

Vertical Interaction in Open Software Engineering Communities

Patrick Wagstrom
Ph.D. Thesis Defense
March 9, 2009

Committee:
James Herbsleb
Kathleen Carley
M. Granger Morgan
Audris Mockus



open source



Open Source is BIG Business

<i>Year</i>	<i>Target</i>	<i>Buyer</i>	<i>Amount</i>
2008	MySQL	Sun	\$1 billion
2008	Trolltech	Nokia	\$153 million
2007	Zimbra	Yahoo!	\$350 million
2007	XenSource	Citrix	\$500 million
2006	JBoss	RedHat	\$350 million
2003	SuSE	Novell	\$210 million
1999	Cygnus	RedHat	\$675 million

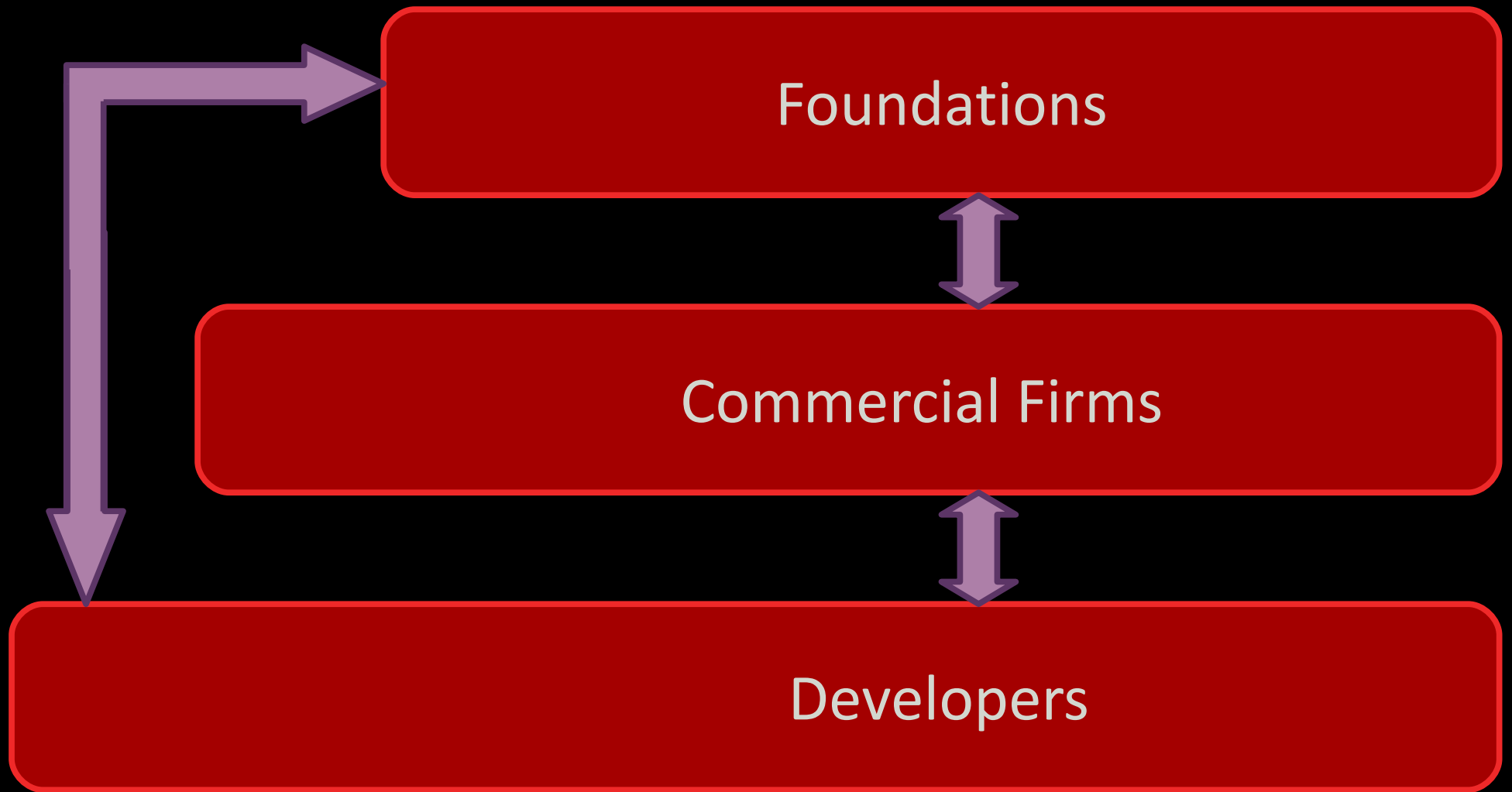
Open Communities are Bigger



From March 2008 Eclipse Executive Director's Report:

