Using Digital Trace Data to Examine the Social Dynamics of Scientific Teams

Laura Anderson, Cheryl Kieliszewski IBM Research – Almaden Accelerated Discovery Lab

Introduction

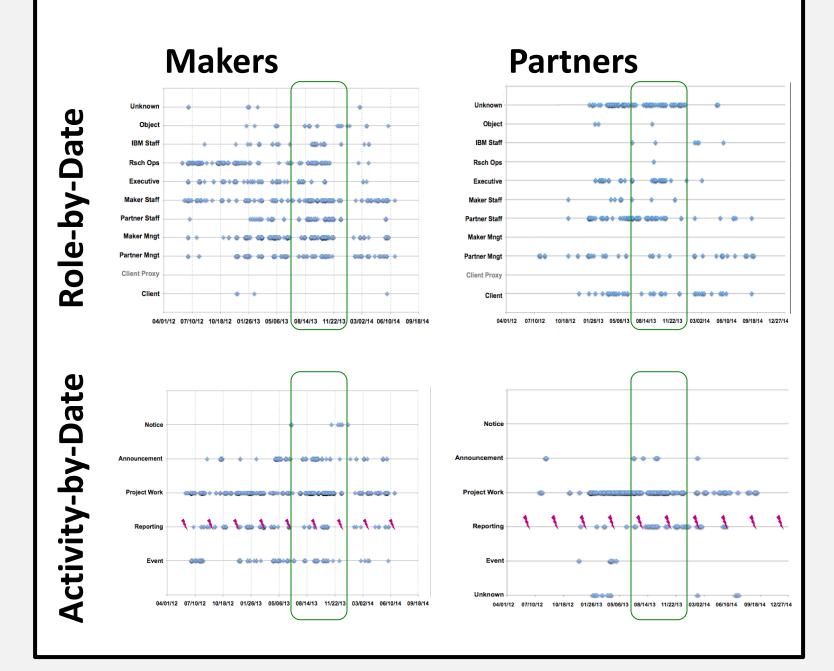
Scientific investigations today are often performed across organizations, disciplines, and geographic boundaries. It is our belief that understanding the social dynamics of team conversation and interaction affords the opportunity to identify efficiencies in (1) team operations, such as information sharing; and (2) relevant tools, technologies, and compute environments. To this end, we are examining digital trace data of scientific teams to identify common and unique patterns of interaction and conversation within and between projects.

Approach

Use a combination of digital trace data (email in this instance) and semi-structured text data sources to study the social dynamics of scientific teams using an analytical and computational approach.

- **Data:** email header/metadata and content
- Phase 1: Identify project cadence and effort.
- Phase 2: Identify topical themes and interaction.
- Phase 3 (future work): Build model using team and social dynamic characteristics based on empirical data.

Cadence and Effort



Topical Themes and Interaction Over Time

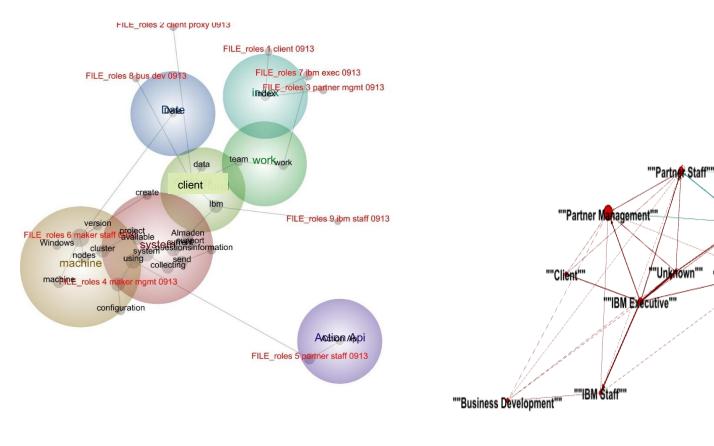
August 2013

- 9 roles create 2 communities that are homogeneous in role affiliation with an average path length of 1.37.
- Core roles each focused on a particular aspect of systems setup, working with data, code, and testing.

