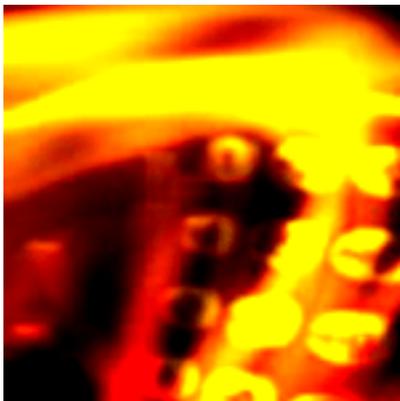


Department Editor: Niek Wijngaards

Trust and Reputation in Dynamic Scientific Communities

Omer F. Rana, *Cardiff University*

Annika Hinze, *University of Waikato*



The formation of collaboration networks (or communities) is an important latent effect in many computational science undertakings. Generally, collaboration networks bring together participants who wish to achieve some common goal or outcome (often over short time frames). Increasingly, scientific collaborations are becoming interdisciplinary, and scientists are working in informal collaborations to solve complex problems that require multiple types of skills. Such networks generally might consist of participants with complementary or similar skills, who might decide to collaborate to solve more efficiently a single large problem. We argue that, given the diverse skills that such collaborations involve, deciding which partners to cooperate with is both critical and difficult. Two particularly important factors in this process are trust and reputation.

COLLABORATION NETWORKS, DISTRIBUTED COMPUTING, AND TRUST

Various analyses of the formation of collaboration networks already exist for various domains, such as the formation of ethnic communities,¹ a school,² a business community,³ and even scientific collaborations.⁴ These analyses generally employ interview- or questionnaire-based techniques, often involving subjective assessment of the data. While studying scientific collaborations, Mark Newman used an alternative automated approach, which constructed a network on the basis of the participants' coauthorship.⁴

Such network structures often demonstrate that not all participants interact in the same way with everyone.⁵ Some establish many interactions; others maintain a limited number. Similarly, some participants will likely have a large connectivity (that is, will be connected to many participants) with small data exchanges, while others have the converse of this. Both of these interaction schemes make such networks more resistant to random unavailability of communication partners and consequently more adaptive to changes in their operating environment.

In recent years, people have been using distributed computing technologies to support such interactions. Here, a key theme has been to identify how effective resource allocation can improve information sharing in the system for participants who interact with many other participants. So, reducing the network latency or increasing highly connected participants' cache sizes will likely improve the system's overall performance. Recently, peer-to-peer (P2P) and agent technologies have gained increasing importance in the context of collaboration networks such as Gnutella.com and JXTA.org. These technologies use a "super peer" (such as a JXTA Rendezvous node⁶) to identify highly active peers.

Often the choice of such "hubs" or "highly social" participants depends on the level of trust others have in them or on their reputation in the network. Trust and reputation are complex, multifaceted issues and are related to other themes such as risk, competence, security, beliefs and perceptions, utility and benefit, and expertise.⁷

DETERMINING TRUSTWORTHINESS

Alfarez Abdul-Rahman and Stephen Hailes discuss three kinds of trust in the context of virtual communities.⁸ *Interpersonal trust* is the direct trust one participant has in another. *System* or

impersonal trust depends on how the participant perceives the institution or system in which he or she is participating. *Dispositional trust* is the participant's general trusting attitude.

Trust is therefore a subjective factor—one that lets one participant or a group of participants rate others within a community. In the selection of trustworthy participants, reputation becomes significant. You can think of reputation as the expectation about a participant's future behavior based on information about the participant's past behavior.

The automatic generation of a trust rating for participants—and therefore the identification of hubs—deserves thorough investigation. Existing trust and reputation efforts in P2P computing are frequently based on the PageRank approach⁹ that the Google search engine has adopted. This approach determines a page's rank by the rank of pages that refer to it. PageRank has led to the definition of useful trust models in P2P computing, such as the EigenTrust approach.¹⁰

THE OPEN SOURCE PARADIGM

However, additional research that considers a wider perspective on trust and reputation is necessary. Many scientific collaborations are generally predefined and static. Often, participants in such collaborations must belong to established institutions (such as universities or national laboratories), frequently because of funding requirements. But with the emergence of P2P and agent-based resource-sharing technologies, we can envision a future where participants can be individuals connected over DSL or cable modems. Such individuals might be enthusiasts who provide specialist expertise or data that isn't available to participants at established sites. The emergence of such ad hoc, dynamic scientific communities might significantly enable the establishment of previously infeasible multidisciplinary collaborations.