<http://www.eclipse.org/org/foundation/membersminutes/20080317MembersMeeting/DirectorsReport.pdf>

Central Players In Open Source



4 Empirical Studies

- Firms and Foundations
- Firms and Firms
- Firms and Individuals
- Individuals and Individuals

Firms and Foundations: Guiding an Ecosystem to Promote Value

The Problem

- Some research has been done about why individual focused OSS projects utilize foundations
- Little research has addressed why commercial firms would participate in foundations
 - Large monetary cost
 - Giving up some control
 - Possibly increased work
- What does the foundation do to drive value?

Data

- Semi-structured interviews with Eclipse Foundation staff and employees of member companies
 - 38 interviews with 40 individuals
- Face-to-face meetings at EclipseCon 2007 and 2008
- Participation in Eclipse members meetings

Driving Value Creation

- Non-market player
- Introduction of process
- Value of the Eclipse brand and marketing
- Organizational structure driving value
- Platform for innovation

Non-Market Player

- Eclipse grew out of IBM's old VisualAge ecosystem
- Small firms had to worry about being stepped on
- Allows innovation without worry about “Gorillas”
- Opens the door for distribution based business models

Platform for Innovation

- Foundation actively recruits new members
- Encourages components to be as modular as possible
 - Modularity == Independence from other components
- Create projects outside of Eclipse and bring inside later
- Push usage outside traditional realms

Takeaways

- Eclipse Foundation has taken concrete steps to build ecosystem
- Governance structure ensures all can provide input
- Non-market nature is very beneficial
- Services provided for members are worth the cost

Firms and

Firms:

