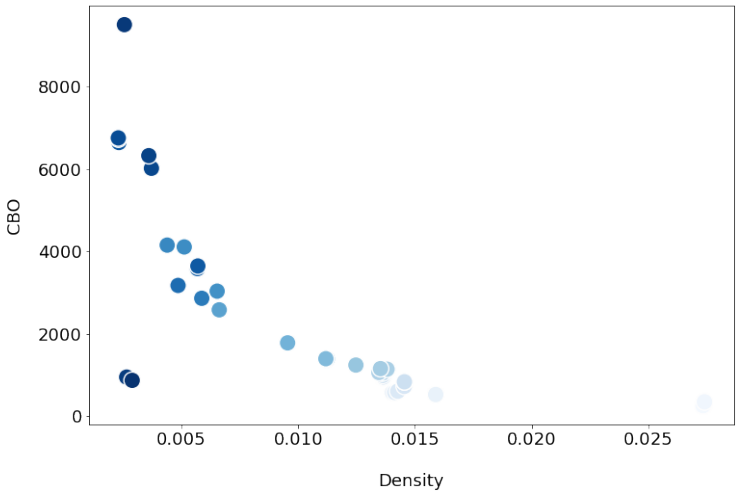


Inheritance network

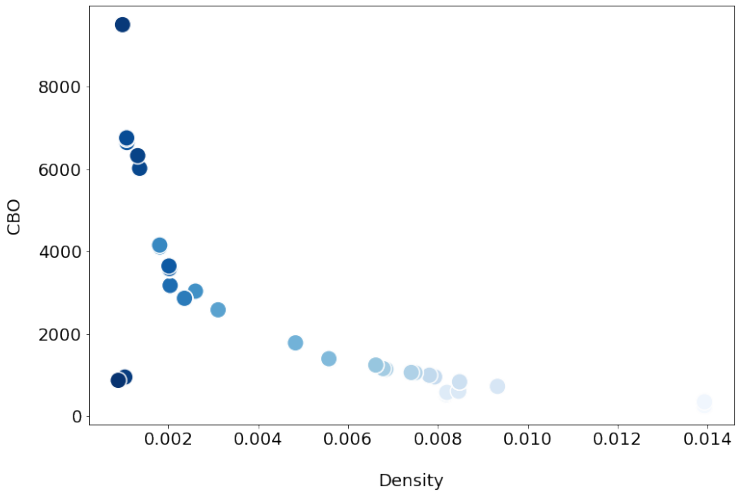
we can see that the clustering coefficient did not change and is equal to zero indicating that this is a bipartite graph.



