Towards a Multidisciplinary Approach to Peer Production: Questions and an Application

David Gutelius

Social Kinetics

gutelius@gmail.com

Abstract

The past decade has seen explosive growth in user-generated, user-mediated content and the web as a fundamentally social space. More recently, however, the web has increasingly become a site of "production" – cultural, social, and economic – in ways few could have predicted. The rise and shifting nature of novel types of peer production and the composition of the resulting artifacts are a largely unexplored phenomenon. This paper argues that a new cross-disciplinary research agenda is vital to understanding how people get things done together online and optimizing peer production. iLink, developed under the DARPA CALO program, is a machine learning engine designed to explore these issues and to help improve peer production in a real military context. The iLink project made clear that we stand at edge of a virtual green field; the right lines of inquiry and exploration could yield fundamental new breakthroughs in social and computer sciences, and provide the foundations for a new generation of intelligent, peer-production enabled applications.

Challenges and Opportunities

Much ink has been spilled over peer production mechanisms as essentially an Information Retrieval problem. More recently, computer scientists and some social scientists have approached the problem from the perspective of social network analysis. While both of these approaches can yield extremely interesting insights, isolated from one another, each misses important elements of how and peer production happens over time. Both approaches have only rarely taken into account the highly dynamic, temporally-specific nature of communities and that this dynamism is reflective of the nature of humankind itself. Graph analysis, alone, tends to consider the aggregate statistics of human interaction and assign overly general, undifferentiated values to both individuals and the groups to which they belong. IR, alone, ignores the social context in which information – and more importantly

knowledge and meaning – is generated and in constant interpretive motion as it traverses networks. In other alone, these disciplines lack the analytic tools and methodological frameworks necessary to understanding the bigger questions around peer production, collective intelligence, and cooperative sensemaking that they seek to answer. More cross-domain inquiry that combines the best of a wide variety of disciplines to ask new kinds of questions is necessary to understanding peer production and technology-mediation social participation.

The importance of effective peer-to-peer production mechanisms seems obvious in the aftermath of the Sept 11 and NASA catastrophes. And yet team members in these contexts are essentially using the same tools and passive web applications that they were a decade ago. Email predominates as the default peer production mechanism among teams working to accomplish tasks. While "Web 2.0" applications like Twitter and Facebook increase the weak-link like ties in networks and the chances for serendipitous discovery (to say nothing of "social prioperception" – in the words of one well-known and over-eager blogger), they also can hurt as much as help.

Some common challenges to effective peer production for which we have inadequate integrated theoretical frameworks and few compelling application solutions include:

The User Generated Web tends to be extremely "noisy". It can be shrill. It is most often personally irrelevant, even in a purely social sense. Closed "work" communities are often not much better. Noise increases the "costs" of usage and engagement. Transaction cost theory and allied insights from behavioral economics, not to mention cognitive psychology, have rich models that could be extended to add insights here.

Beyond the noise communities tend to generate, the sheer volume of information can quickly overload users. Community members often are overwhelmed with potentially relevant information, unable to effectively keep up with and filter for the most personally meaningful and relevant information.

Sensemaking is hard alone, and can be even harder together. Working cooperatively requires cognitive alignment and effort that can be hard to achieve/maintain, and the benefits of working with others do not necessarily outweigh the costs.

Online communities are subcultures, and ever-changing ones at that. There is little appreciation of online communities as both similar to and yet different from offline communities. Economic sociology, behavioral network economics, and anthropology have important insights here, but have been largely ignored by computer scientists seeking to understand change in human societies over time. For example, how do communities define themselves? What signaling do they use to communicate to one another? What are the means of social reproduction and how to members "teach" one another shared value and symbolic systems? What insights from behavioral signaling, linguistics and NLP, could be combined with learning theories and pattern detection to explore these dynamics? In terms of collective problem-solving, what structural traits (if any) makes some communities more or less successful?