Business Collaboration
Through Open Source

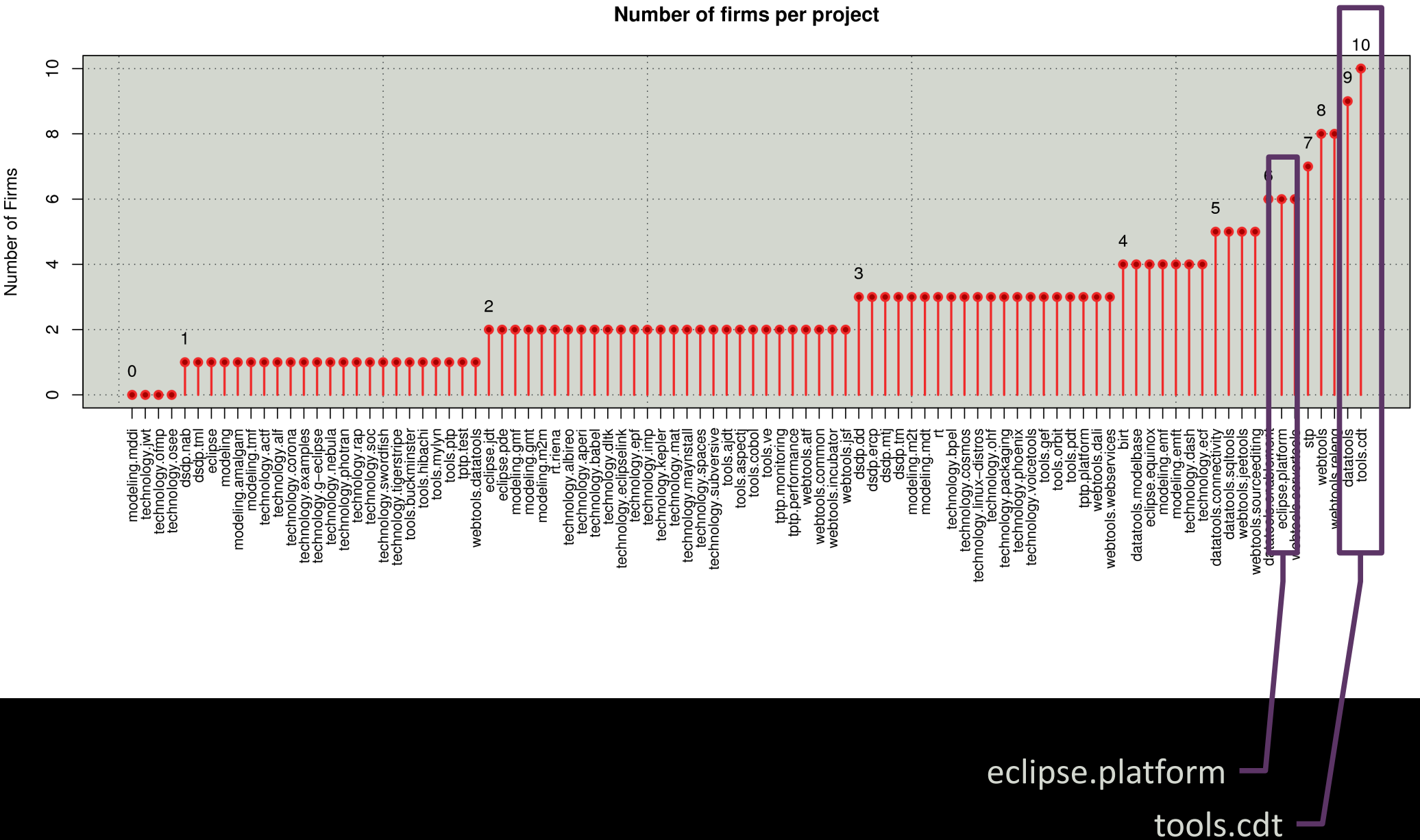
The Problem

- Much data about how individuals interact in OSS
- Little data about how firms collaborate
- Is there an overdependence on single firms?
- How collaborative are OSS ecosystems?

Data