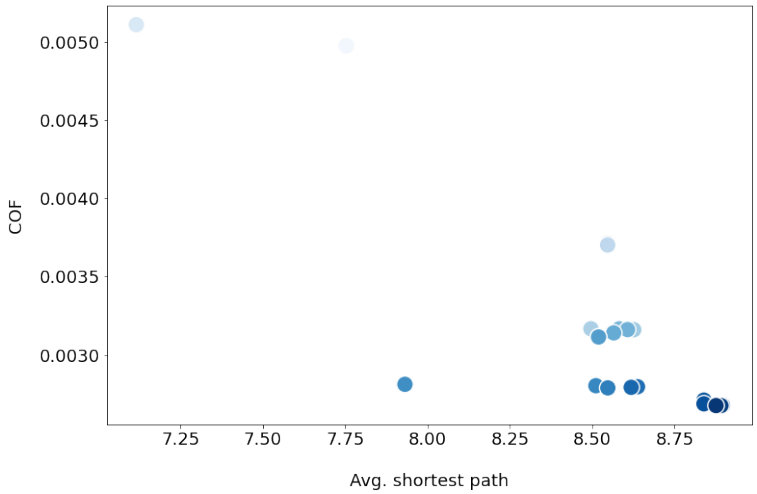
Call network



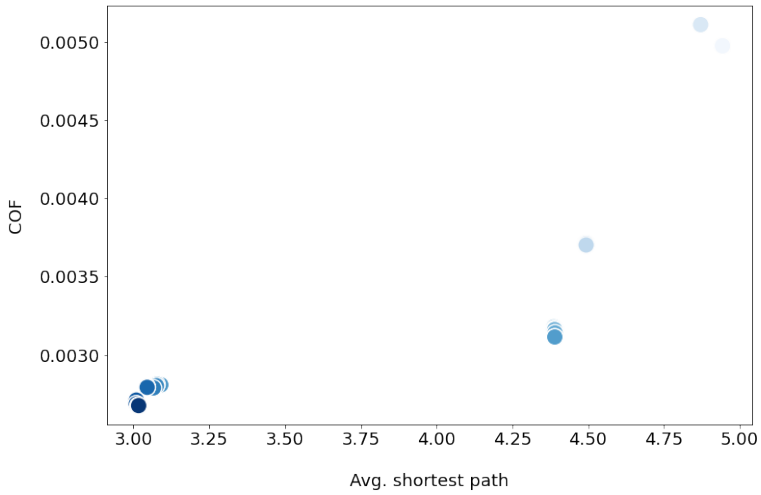
Collaboration network



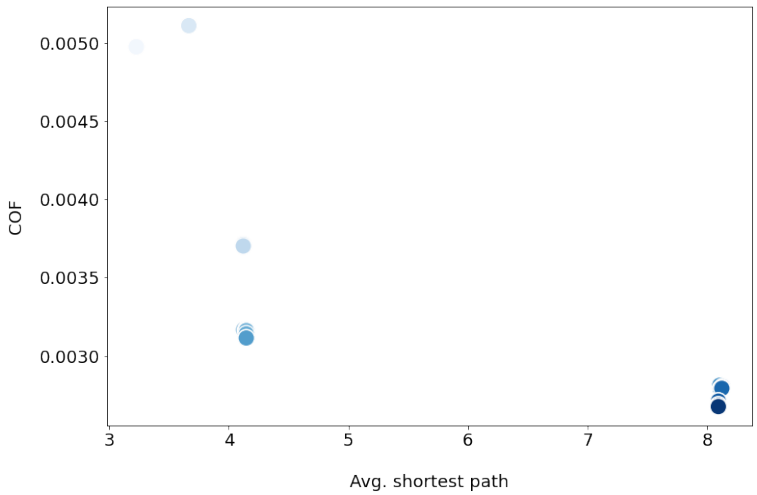
Inheritance network



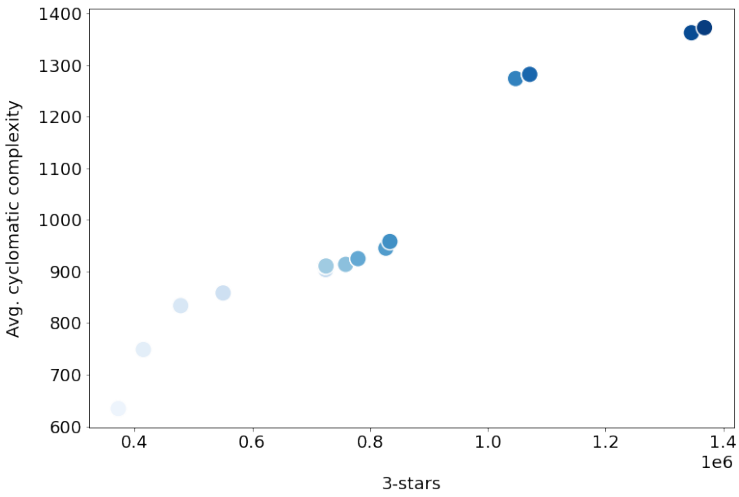
Call network



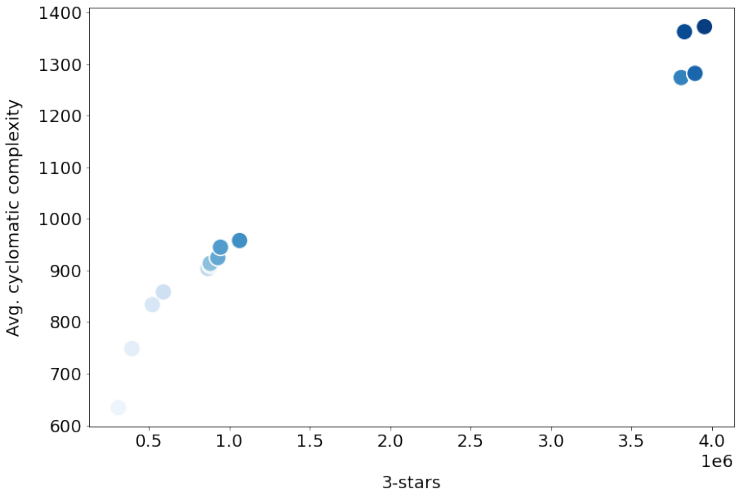
Collaboration network



