

# Network Analysis of Software Repositories

## *The Eclipse Bugzilla Case*

Monika Schubert,

Michel Wermelinger, Yijun Yu

Knowledge Management Institute  
Graz University of Technology

Department of Computing  
The Open University

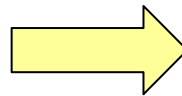
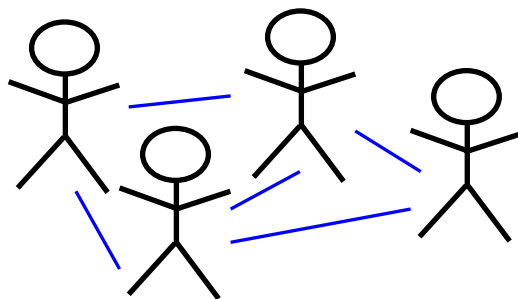
[monika.schubert@tugraz.at](mailto:monika.schubert@tugraz.at)

Milton Keynes, UK

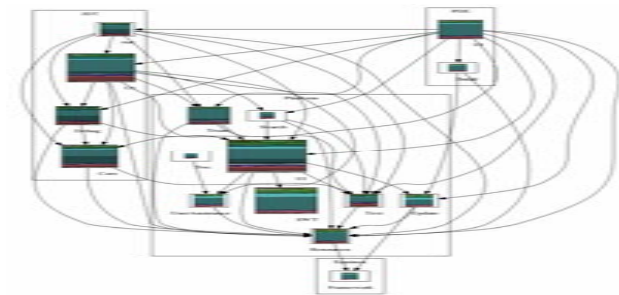
# Conway's Law

“Any organization that designs a system will inevitably produce a design whose structure is a copy of the organization's communication structure” [Con1968]

## Community



## Software Architecture



# Research Questions

- How can we infer **social structure** and **hierarchies** among software engineers from open source software repositories?
- Is there a correlation between the **social** and the **technical aspects** of software development?

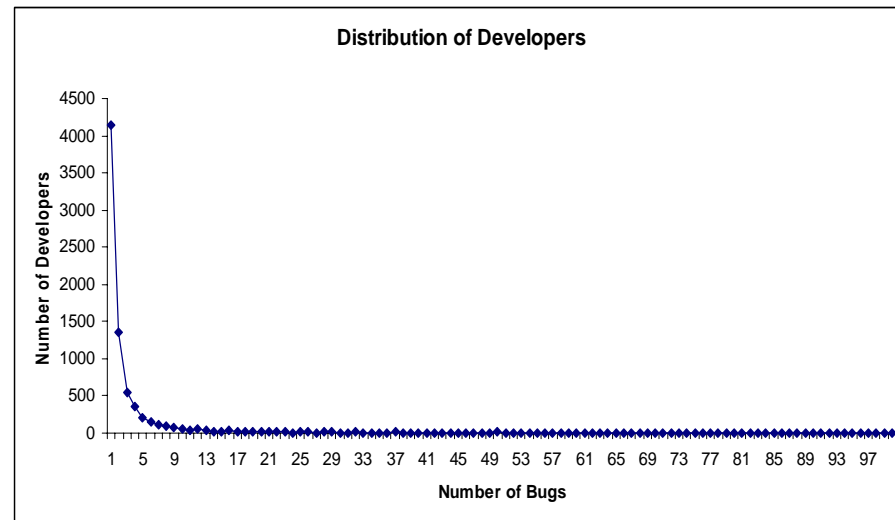
# Eclipse Bugzilla Dataset

Data provided:

- Software component
- Reporter
- Assignee
- Discussants

	Total	SDK
Number of Bugs:	207743	101966
Number of Developer:	25741	16025
Number of Components:	662	18

#Bugs	#Developer
1	4134
2	1356
3	544
4	350
...	...
10	60
...	...



## Related Work

- Work on Conway's Law
  - Analysing the structure of organizations and products of scientific computing projects [Ara2008]
- Work on the Eclipse Bugzilla dataset
  - Bug Prediction [Jos2007]
  - Forecasting the number of changes [Her2007]
  - Author–Topic Modelling [Lin2007]
  - Fixing time of a Bug [Wei2007]

