

Online Knowledge Crowds and Communities

Caroline Haythornthwaite
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign

Paper presented February 2009 at the International Conference on Knowledge Communities organized by the Center for Basque Studies, University of Nevada, Reno, and revised for publication by the Center.

This draft last edited on May 31, 2009

Abstract

While much has been written about virtual knowledge communities, particularly in how to create and sustain long-term, strong-tie relationships, the connection has not been made to newer forms of online organizing such as crowdsourcing. This paper addresses the way knowledge collectives are organized online, considering the organizational and motivational structures that support these new knowledge collectives, and contrasting the social mechanisms that support *crowdsourced* knowledge from those that support *community-based* knowledge. Examination of the literature and cases of crowds and virtual communities suggest a number of important dimensions that distinguish these two forms of organizing, including contribution type and size, personal coorientation and commitment to the knowledge topic, interpersonal ties among contributors, authority and control of contributory practices, and recognition and reward systems. Exploring these different models of organizing knowledge provides insight into the ways to establish and maintain crowd- and community-based knowledge collectives, and also show why strong knowledge communities such as those found in academia come to change their knowledge distribution practices, notably from print to online publication.

Introduction

Since the emergence of the Internet, there has been debate about the nature and quality of online communities, with concern that the lean, text-based media that support online discussion provide a pale alternative to the richness of face-to-face interaction. Nevertheless, social connection has found a way, with online participants developing norms and standards for online interaction that reflect and extend offline practices, with the Internet now supporting vibrant, geographically-distributed knowledge communities. The technologies have also evolved. While still heavily text-dependent, online interaction now extends to the use of images, audio, and video, and linear text has given way to more comprehensive and personal use of screen space for the presentation of self, knowledge, and work. Diversity in the ways and means of presentation of the individual are equally matched in presentations of the collective, and in ways of organizing collectives. Simple, email-based listservs, and web sites modifiable only by designated webmasters, are giving way to collective, participatory methods and forums such as online discussion boards, blogging and commentaries, and collective definition and editing through wikis.

These means contribute to the emergence of *distributed knowledge communities* that coalesce around topics of interest. As Lee Sproull and Sara Kiesler¹ predicted in 1991, communities have been arising around shared interests rather than shared geography or employer. To be sure, such communities have existed for centuries among distributed scholars, from the letter writing of the 17th century among members of the science and philosophy community² to

the pre-internet invisible colleges of academia³. Yet, as in many areas affected by the Internet, quantitative change leads to qualitative difference.

Quantitative change in access has led to qualitative difference in social practices: in who we interact with, from where, how often, and for what purposes. Internet structures make it much easier to start a new platform for a knowledge initiative, be found online, and gain participation. Internet connectivity permits frequent engagement, from anywhere at anytime, shrinking the turnaround cycle of communication and thereby creating a sense of connection and community with others. Further, as Internet connectivity becomes commonplace, it also becomes legitimated. Thus, an online interaction, information retrieval, knowledge exchange, or community is accepted as a legitimate venue, with the knowledge and interpersonal ties gained there considered as legitimate as any found offline.

Although this all seems self-evident, even as little as three years ago, the Modern Languages Association was just forming recommendations about online publication and tenure consideration⁴. Although rapidly adopted in physics, online publication of drafts, pre-prints and peer-reviewed articles have been a hard sell over the last 10 years in many knowledge communities. Surprisingly, two arms of the university have been at odds about these standards: while tenure committees have held to the standard of traditional peer-reviewed, print-based journals and books⁵, libraries have been facing the rising costs of journals, particularly in the sciences, and the unenviable position of buying back the work of the university's own employees. University libraries have thus become the leaders in creating online institutional repositories as long-term storage facilities for the work of their own university faculty, students and staff (e.g., the Association of Research Libraries' Scholarly Publishing and Academic Resources Coalition (SPARC)). As we will return to below, the whole notion of online academic publishing is tied to the larger issue of what motivates academics in their profession, and how an orientation to the ideals of open science merges with ideas of open access and the legal platform of creative commons licensing to provide the underpinning for change.

