# Social dynamics of FLOSS team communication across channels

Andrea Wiggins, James Howison & Kevin Crowston
Syracuse University School of Information Studies

9 September 2008 ~ IFIP 2.13 - OSS 2008

# Introduction

- FLOSS studies using social network analysis (SNA) techniques raise questions of validity
- Methodological issues for SNA on FLOSS
  - Choices of network measures
  - Evaluation of intensity of relationships
  - Effects of time
  - Variations across communication venues
- Building upon *Social dynamics of free and open source team communications* by Howison et al., from IFIP 2.13 in 2006

# FLOSS Networks

- Membership/association networks
  - Developer-project networks, not the focus of this paper
- Communication networks
  - Network based on reply structure of public threads
    - Proxy for direct communication between individuals, leaving external validity questions aside…
  - Link formed between a replier and the immediately previous poster in a threaded discussion
    - Example: Andrea starts a thread, James replies to Andrea, and Kevin replies to James
    - (Directed) links: J -> A, K -> J

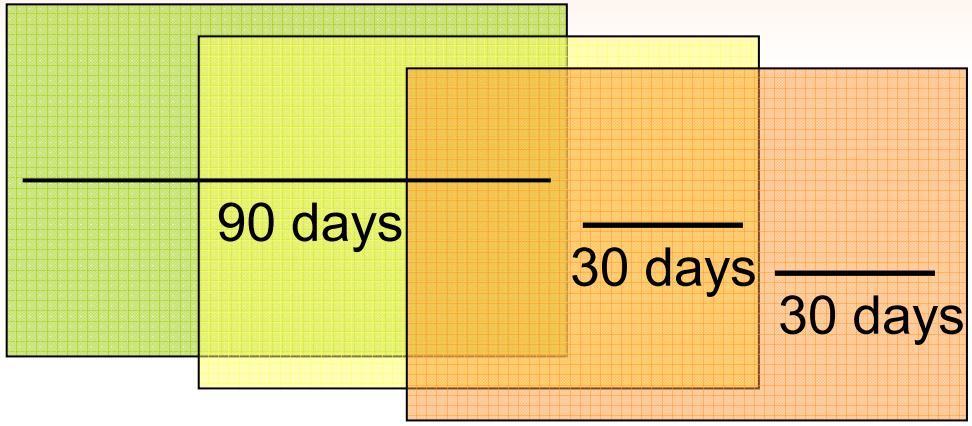
# Methodological Issues

- Choice of measures
  - Are the measures appropriate to the data?
  - Example: broadcast network (e.g. email list) violates assumptions of brokerage measures based on control of information flow
- Time
  - Aggregation is necessary but masks informative details
- Intensity of interactions
  - Most SNA measures are binary
- Variations across venues
  - A significant issue for FLOSS SNA study sampling