Meaning – and by extension, "relevance" -- is understood, mediated, and manipulated in a dynamic social context. As documents or questions are passed around a human network, the original artifact changes subtly as it comes into contact with and is manipulated by that network. Members of the

community may edit the document and pass it along, or ask questions about the document. Members of the community may make clarifying points that alter and focus the nature of the original question. Artifacts may be rated explicitly, or implicitly through the velocity at which they traverse the community, the discussions that happen about the artifact, or by the fact that the artifact is roundly ignored. The meaning attached to artifacts is a function of dynamic social processes that are difficult to characterize. There are a remarkable variety of meta-signals attached to peer-produced artifacts, but we lack coherent ways of detecting, assessing and measuring these signals as they wax and wane.

Trust and reputation are difficult and costly to assess. Yet they remain essential elements of effective cooperation. While these concepts are often subsumed under the term "social capital", they in fact point to somewhat different ideas. How is trust defined and gained in different kinds of communities? How important is reputation? When does it matter (most)? And how would we measure such abstract, subjective value judgments? Do these terms take on different meaning online than offline?

Diversity of interpretation, opinion and contributions is likely key to the quality of peer-produced artifacts, but how can that be enforced – let alone accurately assessed? Recent work suggests that diversity is likely more important than sheer size of community in, for instance, collective prediction. Further, how do we represent and measure diversity over time?

Incentives matter. How can insights from mechanism design be combined with HCI design, cognitive psychology, dynamic graph analysis and behavioral economics to generate novel choice architectures that best motivate collective behavior? How can machine learning and other artificial intelligence approaches be deployed to optimize a group's collective output?

Today's web applications are passive and put all of the burden of collaborating and sensemaking on the individual. Simply exposing users to communities and the artifacts that they generate does not collective intelligence make. Recent studies in fact highlight the quite the opposite – that effective, prolific minorities dominate and bias so called "peer-production". How can we create software that is more "aware" and "adaptive" to a user's context, technology that can learn to improve over time and that is fault-tolerant?

Individuals play different roles in different communities. Further, those roles change frequently as individuals themselves change. How and why do these roles change over time? What do social pressures (positive and negative) have to do with these dynamics? Are there discernable characteristic signals that indicate such shifts? Are those patterns structural, correspondent to a bounded community?

The remainder of this paper describes work under DARPA's CALO (Cognitive Assistant that Learns and Organizes) program that attempted to address a subset of the questions raised above in a real community of military users. While our experience only scratches the surface of these issues, it also points up both the opportunities for further developing the line of inquiry as well as the real importance of creating effective peer production mechanisms.

Background: The iLink Project under CALO

The CALO program was focused on building a cognitive assistant. Initially, this goal was understood as assisting the user with computational resources, either local to the desktop or web-based. Real assistants certainly provide this kind of help (e.g., assisting with the use of on-line scheduling systems). Good assistants also provide help, however, with *human* resources, e.g., human expertise identification. In general, social networks are an essential resource for knowledge workers. In fact, social networks may be the single most important resource for a user both in terms of depth and frequency of use. The CALO program incorporated this notion of assistance as the iLink sub-project.

A canonical problem for social network assistance is expertise and stakeholder identification - who knows what? The idea is that important knowledge is 'somewhere' in the network. The idea is that if there is a catalog of this knowledge then a user could access this catalog and find the right small set of targets for a query. Many companies particularly consulting companies that need the ability to rapidly form pools of expertise for new projects have tried data-entry style knowledge catalogs. For the most part, these catalogs have not been terribly successful since it is hard to standardize and represent knowledge categories and users tend not to be highly motivated to or capable of filling out these forms systematically.

The learning frameworks developed under the CALO program provide a good foundation for a very different approach to the knowledge inventory. By *learning* user attributes through a wide-variety of behavior there is no need for users to fill out forms or keep them up to date. Additionally, by learning how the network itself interacts with a given user or knowledge artifact, it is possible to create a more nuanced and contextually relevant profile. For example, if messages about a particular topic are always routed (by people) to a given user, it can imply some kind of relationship of that user to the topic. If these messages are questions, then we can understand the network as providing a signal about this user's expertise.