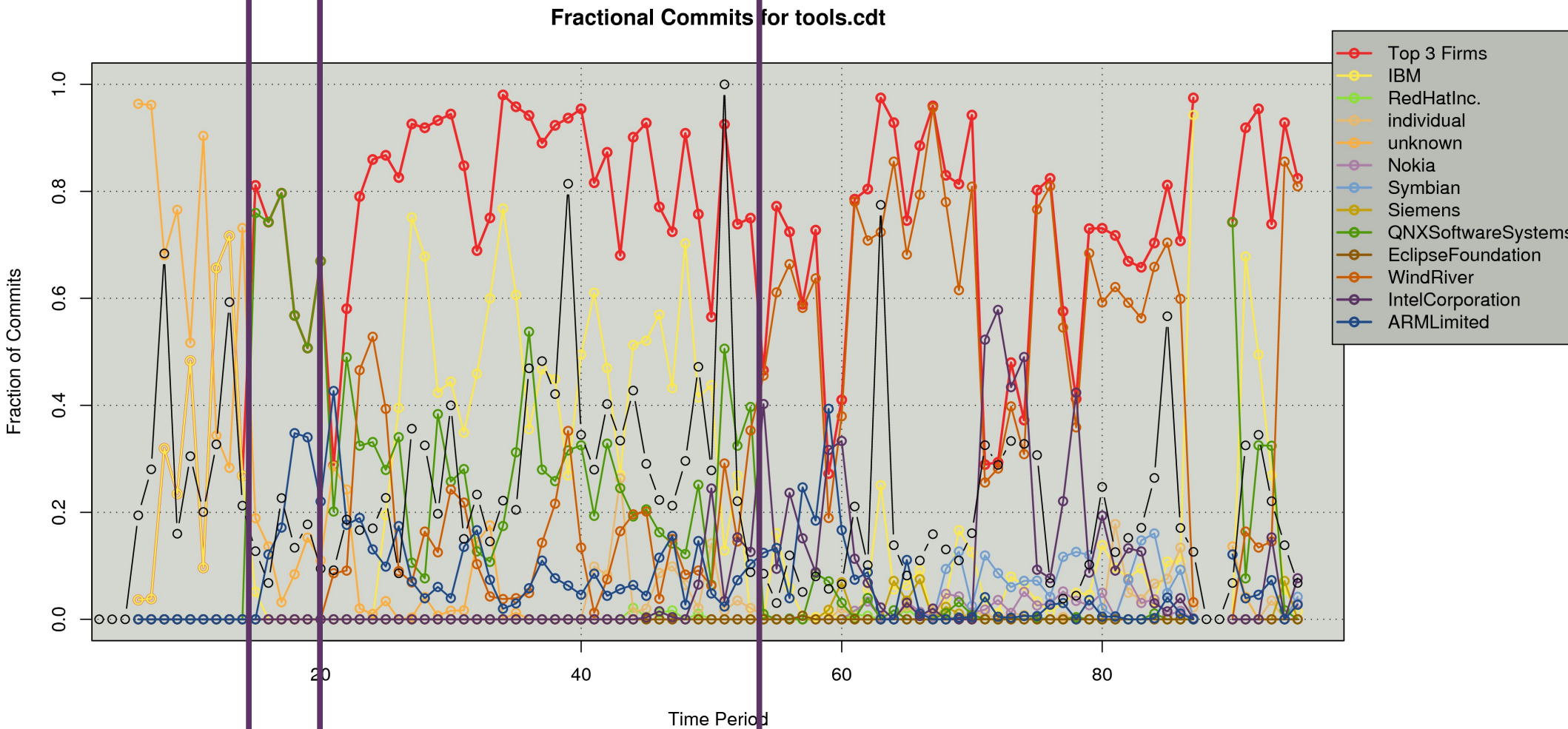
- Projects from Eclipse Foundation
- Two level project hierarchy
 - Top Level Projects (11)
 - Sub Projects (89)
- Collected data from version control system and IP repository
- Ties individuals to code changes and firms
- Compared with data from GNOME