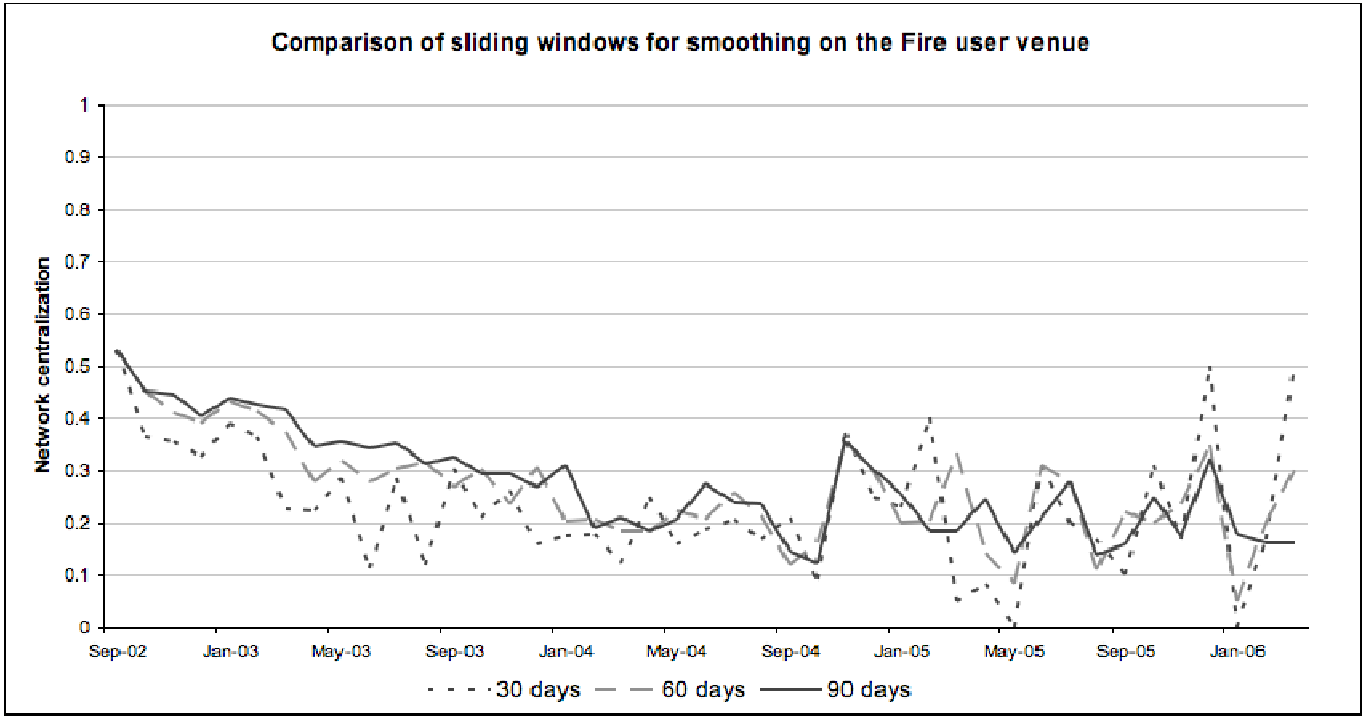
# Measures and Time

- Measures
  - Outdegree centralization: whole network measure of inequality in communication contribution
  - Outdegree (centrality): number of outbound links in a directed network, used for reply-to structure of threads
  - Centralization: relationship between all centralities in the network
    - High values: a few individuals make most responses
    - Low values: more equal communication levels
- Time
  - Series of snapshots of network, with sliding window to handle low-volume time periods