As the iLink project worked on learning-based approaches to expertise identification, a second notion of social network assistance emerged. Imagine that some member of a community has submitted a question to a community. Suppose moreover that the expertise catalog is very incomplete and that the expertise to answer this question is not resident in any single person. We can now think of the problem as a question being addressed *to the network*. We can then imagine a new goal for assistance, namely the goal of assisting the network in getting a question answered, i.e., producing some new knowledge such as an entry in an FAQ, or frequently asked questions. This way of posing the problem led to an approach to message routing in social networks that was the basis for all of the applications that the team developed. The work is described in Davitz et al.

The basic goals for the iLink project were thus straightforward: to understand some of the conditions under which peer production mechanisms worked effectively and use this understanding to create new adaptive, aware, web-based applications that could help create these conditions. In attacking the problem, we a) extended the notions of node and link models and b) created an associated learning framework optimized to the models and social dynamics of the community. The network of nodes and links had to acquire a variety of different – and dynamic -- property vectors. Both node and link

properties could be learned or exogenously defined, and both probabilistic and dynamic. In developing our learning approach, dubbed the Social Query Model (SQM), it became clear that we needed a semi-supervised method that allowed both batch- and online- learning to happen in an optimized fashion over these heterogeneous sources, and at significant scale. The resulting approach used insights from computational linguistics, probabilistic learning, social network analysis, economic sociology and behavioral economics. The technical approach is described in Davitz et al. (2007) and Basu (2008)

The team's interest in social networks and intelligent message routing had been consistently informed by the growing interest in and use of social media in the military and intelligence agencies. It may not be widely known, but the deployment and adoption of social media is very widespread in both of these communities. There are, of course, many reasons including the need for intense collaboration, the reliance on networks for operational effectiveness and the constant demands for learning and training. The remained of this section describes the work done by the iLink team in collaboration with a group that pioneered the use of social media in the military, the Center for Company-Level Leadership at West Point.

Background on CCL and CompanyCommand

A volunteer army is a reflection of the society it serves, and nowhere is this more evident than in the rise and spread of social media. As the Internet grew from curiosity to commonplace in the American home and workplace, a similar revolution was happening in the ranks of the Armed Forces. In the spring of 2000 a team of officers developed and launched CompanyCommand.com as a means to connect past, present, and future company commanders in an ongoing conversation about leading Soldiers and building combat-ready units. This website sought to duplicate what the founders identified as critical to their learning as company commanders—countless front porch conversations in which they were collaboratively learning from each other and addressing their greatest challenges.