How Much Collaboration Really Exists?



Collaboration in CDT

IBM Leaves/QNX Lead

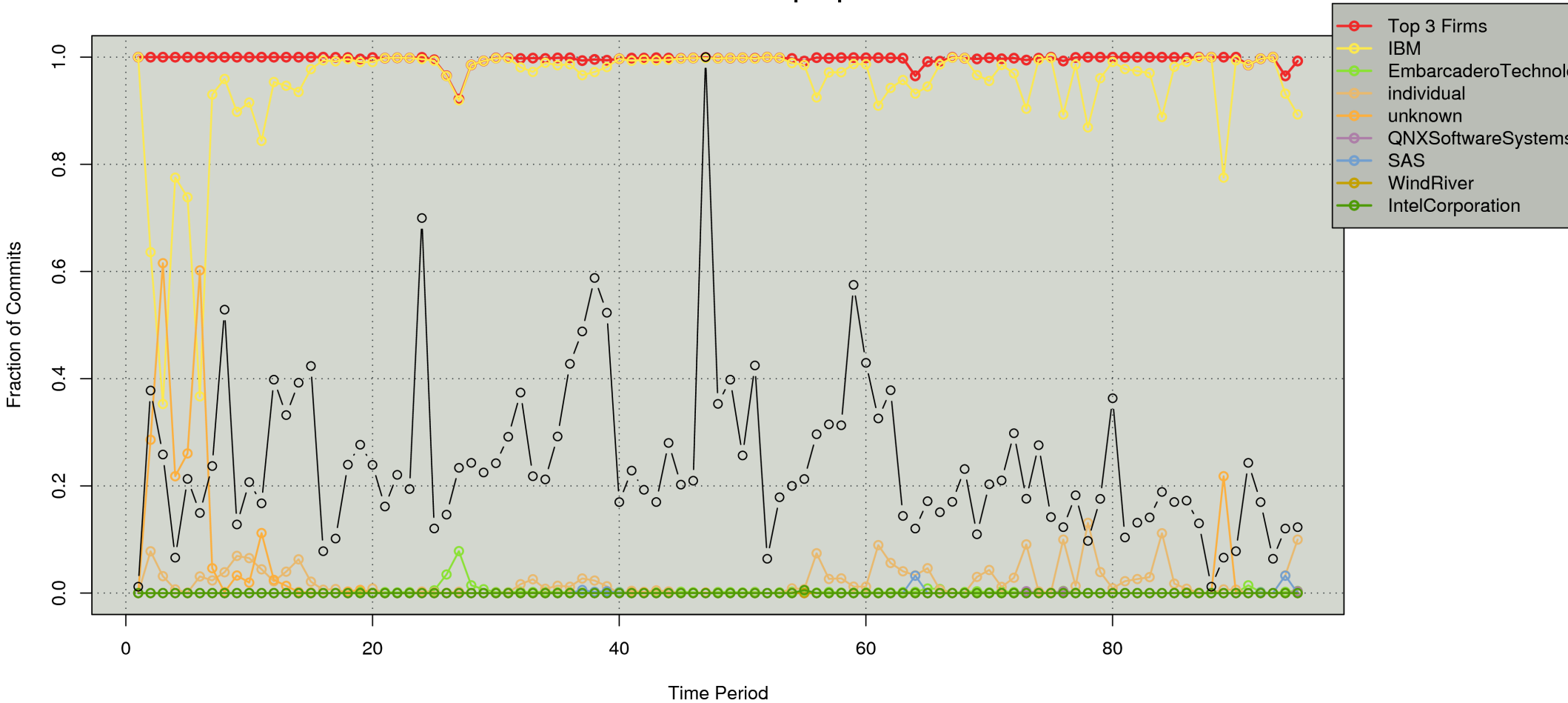


WindRiver Joins/IBM Lead

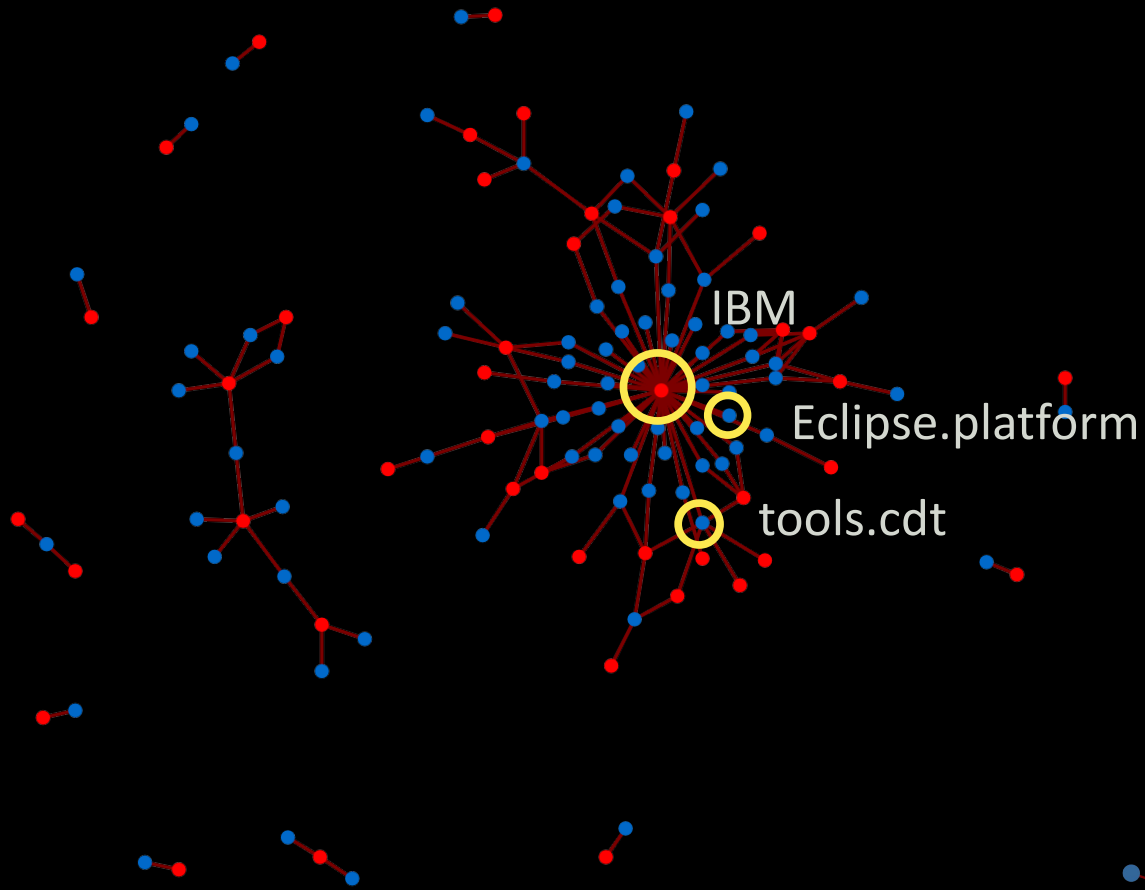
WindRiver Leads

Who Builds the Platform?