# Sliding Windows and Smoothing



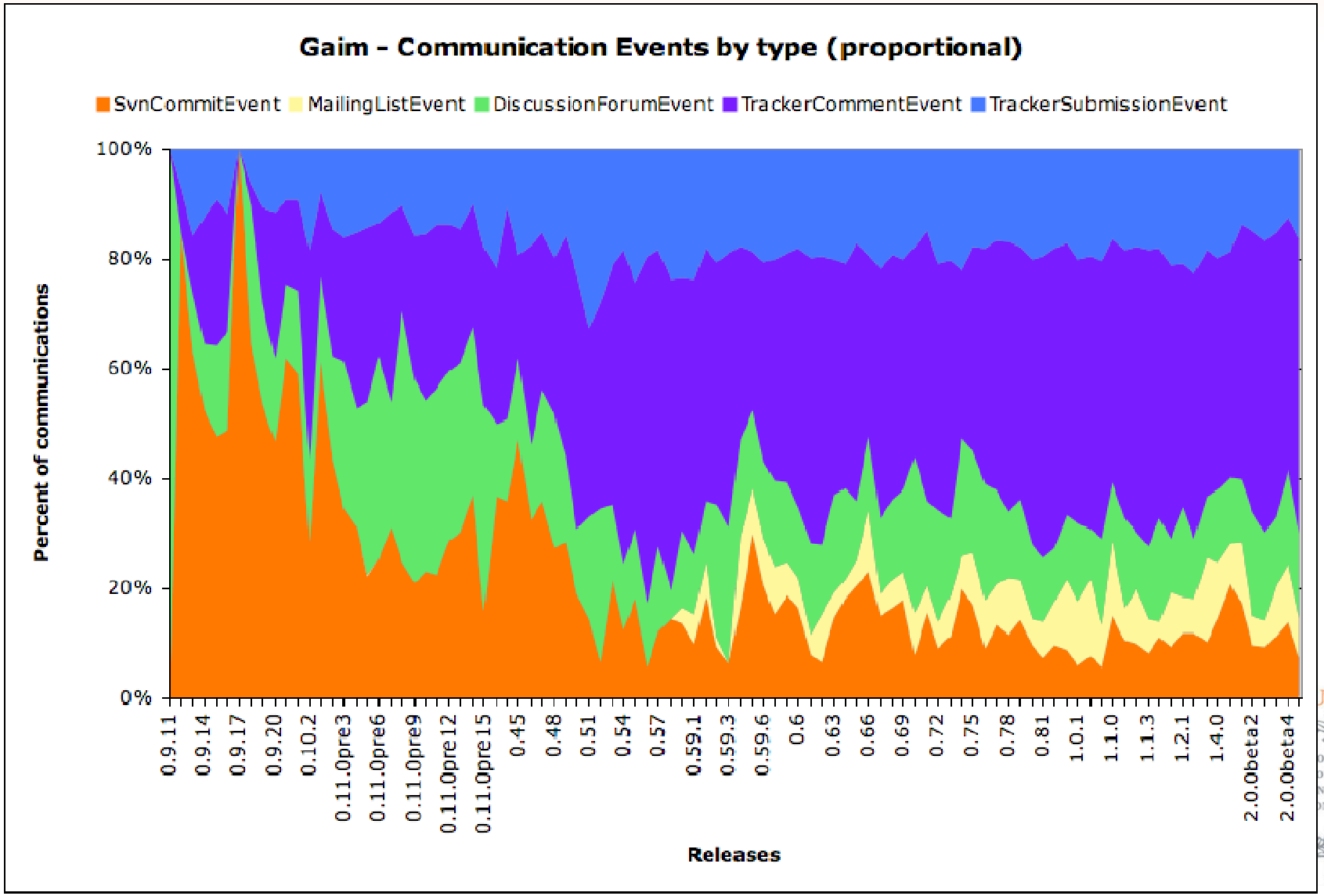

90 day window, sliding forward by 30 days at a time



# Intensity and Variation Across Venues

- Intensity
  - Created an intensity-based smoothing function
  - Exponential decay of interaction weight as time passes: more recent events more heavily weighted
  - Threshold dichotomization for use with binary SNA measures
- Variation in venues
  - Analysis of all venues for each project
  - Grouped by target audience/purpose for venues: users, developers, and trackers
  - Trackers include bugs, feature requests, etc.