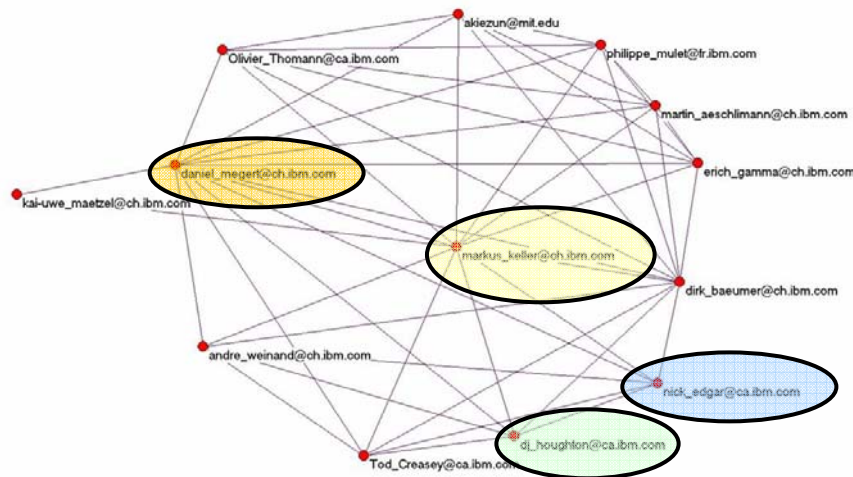
# Analysis Concepts

- Analysis of Community
  - Folding, Cooccurrence [Was1994]
  - Formal Concept Analysis [Wil2005]
- Analysis of the Architecture
  - Static and Dynamic Dependencies
- Correlations
  - Degree Centrality
  - Centrality Rank [Spe1987]