Fractional Commits for eclipse.platform



Community Network Structure

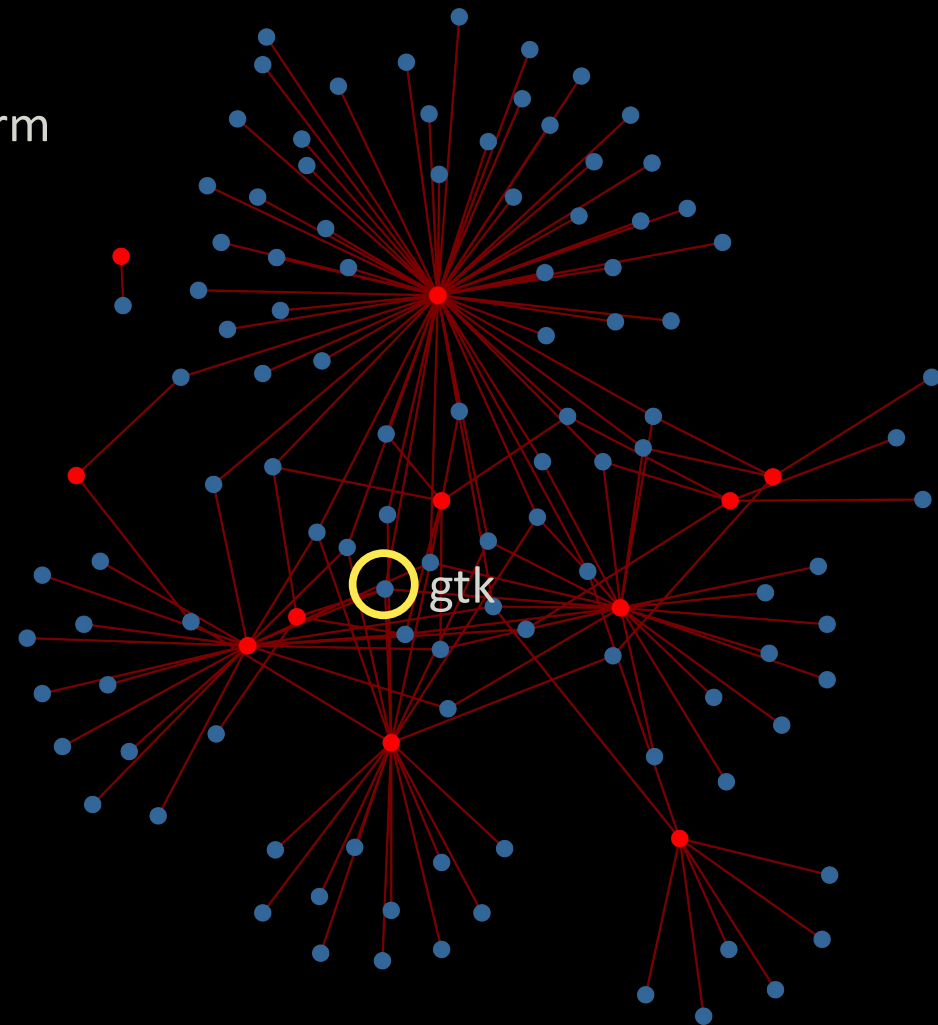


Eclipse

May 2008

GNOME

May 2005



Takeaways

- Participation in an OSS ecosystem may require little collaboration with other firms
- Many key portions of Eclipse are centered on IBM
- Allows IBM to exert great influence, even though no longer at the center
- The organic community around GNOME shows much more collaboration

Firms and Individuals:

The Impact of Commercial
Participation on Volunteer
Participation

The Problem

- Commercial firms have different interests than volunteer OSS developers
- Firms bring many resources to projects that benefit projects
- What impact do these firms have on volunteer participation?

Data

- Source code version control, bug tracker, and email lists from GNOME project
- Individuals are disambiguated and identities linked
- Commercial affiliation for developers identified
- Face to face interviews with 18 developers

Firm Classifications

- 9 major firms in community
- Divided into two categories -
 - Product focused
 - Community focused
- Validated through interviews
- Developers from community focused firms generally more active within the community

Do commercial developers drive away volunteers?

- Designed a multilevel model to predict current volunteers based on previous participation

$$VolDevs_{i,t} = \beta_0 + \beta_1 VolDevs_{i,t-1} + \beta_2 ComDevs_{i,t-1} + \beta_3 Commits_{i,t-1} + \nu_i + \epsilon_{i,t}$$

<i>Variable</i>	<i>Estimate</i>	<i>Std Error</i>	<i>P-Value</i>
<i>Intercept</i>	0.5643	0.1397	0.0001
<i>VolDevs</i>	0.4562	0.0442	<0.001
<i>ComDevs</i>	0.0817	0.0389	0.0360
<i>Commits</i>	0.0601	0.0242	0.0130

No! They actually have a slight positive impact on the number of volunteers!

Do commercial developers drive away volunteers (by firm)?

<i>Variable</i>	<i>Estimate</i>	<i>Std Error</i>	<i>P-Value</i>
<i>Intercept</i>	0.6032	0.1381	<0.001
<i>VolDevs</i>	0.4212	0.0443	<0.001
<i>ComDevs(CF)</i>	0.2050	0.0432	<0.001
<i>ComDevs(PF)</i>	-0.0433	0.0388	0.264
<i>Commits</i>	0.0711	0.0234	0.003

Developers at community focused firms have a significant attractive power while developers at product focused firms have no relation.