# Venue Volumes Over Time



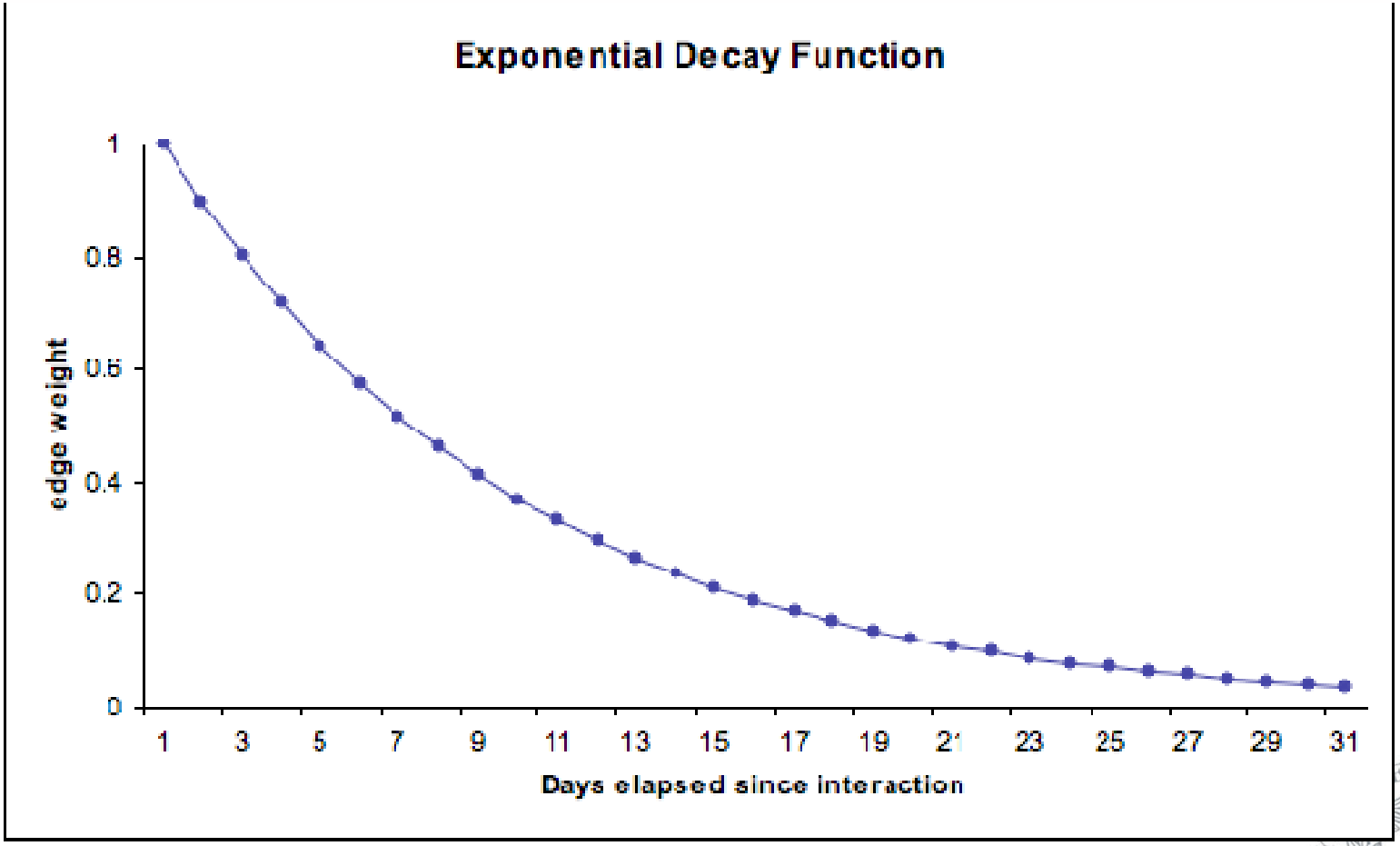
# Data

- Data from FLOSSmole
- Compared two projects: Fire & Gaim, both IM clients
- Gaim
  - November 1999 - April 2006, when project identity changed to Pidgin
  - 4 trackers, 1 user forum, 2 developer email lists
  - Considered successful
- Fire
  - 2001 - March 2006, project's final release
  - 2 trackers, 1 user email list, 2 developer email lists
  - Not considered successful

# Intensity-Based Smoothing

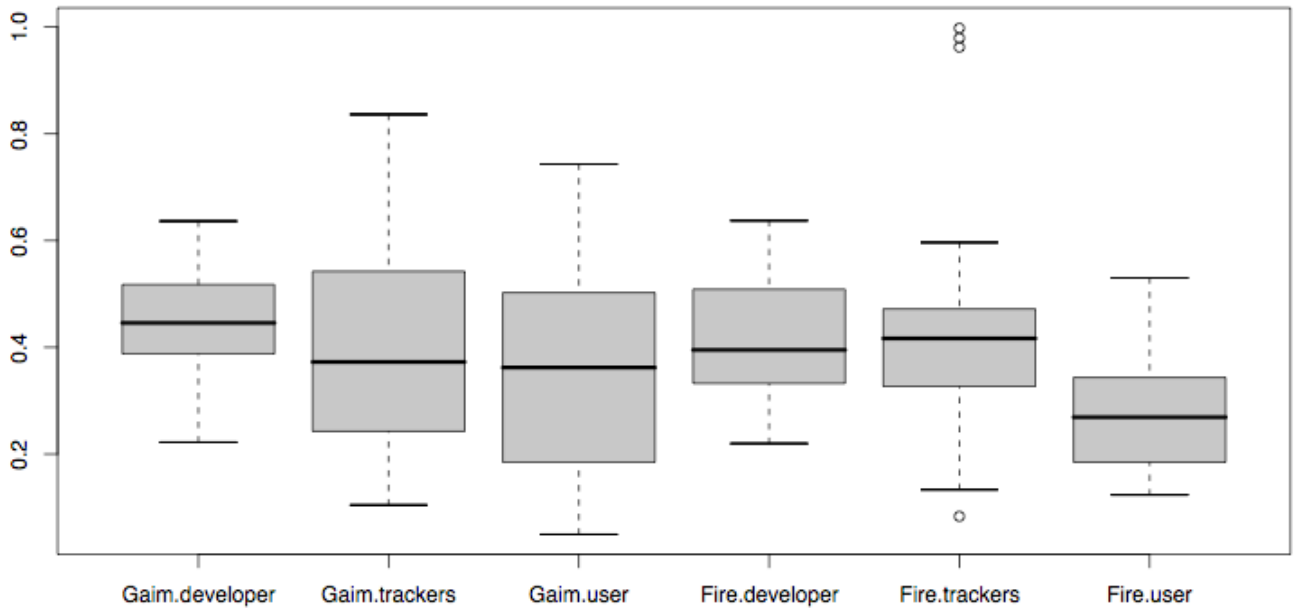
- (simplified) R script for calculating edge weights:
  - end.date, first.date, event.date: inputs for beginning and end dates for period, plus date of event
  - total.time <- end.date - first.date + 1
    elapsed.time <- event.date - first.date +1
    event.rate <- (total.time - elapsed.time)/total.time
    event.weight <- exp(-log(total.time)*event.rate)
  - Sum up interaction weights for each dyad in time period for edge weight
- Intent is to reduce undesirable effects of smoothing from overlapping windows

