

Mining Email Social Networks

Christian Bird, Alex Gourley,
Prem Devanbu, Michael Gertz

目次

- 背景
- 目的
- 実験方法
- 実験結果
- まとめ

背景

- 大規模ソフトウェアプロジェクトではコミュニケーションや調整活動の観察が難しい
- OSSでは全ての活動をEmailなどのネットワークを用いて公的に行う
- メーリングリストには開発者同士のやり取りが残されている

目的

- 開発者のソーシャルネットワークの特徴は何か
- 活発な開発者はソースコード変更時でも活発であるか
- 開発者はソーシャルネットワークの開発以外の役割を果たすか

以上の問題解決を目指す

実験方法

- 1999年から2006年までのメーリングリストを対象
- 101637件のメッセージを解析
- 送信者, 送信時刻などを抽出

実験方法

- 同一人物でも違ったメールアドレスを使う人がいる

rasmus@apache.org

rasmus@bellglobal.com

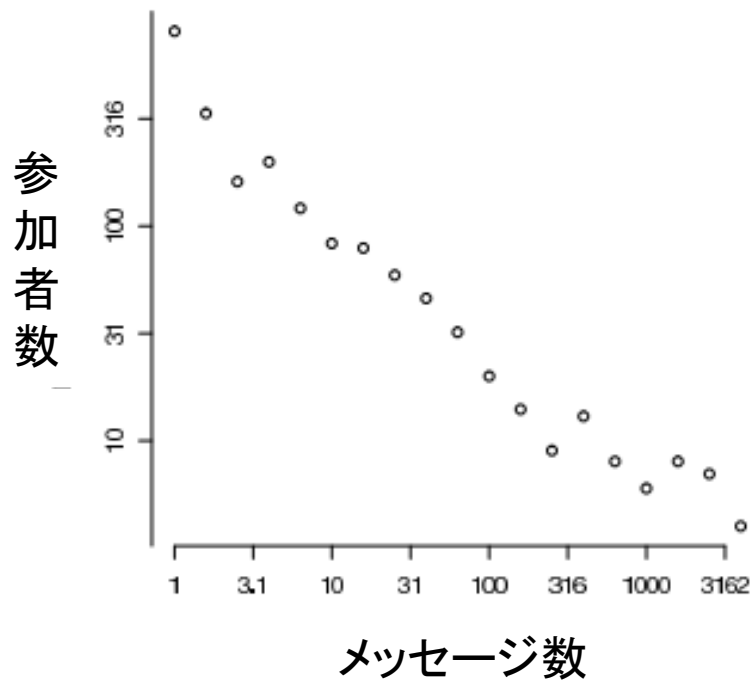
rasmus@lerdorf.com

rasmus@linuxcare.com

これらすべて同一人物として捉える

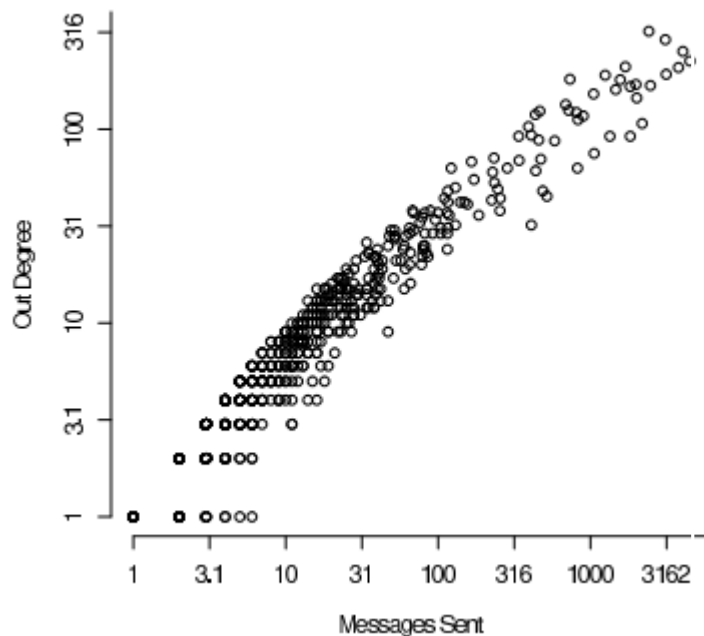
- 電子メールと名前が一致することがある
メールにファーストネーム, ラストネームが含まれている
場合, 同一人物としてみなす.

実験結果



- ✓ 大多数の開発者は1通しかメッセージを送信していない
- ✓ 一部の開発者は多数のメッセージを送信

実験結果



- ✓ Message Sent:
個人が送信したメッセージの数
- ✓ Out-Degree:
個人のメッセージに返答した
回答者数
- ✓ 強い相関があることがわかる
- ✓ スピアマンの順位相関係数=0.97

実験結果

	Developer	Non-Developer	T-value	Significance
Betweenness	0.0114	0.000140	5.07	$p < 0.001$
Out-degree	0.00666	0.000451	8.14	$p < 0.001$
In-degree	0.00794	0.000367	7.54	$p < 0.001$

Betweenness:メッセージを送信する際に仲介しているかどうかの値

In-degree:個人がメッセージを返答した回答者数

- T検定を行った結果, BetweennessからT値5.07と有意な結果が出た
- 開発者と非開発者で大きな差があることから, 開発者のほうがステータスが高いといえる

実験結果

	changes	srcChanges	docChanges	
changes	1	0.789	0.932	
srcChanges	0.789	1	0.514	
docChanges	0.932	0.514	1	
outdegree	0.520	0.712	0.308	
indegree	0.474	0.679	0.263	
betweenness	0.553	0.757	0.327	

ソースの変更とドキュメントの変更にあまり相関は見られない

→ 開発者が両作業に同じレベルに従事しているわけではないと予測

実験結果

	changes	srcChanges	docChanges	
changes	1	0.789	0.932	
srcChanges	0.789	1	0.514	
docChanges	0.932	0.514	1	
outdegree	0.520	0.712	0.308	
indegree	0.474	0.679	0.263	
betweenness	0.553	0.757	0.327	

ソース変更のbetweennessが一番高い

→ ソース変更作業において、メッセージを仲介することは一番多い

まとめ

- 大多数の開発者はメッセージをほぼ送信していない
- 個人が送信したメッセージとそれに返答してくれる回答者数には強い相関がある
- 開発者はいろいろな作業に従事しているとはいえない