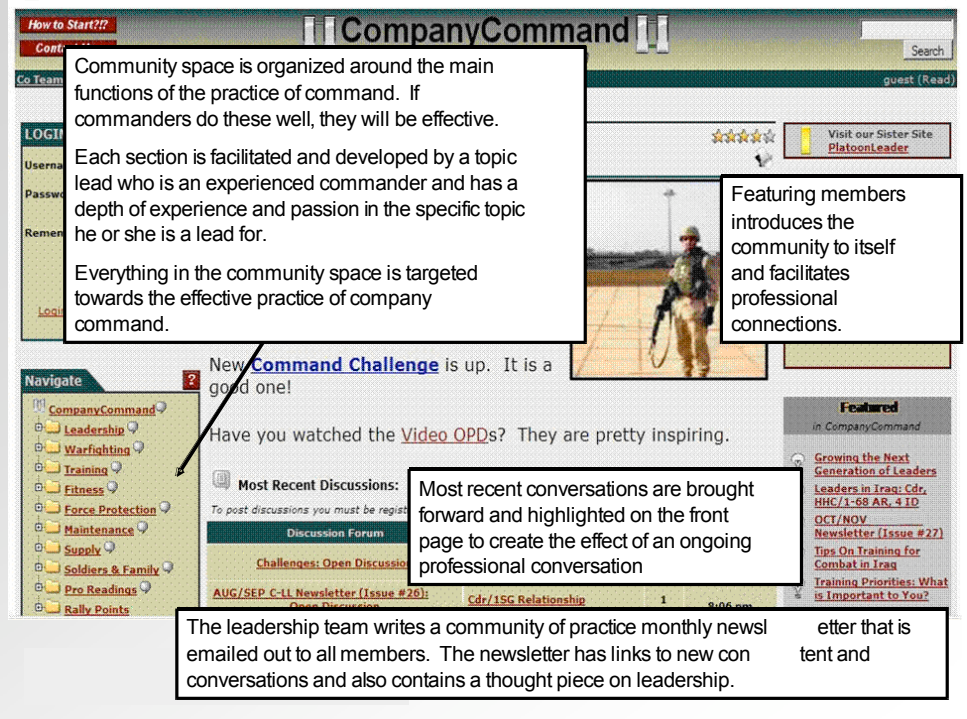


Exhibit 1: Example screen shot of the Company Command professional.

Subsequent to the successful launch of the Company Commander professional forum, this same team launched a sister forum with a similar vision for Platoon Leaders.[1] The founders described these environments as "professional forums." The term forum brings to mind the ancient Roman Forum, a gathering place for conversation that served as an incubator for ideas that advanced the entire society. And, including the word "professional," places a forum in the context of a particular profession's advancement, in this case, that of the Army. "Professional Forum" thus communicates who participates (members of the profession), why they participate (to improve the profession's effectiveness), and how they participate (with candor and mutual respect). Most of all, it identifies conversation and the exchange and creation of professional knowledge as the defining characteristics (Dixon et al).

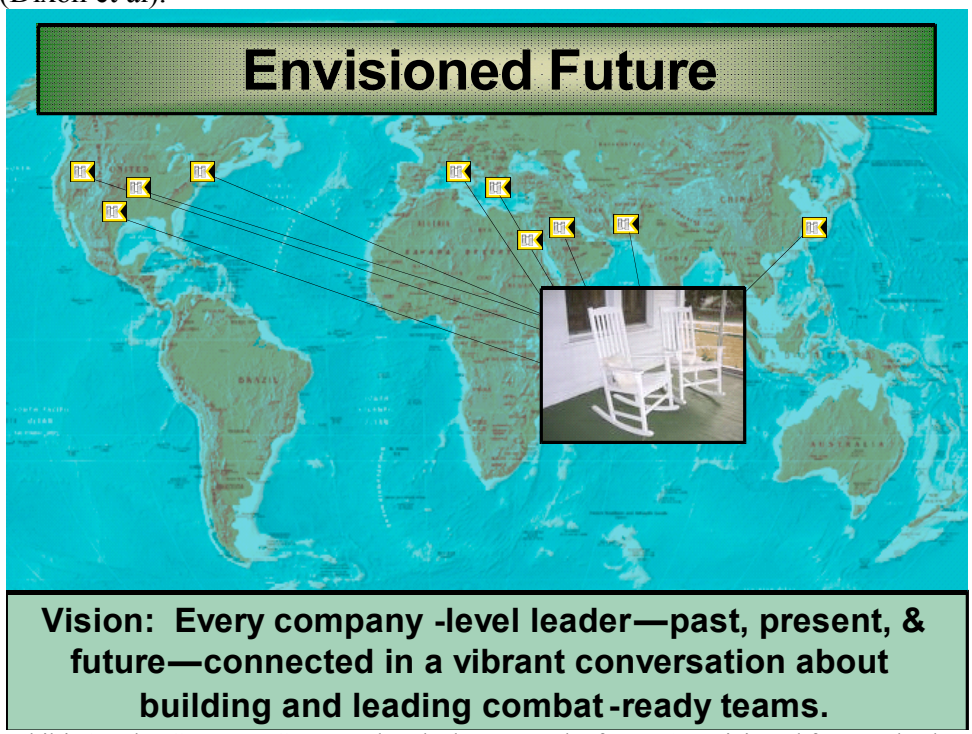


Exhibit 2: The CompanyCommand and PlatoonLeader forums' envisioned future—leaders distributed globally in over 120 countries in an ongoing, virtual "front porch" conversation about what they are learning and collectively addressing the contemporary challenges they are facing.