Strong adherence to existing practices may be one reason that online communities were greeted with such suspicion when they first arrived on the scene. The debate about whether online communities are 'real' communities has centered on whether these initiatives can sustain the kind of multi-threaded, interpersonal ties that lead to commitment to community goals and values.⁶ While some still question the application of the term "community" to online venues, there has also been much work substantiating their place as suitable environments for collaboration, knowledge co-construction, and communities of practice.⁷ This work has given insight into what it takes to build a community online, and many examples now provide evidence and models of sustained commitment to communal goals (e.g., Howard Rheingold's descriptions of the Well; Nancy Baym's description of soap opera fans; or Christina Preston's description of Mirandanet, a knowledge community for learning about educational technology, <http://www.mirandanet.ac.uk/home.php>⁸).

The debate about the legitimacy of online communities carries over to knowledge communities where the challenge is to the legitimacy of online knowledge activity. Not only is the knowledge that is posted online suspect, so too are people's commitments to collaborative activity. What kind of trust can be placed in an online relationship? What hold or expectation can we have that the individuals involved will contribute their part of the work, maintain civil discourse online, and not steal the ideas and sell them for profit? Indeed, many of these cautionary notes are necessary, but they are not unique to online venues. But our measures of trust and relationships have had to be transformed for the online world. Thus, it is of importance

to consider what motivates and sustains commitment in online venues where the supposed social controls of traditional face-to-face interaction and community are seemingly absent.

Meantime, while we've been grappling with the question of how to gain strong, long-term, high overhead commitment to knowledge communities, another form of collaborative activity has arisen premised on exactly the opposite set of principles – weak, short-term, low overhead contributions to knowledge. *Crowdsourcing*⁹ has become a new form of knowledge activity, particularly notable in the creation of Wikipedia, but also prevalent in blogging and commentaries associated with blogs, and collaborative writing projects that encourage reader contributions.

At one level, the promise of online knowledge crowds is the promise of participatory democracy,¹⁰ newly minted as the promise of citizen journalism and participatory culture.¹¹ This view sees the Internet as providing a platform for *netizens*¹² to voice their opinions and create an alternate structure for engaged citizenry – an alternative, complement or watchdog to government or to the media depending on the particular discussion. In this case, crowdsourcing is the beginning of some larger change or action, with an idea of continued attention and action around the topic of interest. Although this crowd is not a community, it is assembled *in the interests of community*, i.e., of society, and members of this crowd are likely to pay attention to the actions of other members of this crowd even while not interacting with them directly.

This crowd of concerned netizens sits somewhere between notions of a virtual community and the leaner, more independent, and often commercial perspectives on crowdsourcing. This kind of crowdsourcing looks to individuals as independent contributors to a collective enterprise, but not a collaborative one. Individuals contribute information to a computer program, but have little or no need to engage with fellow contributors. Whether for commercial or scientific purposes, the collective end result belongs to and is made use of by an *authority* outside the contributor pool (e.g., SETI@home, GalaxyZoo, AfricaMap, NASA clickworkers, 23andMe). This contrasts with a richer type of crowdsourcing that generates modifiable and updatable content, such as the encyclopedic knowledge brought together on Wikipedia, or the geographic information on OpenStreetMap, and puts these kinds of enterprises closer to the netizens than to the independent contributor crowds.

Borrowing the idea of rich and lean media first described by Daft and Lengel¹³ in 1986, *richer media environments* contain more opportunities and means for discussion, evaluation, and correction than *leaner* media environments. Daft and Lengel described rich media as better communication means for discussion, negotiation and interaction. They are better means for dealing with problems of equivocality where the even the questions to be asked must be negotiated. Lean media are best used for known data exchange, e.g., for dealing with problems of uncertainty where the questions are clear but data must be found to answer them.¹⁴ As will be discussed further below, richer collective environments, online or off, contain structures for internal, self-governance that allow for negotiation and adjustment of goals, whereas lean environments support more authoritative control on information use and goals.