*In cooperation with  
The Open University*

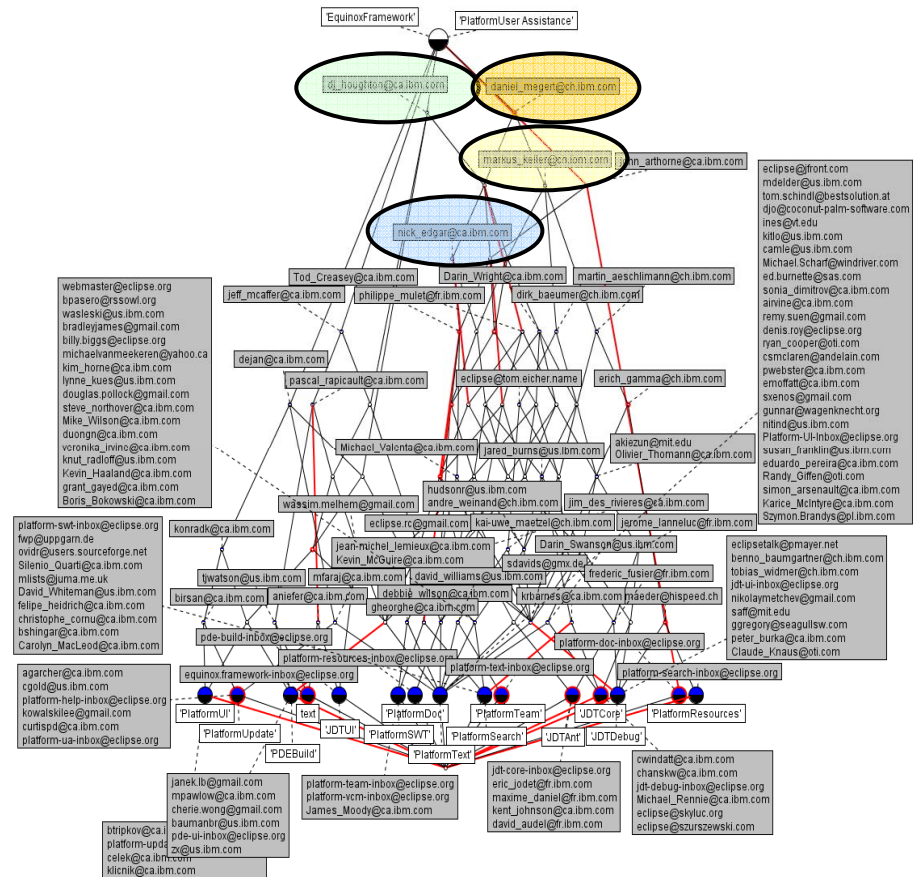
# Social Structure and Hierarchies

- Single entity dominance
- Geographic clustering

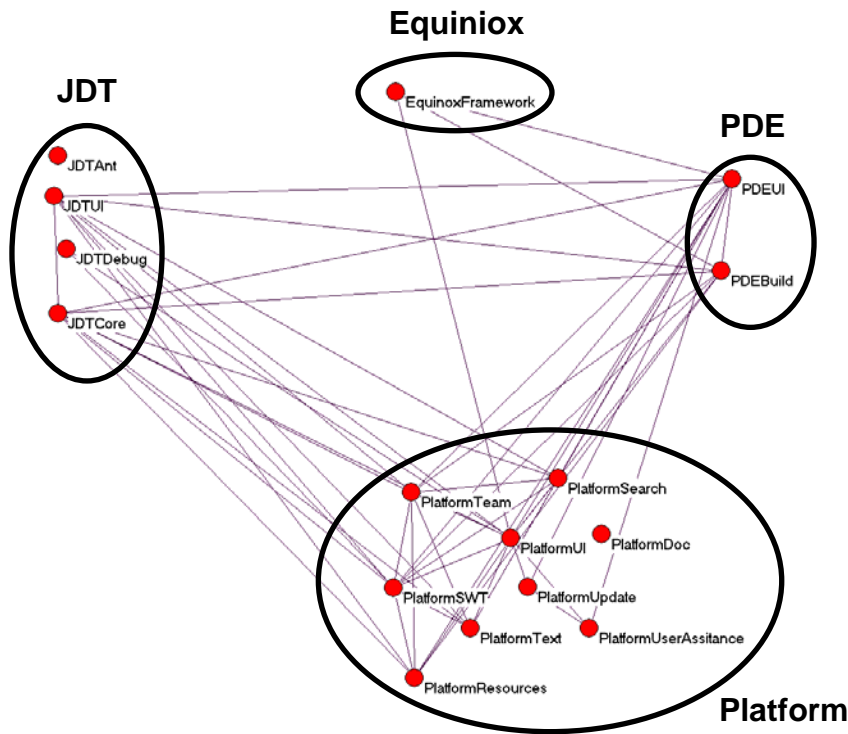


Network of Developers

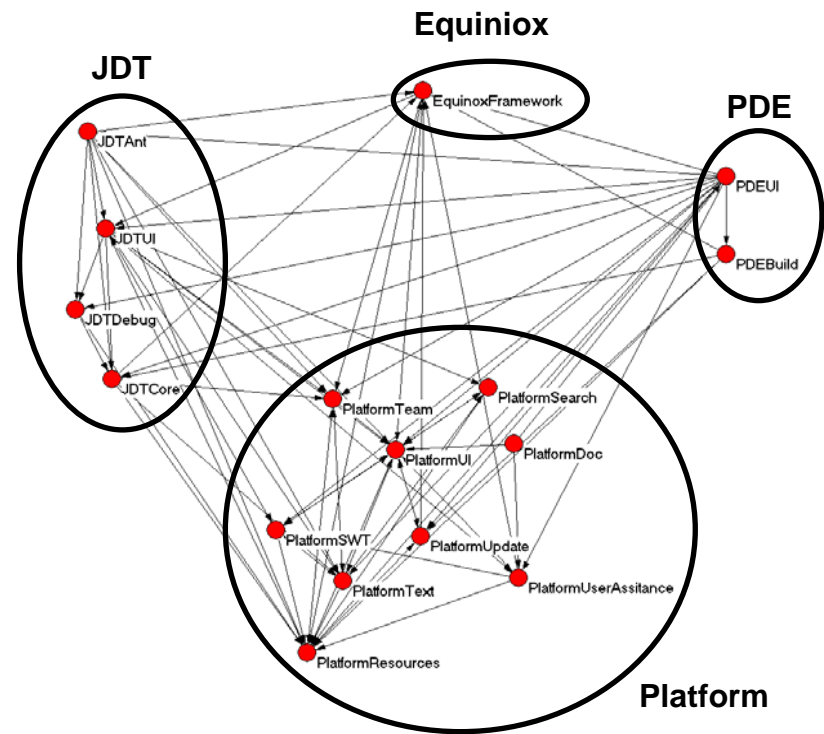
Created by folding a Component-Developer Graph