# Applying the Smoothing Function



# Findings: Variations in Dynamics

- Compared mean and standard deviation of centralizations in aggregated venues
- For both projects, different venues showed different communication dynamics




# Findings: Gaim

- Average centralizations lowest for user email list and highest for the developer lists: number of participants matters
- Standard deviations of the centralizations for the user forum and the trackers are comparable, while the standard deviation for the developer lists was much lower: consistency of participation
- Centralization trends reflect more varied participation dynamic in user forum and trackers, more regular for developer lists
- Periodic spikes in tracker activity (to be continued…)

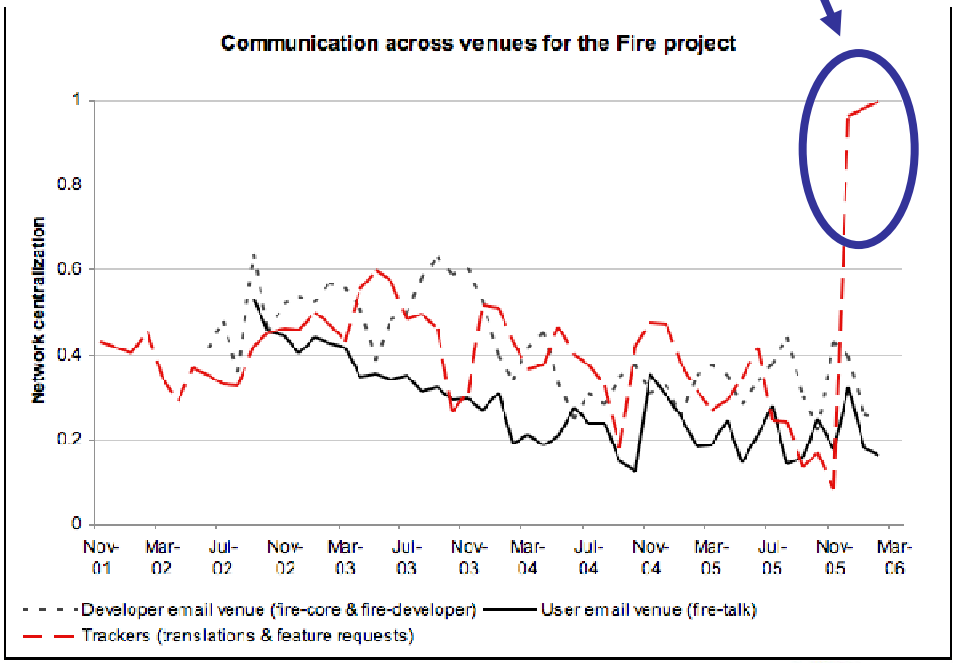
# Findings: Fire

- Comparable mean values for centralizations for user and developer venues, but higher for tracker
  - Anomalous data pattern, to be explained momentarily
- Excluding the anomalous tracker data, means and standard deviations of centralizations for trackers and email lists were comparable
- Suggests a degree of regularity across the communication venues
- All venues tended toward decentralization over time
  - Except for that anomaly…

# Tracker Housekeeping Behavior

- Observed anomalous patterns in trackers for both projects: periodic centralization spikes
- A single user makes batch bug closings (up to 279!)
  - Fire's (feature request) tracker housekeeping appears to be preparation for project closure
  - Gaim's tracker housekeeping was more regular and repeated

Cleaning up before shutting down



# Observations on Dynamics

- Different levels of correlation between venues, suggesting different types of interactions
- Overall trend toward decentralization over time could be a result of different influences
  - Fire: decentralization due to loss of project leadership
  - Gaim: decentralization due to growth in user participation
  - Highlights duality of centralization measure: can be affected by leadership and/or number of participants
- Variations indicate reason for concern over validity

# Conclusion

- Contributed an original method for exponentially decayed edge weightings in dynamic networks
- Variation in communication centralization dynamics across venues has implications for research design
- Periodic project management activities are apparent in batch bug closings by few individuals, which cause spikes in tracker centralization
  - Interesting housekeeping behavior, but also a potential confound for analysis based on trackers
- All venues in both projects tended toward decentralization over time