This concept quickly grew beyond the CompanyCommand and PlatoonLeader forums and spread throughout the Army and the other U.S. Armed Forces. Today, professional forums are found in every branch of service, and within the Army, are represented by the Battle Command Knowledge System (BCKS), which describes its mission as: support[ing] the online generation, application, management and exploitation of Army knowledge to foster collaboration among Soldiers and Units in order to share expertise and experience; facilitate leader development and intuitive decision making; and support the development of organizations and teams (bcks.army.mil). In short, the long-term vision of BCKS involves every leader in the Army tapping into a peer-to-peer professional forum related to every job responsibility they are given throughout their career.

1 Platoons and companies are small Army units containing between 15-40 and 60-200 soldiers, respectively. They are the "building block" units of the Army; every soldier in the Army belongs to a platoon and a company. They are also the first leadership positions held by commissioned officers.

As these professional forums prosper and grow, so does the need to make the knowledge that resides within them more readily accessible to the community. Members join in order to connect with like-minded individuals about their challenges, ideas, and desires; as such, they need a way to quickly find and comprehend relevant content and conversations that address their unique context and professional development. Furthermore, as these forums have grown to several thousand members and several times that number of knowledge objects, the member experience has potential to be overwhelming and lacking in relevance. This requires more than just a search engine—search engines can produce reams of data, but they require the individual to know what it is they're searching for in the first place. Especially, for the young officer just starting out in command of a platoon or company, the sheer magnitude of the tasks confronting them can seem overwhelming. Add to that the rapid tempo of an Army that has been at war for nearly a decade, and one begins to get a picture of the daunting task confronting new leaders in today's Army. The challenge at hand was to create an intelligent learning aid that adapted to encompass each individual's changing work context and understanding of their profession as they worked through the challenges of their service.

iLink Deployment into CompanyCommand

The iLink team reviewed the three online communities within CCL to both understand how these communities functioned as well as how CALO iLink technology could most effectively improve their functioning. Early on, the teams identified three "C's" – content, conversation, and connection – as building blocks to a successful community and where targeted learning and personalized relevance could have a major impact on peer production. In concert with the CCL Leadership Team, we identified the following goals for the iLink deployment in order to deal with these challenges:

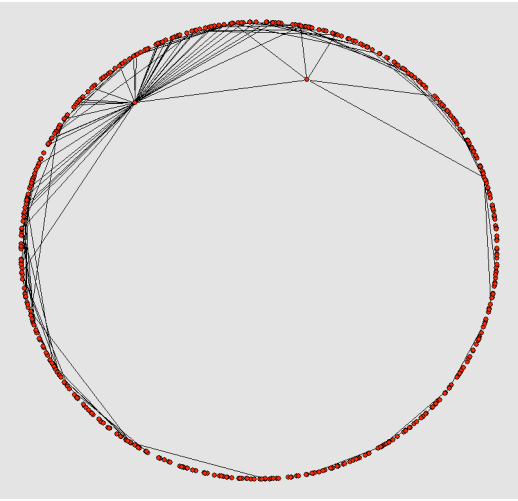
1. Improve user cognition by focusing on providing users quick, convenient suggestions of the most relevant people and knowledge in the community.
2. Solve problems quickly by suggesting the most relevant teams in the community for a given topic, issue, or discussion.
3. Recognizing niche expertise. To a user who has a question, it's difficult to know who to ask and who to trust on a particular topic in large online communities.
4. Improve the quality of discussions by assembling the right team and resources around the right problem, on demand
5. Encourage users to "grow" from "lurkers" into active community participants.
6. Build the sense of trust in the community itself.