# Social vs. Technical Aspects

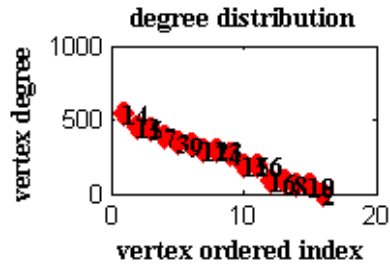


Network of Components created by folding a Component-Developer Graph K=256

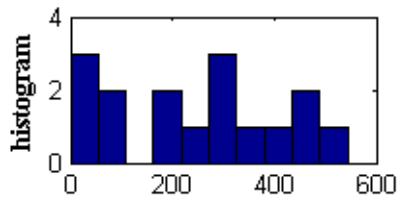


Connections between components from the architecture

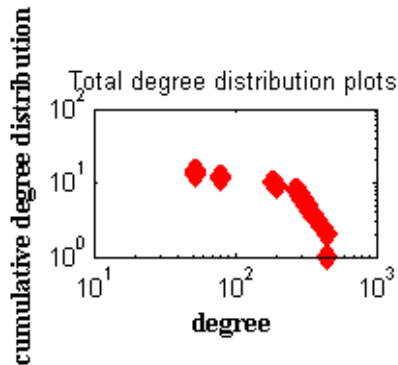
# Degree Distribution



Degree Distribution: ranking according to the degree of each node



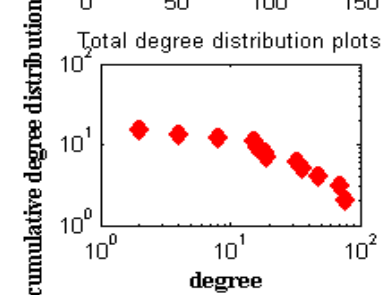
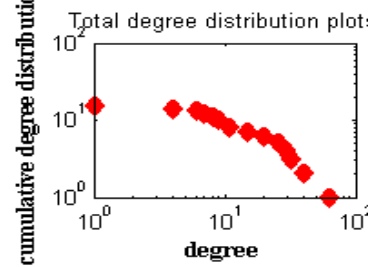
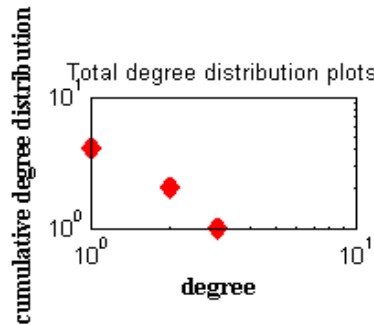
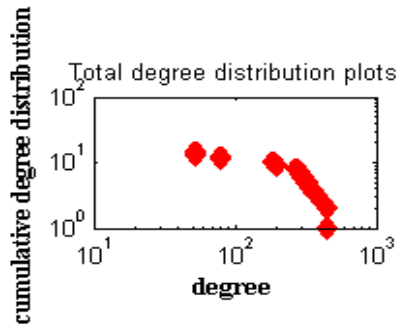
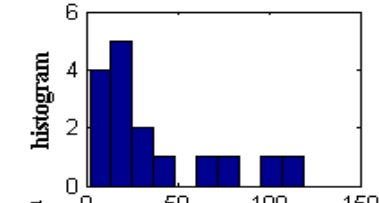
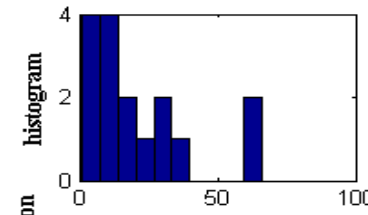
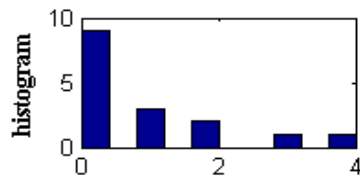
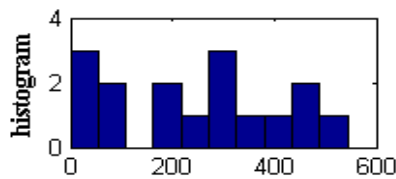
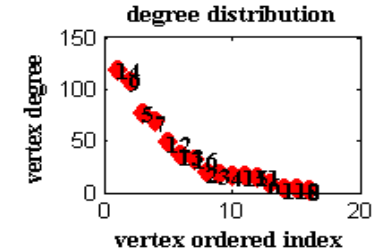
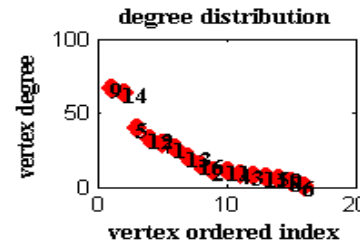
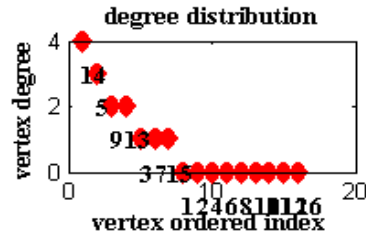
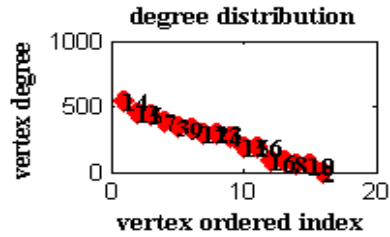
Histogramm: clustering nodes to degree intervals



Total Degree Distribution: cumulative degree distribution for a given number

Social inferred component network with k=32

# Degree Distribution



k=32

k=1024

static undir.

dynamic undir.