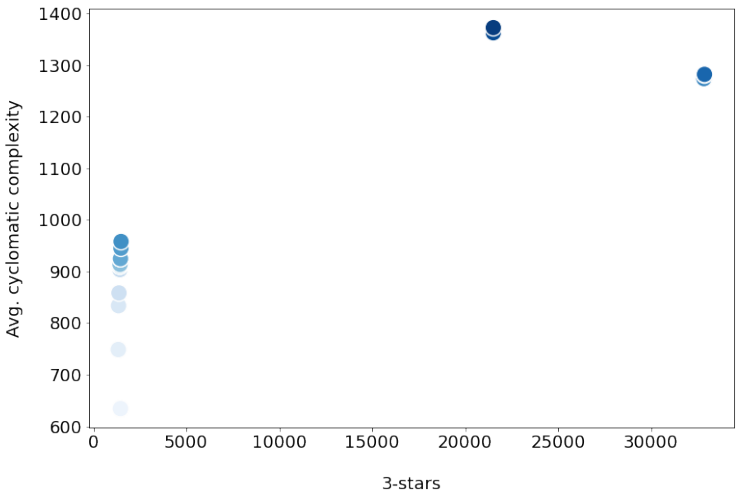
Inheritance network



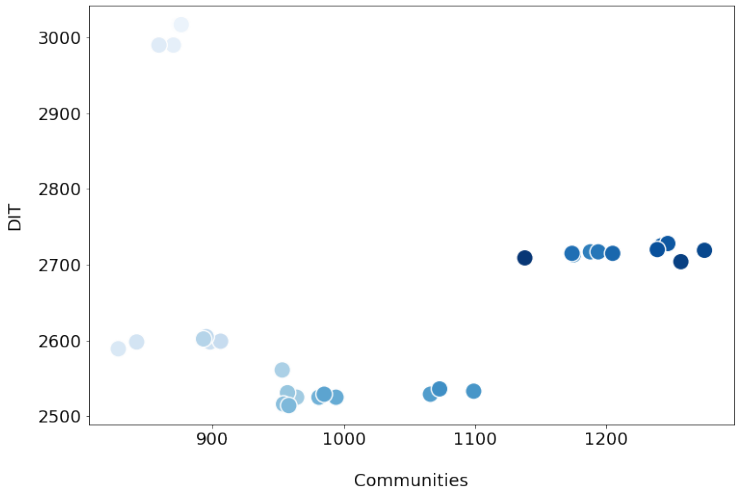
Call network



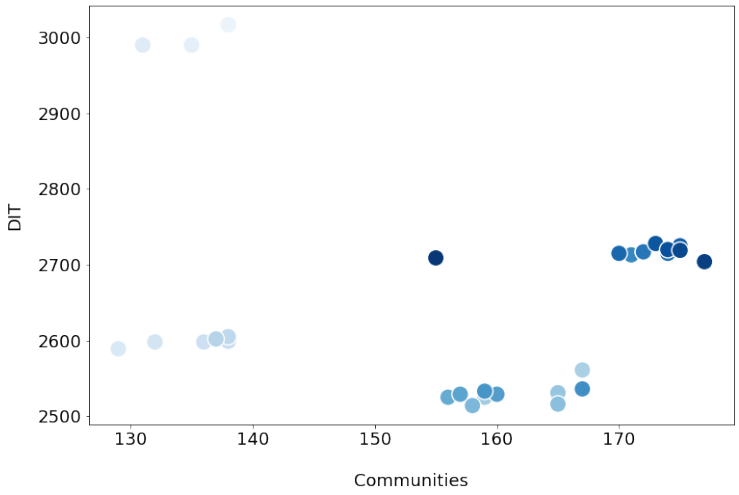
Collaboration network



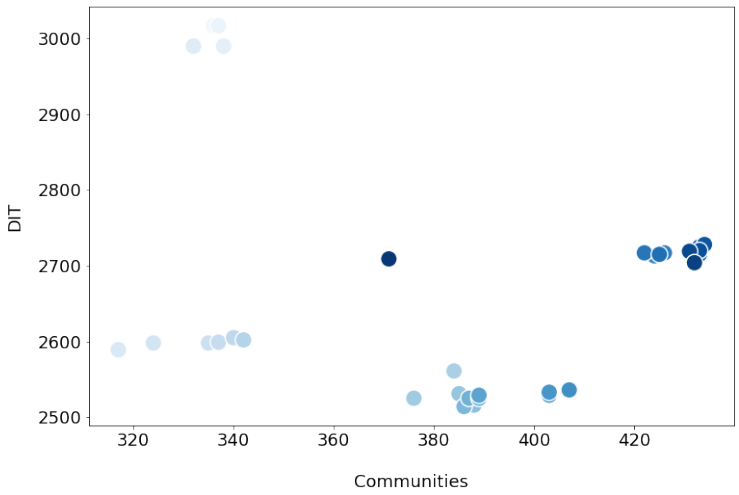
Inheritance network



Call network



Collaboration network



Inheritance network