Knowledge crowds provide an interesting contrast to knowledge communities and one that helps tease out the motivations of contributors to each. Questions we can ask are: How are these different kinds of knowledge initiatives organized? What distinguishes a successful, free-wheeling, online, crowdsourced knowledge collective from the formal dynamics of a knowledge community? How does each support knowledge processes of learning, debate, data collection, analysis, dissemination and evaluation? To explore this, we turn to some examples of these

forms of organizing, examining the features that distinguish them, and how these facilitate contribution and participation in these different knowledge systems.

Modes of Contribution and Participation

Crowds and communities may be considered at two ends of a spectrum of organizing. At one end are efforts that harness the knowledge and talents of many (relatively) anonymous individuals through online systems that aggregates discrete contributions into a whole. At the other end are communities, that meld, form, and define knowledge through the continued efforts of among a set of known participants. Each has different patterns of contribution, participation, aggregation and evaluation in their organizational structures.

Knowledge Communities

At the community end of the spectrum is a mode of contribution and participation I will describe as *heavyweight*.¹⁵ “Weight” in this case is used to refer to the commitment an individual has to the collective enterprise, not to the significance of it. Commitment to a knowledge community entails the heavy burden of learning about the topic, equipment, methods, and norms of production around this domain of knowledge¹⁶. To this I would add the need to learn about the *networks of knowledge and expertise*.¹⁷ This entails learning about the *knowledge structure*: How are knowledge concepts related? What references are most used? What standards exist for comparing references? It entails learning about the *actors* and their place in the knowledge network: Who is a novice, and who the expert? How are novices and experts tied? What relationships exist among actors? Who performs what roles: instructor, researcher, writer, editor, gatekeeper, guru? And it entails learning about *network practices* around these networks, e.g., how are relationships to others signaled (university affiliations; association memberships; review committee memberships) and to others’ ideas signaled (e.g., citations, weblinks, RSS feeds)? How are contributions evaluated and contributors rewarded for their work?

A prime example of a heavyweight knowledge community is academia. It should not be a surprise given the kinds of overhead listed above that it takes so long to become a member of an academic community (earning a doctorate and academic tenure), and also why such a community can be expected to cling to its existing recognition and reward practices (peer review, tenure review, book and print productions). As noted above, it has been a strain for tenure committees to adopt new online publication practices, and, at present, that change accepts only the change in medium rather than any substantive change in practice. The peer-reviewed online journal is now (generally) accepted, but the place of a website, blog or wiki is undetermined. How then is a tenure committee to deal with an island in Second Life, software of any kind (outside engineering and computer science), collaborative publishing ventures (e.g., a crowdsourced, wiki produced paper), or products with no fixed state (an ongoing wiki site)?

So, why do academics place their work online? What motivates the individual to distribute their ideas freely and openly, and to create platforms for others to do the same (e.g., the Public Knowledge Project: <http://pkp.sfu.ca/>)? What is the relationship between individual practices in dissemination of personal work products and their knowledge community?

I suggest that the concept of coorientation to a community’s values provides a strong and enduring base for community behavior. Individuals spend time, energy, and personal commitment to the working of a community. Hence it is not surprising that it endures beyond day to day activity. During times of change, we may get a rare insight into the enduring principles that underpin a community.