This more open view of collaborative scientific research has a precedent in the open-source software community, which makes source code available for auditing, vetting, and analysis by others. Such a process enables the quicker discovery and patching of software bugs (tools such as Bugzilla.org already feature heavily in open source software releases). Such a process also provides disincentive for programmers to insert "back doors"—essentially malicious code that they can exploit. In this way, open source software can be more trustworthy than proprietary software¹¹ (provided that the developer using the open source software has enough experience to fully understand the source code).

In reality, a problem with open source is that developers often face time constraints for exploring the provided source code. Furthermore, they might not be able to easily grasp the effects of

upgrades or hosting environments (such as when interpreting and running a Perl 4 source code on Perl 5 or a different platform). So, a developer downloading source code still must place significant trust in the initial designer and implementer. The initial designer's reputation therefore also becomes significant.

Perhaps open source's major advantage is that it allows contributions to source code from both individuals belonging to major academic or industry institutions and enthusiasts armed with a powerful laptop or a home machine. A significant aspect of this movement is the recognition that good ideas can come from a variety of sources—and not necessarily from well-established institutions.

CONCLUSION

If an open-source-like reality is to hold true for future scientific collaborations, trust and reputation issues will play a significant role. This means that ascertaining and using these ratings will be a necessary requirement in future distributed computing infrastructure.

References

1. H.R. Bernard , et al., "Studying Social Relations Cross-culturally,"*Ethnology* , vol. 2, no. 27, 1988,pp. 155-179.
2. T.J. Fararo and M. Sunshine , *A Study of a Biased Friendship Network*, Syracuse Univ. Press, 1964.
3. P. Mariolis , "Interlocking Directorates and Control of Corporations: The Theory of Bank Control,"*Social Sciences Quarterly* , vol. 56, 1975,pp. 425-439.
4. M.E.J. Newman , "Who Is the Best Connected Scientist? A Study of Scientific Co-authorship Networks,"*Physical Rev. E* , vol. 64, no. 1, July 2001; <http://aps.arxiv.org/abs/cond-mat/0011144>.
5. U. Schimank , "From 'Clean' Mechanisms to 'Dirty' Models: Methodological Perspectives of the Scalability of Actor Constellations," to be published in *Socionics: Contributions to the Scalability of Complex Social Systems*, K. Fischer and M. Florian, eds., Springer Verlag, 2004.
6. B.J. Wilson , *JXTA*, New Riders, 2002; www.brendonwilson.com/projects/jxta.
7. J. Sabater and C. Sierra , "REGRET: A Reputation Model for Gregarious Societies,"*Proc. 5th Int'l Conf. Autonomous Agents*, ACM Press, 2001,pp. 194-195; www.iiia.csic.es/~sierra/articles/2001/reputation.pdf.

8. A. Abdul-Rahman and S. Hailes , "Supporting Trust in Virtual Communities,"*Proc. 33rd Ann. Hawaii Int'l Conf. System Sciences (HICSS 33)*, vol. 6, IEEE CS Press, 2000, p. 6007; <http://citeseer.nj.nec.com/235466.html>.

9. L. Page , et al., *The PageRank Citation Ranking: Bringing Order to the Web*, tech. report SIDL-WP-1999-0120, Stanford Digital Library Technologies Project, Stanford Univ., 1999; <http://dbpubs.stanford.edu:8090/pub/1999-66>.

10. S.D. Kamvar , M.T. Schlosser, and H. Garcia-Molina , "The EigenTrust Algorithm for Reputation Management in P2P Networks,"*Proc. 12th Int'l World Wide Web Conf. (WWW 2003)*, Int'l World Wide Web Conf. Committee, 2003; www.stanford.edu/~sdkamvar/papers/eigentrust.pdf.

11. S.T. Ross , *Unix System Security Tools*, McGraw-Hill Osborne Media, 1999; www.albion.com/security/intro-7.html.

Omer F. Rana is a senior lecturer at Cardiff University's School of Computer Science and the deputy director of the Welsh eScience/Grid Computing Center. Contact him at o.f.rana@cs.cardiff.ac.uk.

Annika Hinze is a lecturer at the University of Waikato's Department of Computer Science. Contact her at hinze@cs.waikato.ac.nz.