The CCL and CALO iLink teams worked to design a scalable solution that included technical integration with the existing CCL community software, interface integration, and customization of iLink natural language processing capabilities. We ultimately created lightweight recommendation widgets that could make remote web services calls to a modest iLink server. New user events such as interactions with the iLink widgets and their recommendations, authoring new content, responding to existing content, etc. were all piped to the iLink server for processing model updates every three minutes. The widgets themselves were of four simple types:

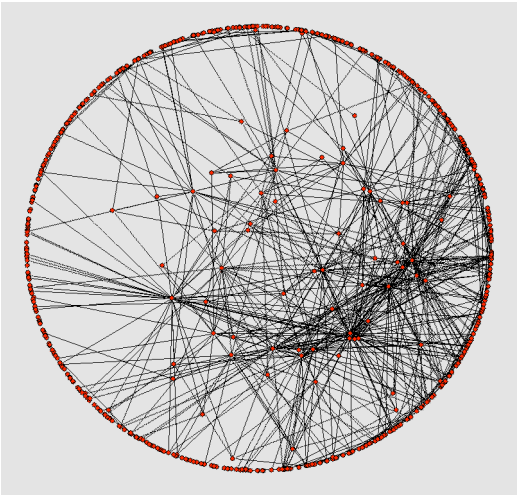
1. Recommending other community members to meet for the user, based on similar expertise and interests (which appears on a user's homepage)
2. Recommending highly relevant new posts and documents for the user (also appearing on a user's homepage)
3. Recommending highly relevant community experts for a given ongoing discussion
4. Recommending highly relevant documents, discussions, and other materials a given ongoing discussion

We developed a basic quantitative and qualitative metrics framework to assess usage and performance of iLink against our goals. We created a customized data processing pipeline for our quantitative analysis needs and separated the population into four categories: 1) all users, 2) all those served iLink, 3) those who were served iLink and interacted with it and 4) those served iLink who did not interact with it. Quantitative measures included a range of customized social network analysis, overall traffic and specific usage statistics. We also developed a questionnaire with the CCL Leadership team, who then picked a random sample of CCL community members as our response group.

The outcomes after two months of iLink being deployed were encouraging, both quantitatively and qualitatively. From a network perspective, key metrics indicated a more vibrant community for those who used iLink versus those who did not. Density and closeness were both substantially higher for iLink users than non-users. Hierarchization, degree and group degree centralization were markedly lower for users versus non-users. Reciprocity and sentiment dynamics measured by directionality and group clustering were similarly significantly different between the two groups, as iLink users tended to have more opportunities to "give back" inside the communities.



Non-iLink users



iLink users

Anecdotal evidence from community members supported the quantitative results. iLink was reportedly able to identify niche experts that had previously not been active in the community, and made it easier for community moderators to do their job. It was also able to recommend high-quality older discussions and documents as resources for questions previously well-addressed. As one survey

respondent put it one day after iLink's initial deployment, "iLink is awesome. There is so much knowledge out there that I did not know about. Where does it come from... where has it been hiding... It's almost like coming to CC for the first time."

Summary

We have described how a novel line of research that drew on insights from multiple disciplines was quickly adapted to a set of immediate peer production needs inside the military. While iLink's original goals were modest, the project's success underlines both the importance of and opportunities in developing an underlying foundation of cross-domain research in peer production and technology-mediated social participation. Social media continues to grow and evolve in the military and is becoming understood as a core management/communication tool, particularly in the case of asymmetrical warfare settings. iLink applications have been running for 2 years in these communities and have since been deployed more widely in DoD production systems. The demand for effective online peer production mechanisms clearly continues to grow – not simply within the U.S. government, but also in work settings, civic participation and everyday life. Meeting that demand requires a new kind of interdisciplinary research agenda.

References

Arindam Banerjee, Sugato Basu . **A Social Query Model for Decentralized Search.** *Second ACM Workshop on Social Network Mining and Analysis at the KDD Conference* (SNAKDD-08), Las Vegas, Nevada, August 2008.

Davitz, J., Yu, J., Basu, S., Gutelius, D., Harris, A. iLink: Search and Routing in Social Networks, in Proceedings of The Thirteenth International Conference on Knowledge Discovery and Data Mining (KDD) August 2007.

Nancy Dixon, Nate Allen, Tony Burgess, Pete Kilner, and Steve Schweitzer, *CompanyCommand: Unleashing the Power of the Army Profession,* Center for the Advancement of Leader Development and Organizational Learning, West Point, NY: 2005.

"Mission and Objectives – Battle Command Knowledge System", as accessed on bcks.army.mil, October 24, 2009.