Takeaways

- Commercial firms do increase volunteer participation in Open Source
- Community focused firms have a much greater attractive power than product focused firms

Individuals and Individuals:

Evolution of the Socio-
Technical Congruence
Metric

The Problem

- STC hasn't been replicated in OSS
- Difficult to distill to individual level
 - Typically done at network level
 - Ratio muddles effects of coordination requirements and actual coordination
- Original analysis looked only at short term
 - Most software projects are long term

Data

- GNOME project
- Filtered for projects that had CVS, bug tracker, and mailing list archives
- Do not have as much developer information as Cataldo et. al.
- Examine time to resolve bugs
 - Only include those bugs marked as defects

Individualized STC

$$\frac{\sum (C_A \wedge C_R)}{\sum C_R}$$

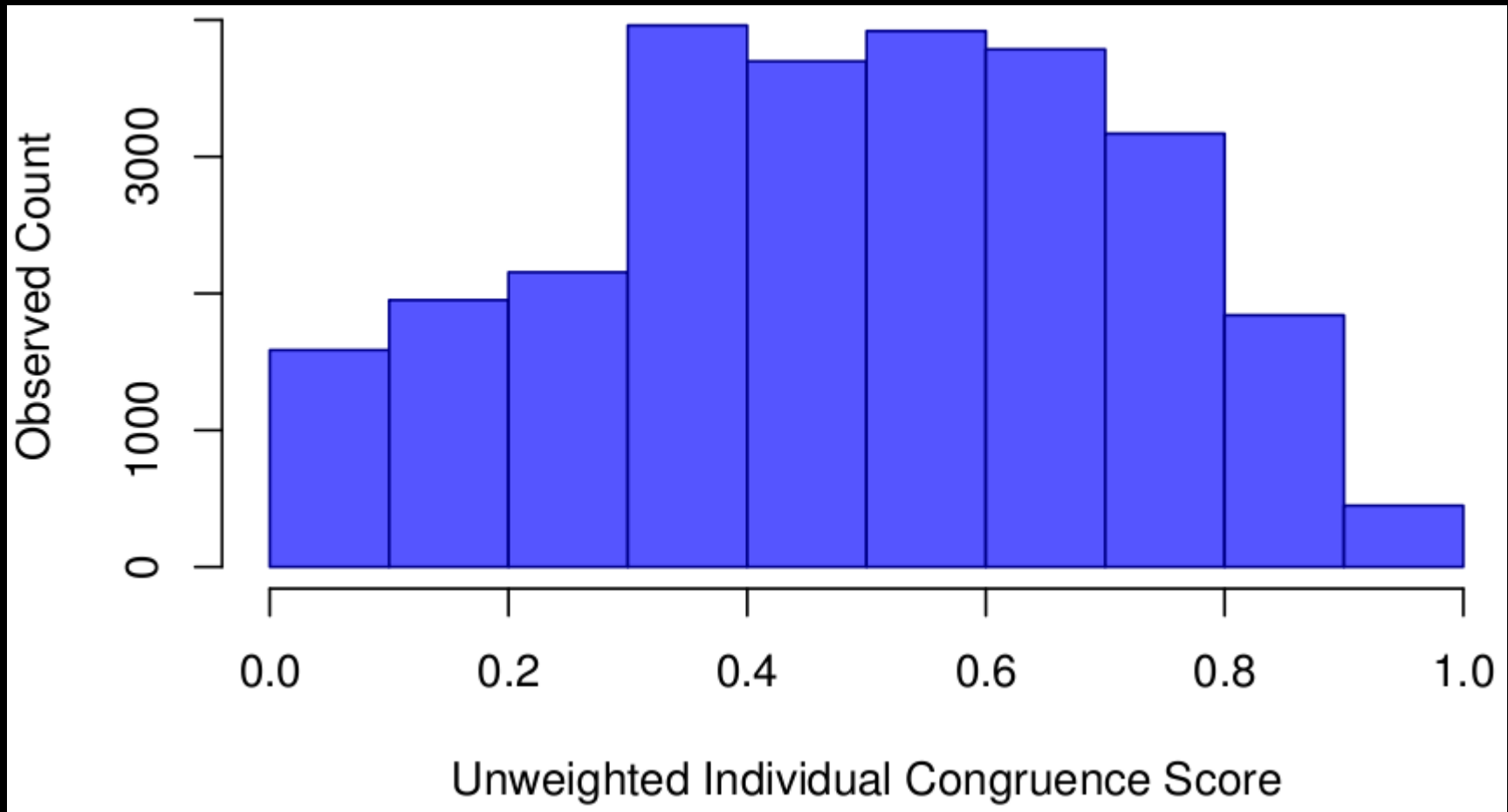
Proportion of coordination requirements that are mirrored in the actual communication network.