Persistent Systemat Itd. ""IBM Executive"" FILE_roles 2 client proxy 0813 ""Business Development"" FILE_roles 3 parting group 0813 ""Client" Information ""Client" Client Cli

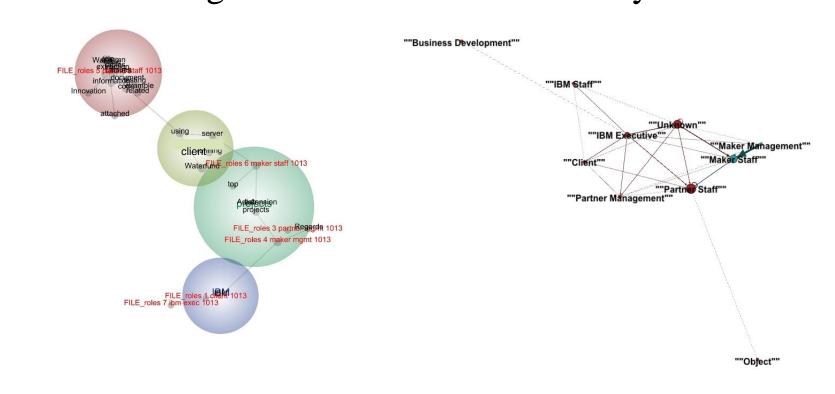
September 2013

- Increase to 10 roles that create 2 communities with an increased average path length of 1.78.
- Discussion focused on using computer systems and data led by Partner Management and Staff.



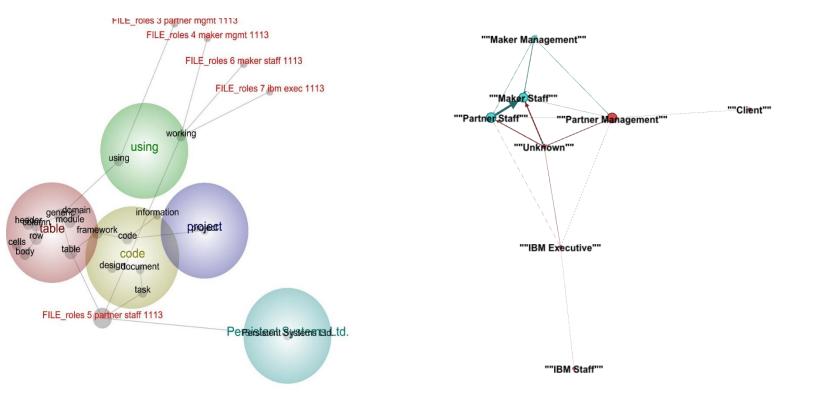
October 2013

- 10 of 11 roles contained in 2 communities with an average path length of 1.97.
- Fewer conversational themes and a shift to extracting and harvesting data from documents led by 6 roles.



November 2013

- 8 roles spread across 2 communities and close connections with an average path length of 1.5.
- 5 roles lead conversations with continued focus on data and emergence of architecting the deliverable.



Conclusions

- Analysis and juxtaposition of (1) social network and (2) content analysis provides two perspectives from one data source.
- Nuances in the social network become visible, particularly heterogeneity of composition and closeness of the connection.
 - illustrates the arrival and departure of project members
 - provides partial picture of work and interaction
 - content analysis identified natural language concepts and themes, and emerging and fading topics of conversation.
- Social dynamics mined from empirical data and to be mapped to model parameters.
 - evidence of changes in community composition
 - number of communities
 - average path length between members
 - number and proximity of role categories engaged in discussion
 - change/persistence in themes
 - variation in conversational focus at particular project phases

Literature cited

Anya, O., et al. (2015). Understanding the practice of discovery in enterprise big data science: An agent-based approach. In proceedings of 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, (pp. 4389 – 4396).

Dumais, S., et al. (2014). Understanding user behavior through log data and analysis. In *Ways of Knowing in HCI* (pp. 349-372). Springer New York. Haas, L., et al. (2014). The IBM Research Accelerated Discovery Lab. *ACM SIGMOD Record*, *43*(2), 41-48.

Pentland, A. (2014). Social physics: How good ideas spread—The lessons from a new science. New York: Penguin.

Acknowledgments

Special thanks go to the Partner and Maker members of the project used for this research and to all of our colleagues in the IBM Research Accelerated Discovery Lab for critical research discussions of this effort.

Further information

Contact either author through email: Laura Anderson at lca@us.ibm.com Cheryl Kieliszewski at cher@us.ibm.com