# Rank Correlation

k=32		k=1024		static undirected		dynamic undirected	
<u>PlatformUI</u>	1	<u>PlatformUI</u>	1	<u>PlatformUI</u>	1	<u>PlatformResources</u>	1
PlatformSWT	2	JDTUI	2	<u>PlatformResources</u>	2	<u>PlatformUI</u>	2
JDTUI	3	<u>PlatformResources</u>	3	JDTUI	3	JDTUI	3
PDEUI	4	PlatformText	3	PDEUI	4	PlatformTeam	4
JDTCore	5	JDTCore	5	PlatformTeam	5	PDEUI	5
<u>PlatformResources</u>	6	PDEUI	5	PlatformText	6	EquinoxFramework	6
PlatformTeam	7	PlatformUpdate	5	PlatformUser Assistance	7	PlatformText	7
PlatformText	8	EquinoxFramework	8	JDTAnt	8	PlatformUser Assistance	8
JDTDebug	9	JDTAnt	8	JDTCore	9	JDTAnt	9
PlatformUpdate	10	JDTDebug	8	JDTDebug	10	PlatformSWT	9
PlatformUser Assistance	11	<u>PDEBuild</u>	8	PlatformUpdate	10	JDTDebug	11
EquinoxFramework	12	PlatformDoc	8	PlatformSWT	12	JDTCore	12
<u>PDEBuild</u>	13	PlatformSearch	8	<u>PDEBuild</u>	13	PlatformUpdate	13
PlatformDoc	14	PlatformSWT	8	EquinoxFramework	14	PlatformSearch	14
PlatformSearch	15	PlatformTeam	8	PlatformSearch	14	PlatformDoc	15
JDTAnt	16	PlatformUser Assistance	8	PlatformDoc	16	<u>PDEBuild</u>	16

# Rank Correlation

Spearman: 
$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

- Compared all different *social-inferred* and *code-inferred* graphs with each other

## Results:

- up to 0.7368 correlation
- between k=1024 and static undirected

- Compared the *social-inferred* with *random* graphs

## Results:

- Up to 0.1114 correlation

# Contributions

- Provide a large-scale study of the relationship between social systems and the software architecture
- Exploring evidence that speaks for and/or against Conway's Law

# Discussion Points

- Conway's law is incomplete
  - What is a communication structure?
  - What is the structure of a product or source code?
- Degree centrality versus graph structure
  - The degree centrality is an indication of the importance of a node
  - The graph structure is represented by the edges
- Rank correlation
  - How do the tied ranks effect the interpretation?

# Monika Schubert

Graz, University of Technology

[monika.schubert@tugraz.at](mailto:monika.schubert@tugraz.at)

# References

- [Con1968] Conway M.E. (1968). **How do committees invent. Datamation**, (14)4:28—31.
- [Her2007] Herraiz, I.; Gonzalez-Barahona, J. M.; Robles, G. (2007), **Forecasting the number of changes in Eclipse using time series analysis**, in 'Proceedings of the 29th International Conference on Software Engineering Workshops', IEEE Computer Society.
- [Jos2007] Joshi, H.; Zhang, C.; Ramaswamy, S.; Bayrak, C. (2007), **Local and Global Recency Weighting Approach to Bug Prediction**, in 'Proceedings of the Fourth International Workshop on Mining Software Repositories', IEEE Computer Society.
- [Lin2007] Linstead, E.; Rigor, P.; Bajracharya, S.; Lopes, C.; Baldi, P. (2007), **Mining Eclipse Developer Contributions via Author-Topic Models**, in 'Proceedings of the Fourth International Workshop on Mining Software Repositories', IEEE Computer Society.
- [Was1994] Wasserman S.; Faust K. (1994). **Social Network Analysis: Methods and Applications**. Cambridge University Press.
- [Wei2007] Weiss, C.; Premraj, R.; Zimmermann, T. & Zeller, A. (2007), **How Long will it Take to Fix This Bug?**, in Harald Gall & Michele Lanza, ed., 'Proceedings of the Fourth International Workshop on Mining Software Repositories'.
- [Wil2005] Wille R. (2005). **Formal Concept Analysis as Mathematical Theory of Concepts and Concept Hierarchies**. Formal Concept Analysis, 1--33, 2005.