0	1	1	0	0	0	1	1	0	0	1	0
1	0	0	1	0	0	1	1	0	0	0	1
1	0	0	1	1	1	0	1	1	0	0	1
0	1	1	0	1	1	1	0	0	1	1	0
C_A				C_R							

$$\frac{6}{10} = 0.6$$

$$\frac{2}{4} = 0.5$$

Individualized STC



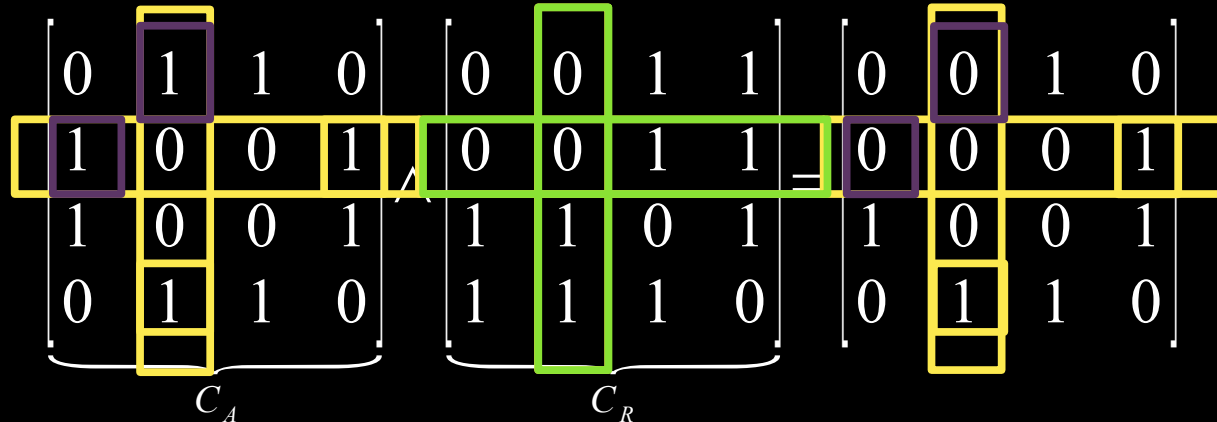
Testing Individualized STC

- Predict log2 of time to resolve defect
- Independent variables
 - Number of developers active on defect
 - Number of people changing defect status
 - Number of comments made
 - Individualized STC for developers

<i>Variable</i>	<i>Estimate</i>	<i>Std Error</i>	<i>P-Value</i>
<i>Intercept</i>	1.9707	0.0581	<0.0001
<i>NumDevs</i>	0.2846	0.0301	<0.0001
<i>DeltaPeople</i>	0.8074	0.0176	<0.0001
<i>Comments</i>	-0.0142	0.0036	<0.0001
<i>UIC</i>	-1.2140	0.0770	<0.0001

$R^2=0.134$, $DF=26507$, $p < 0.0001$

Disambiguating Results



Extra Communication Coordination Requirements Matched Communication

Variable	Estimate	Std Error	P-Value
<i>Intercept</i>	1.4590	0.0568	<0.0001
<i>NumDevs</i>	0.2500	0.0306	<0.0001
<i>DeltaPeople</i>	0.8020	0.0177	<0.0001
<i>Comments</i>	-0.0125	0.0036	0.0006
<i>MatchedComm</i>	-0.0524	0.0056	<0.0001
<i>CoordReq</i>	0.0314	0.0032	<0.0001
<i>extraComm</i>	-0.0119	0.0035	0.0006

$R^2=0.132$, $DF=26505$, $p < 0.0001$

Takeaways

- Demonstrated a method to individualize STC
- Should break apart STC metric into its constituent portions
- Extra communication, not related to coordination requirements, improves task performance

Conclusions

Building OSS Communities

- Not a matter of just throwing code out there
- Designating non-market player for head is helpful
- Need to find way to drive additional value to members, beyond just software
- Enable members to work independently
- Watch the centralization of components
- Invite firms to participate with volunteers
- Encourage discussion in the community

Thank You!

This work was supported in part by a National Science Foundation graduate research fellowship, the National Science Foundation (IIS-0414698), the IGERT Training Program in CASOS(NSF,DGE-9972762), the Office of Naval Research under Dynamic Network Analysis program (N00014-02-1-0973), the Air Force Office of Sponsored Research (MURI: Cultural Modeling of the Adversary, 600322), the Army Research Lab (CTA: 20002504), and the Army Research Institute (W91WAW07C0063) for research in the area of dynamic network analysis. Additional support was provided by CASOS - the center for Computational Analysis of Social and Organizational Systems at Carnegie Mellon University. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the the National Science Foundation, the Office of Naval Research, the Air Force Office of Sponsored Research, the Army Research Lab, or the Army Research Institute.

And more folks than I can fit on a single slide.

Thanks!