Academic response to the Internet in relation to academic publishing seems to be just such an opportunity. For the academic case, I suggest the answer to the questions above can be found in the “academic calculus” performed by each academic as they consider where, how and when to publish. One part of the calculus equation is the extent to which the individual adheres to principles of *open science and open access* – i.e., their personal stance on whether information should be free. ‘Free’ is interpreted both in Richard Stallman’s open source sense of ‘free as in free speech,’ i.e., free to use, reuse, modify, and transfer, and in the academic, education and library sense of ‘free as in free beer,’ i.e., available easily, universally and without cost to readers. Willinsky¹⁸ articulates well how the confluence of open source, open access, and open science sets the stage for the use of the Internet as an open platform for publishing.¹⁹ Relevant to our discussion of communities, we owe another debt to 17th century science for our orientations to open source. In a recent article, Steve Shapin,²⁰ describes how the habit of the 17th century scientist was to work with a community of interest rather than a set of co-workers. This we see living on in contemporary academic practices – cross-university disciplinary associations and participation in peer review – and attitudes – the loyalty, orientation, and *coorientation* to providing open access to work products, whether these are ideas, theories, experimental results, software, data or courses, and whether wrapped in book covers or web frames.

Until very recently, publishing companies and print media were that access point. They took on the burden of collecting, editing, printing, promoting and disseminating academic texts. But somewhere recently they took a wrong turn. First, journals became more expensive. Then, when journals went online, pricing schemes changed for academic libraries from one-time payment for a permanent resource to yearly payments for ephemeral resources (as a response to the ephemeral part, see the LOCKSS initiative (Lots of Copies Keep Stuff Safe), www.lockss.org/lockss/home). To protect budgets, libraries looked to cut subscriptions. However, to protect their stakeholders, they were obliged to keep them. This dilemma created a movement among librarians to make their stakeholders aware of the costs of publishing in particular journals, and initiated the developing practice of maintaining institutional repositories (see the SPARC site, <http://www.arl.org/sparc/about/index.shtml>, and in particular the copyright addendum form for amending authors rights in keeping with principles of open access, <http://www.arl.org/sparc/author/addendum.shtml>). Now academics pause in their calculus to consider what a publication choice costs their institution, and the copyright arrangement they are signing.

Another factor in the equation is the competition from all those newbies and non-academics who can create websites, journals and diaries online at the drop of a hat, while academics wait through the cycle of peer review and publication (online or off). Why not just post it online in the first place and be done with it? The development of the creative commons licensing scheme²¹ has been highly important for overcoming a key reason not to post in this way, i.e., that someone would steal the idea. The creative commons licensing provides a way to say just how free you want your newly posted online text to be, and is a highly important development that promotes online dissemination of works.

All this online posting has actually created another change in the publishing landscape, and thus in the academic landscape. With peer review, only those who pass review can be published, hence publication alone indicates acceptance into the community and recognition by the community. But, when anyone can post to the web, the scales tip to make *retrieval* the important dimension. Information doesn’t so much want to be free as it wants to be found, cited, linked to, referenced and indexed. And thus, the online venues again become more valuable in an

academic calculus as papers and information reach others earlier and more easily when online, are retrievable by search engines, and gain visible popularity through use.

The long investment in joining a community – including, for academics, an understanding of the nuances of publishing calculus – also highlights the amount of work that goes into defining the boundaries and practices of knowledge domains, and thus how knowledge communities face yet another contemporary challenge, that of multi- and interdisciplinary work, and the merging and combining of knowledge across disciplines.²² Treating this topic in depth is beyond the scope of this paper, but it is interesting to observe that while ‘the world is [becoming] flat,’ it is also becoming closer – we have more frequent contact, interaction and engagement with different places, cultures, and disciplines. And frequency affects attention. It drives our ‘attention structures,’ i.e., that which we observe, value, and count as outcomes,²³ and it builds stronger ties between people, places and discipline.²⁴

In academic communities, we are not only more tied to other disciplines directly in interdisciplinary endeavors, particularly with computing – and even the humanities are joining up with new cyberinfrastructure initiatives directly related to issues of “what to do with a million books”²⁵ – we are also tied by exposure to the practices of other disciplines. When physicists easily accept publication of preprints and online papers and computer scientists consider the conference paper as the finished product, social science researchers take notice and ask why do we have to keep our papers offline, unpublished, uncirculated before, during and after publication in traditional venues? And humanities scholars notice that when books cannot get published within the timeframe for getting tenure at a university, other forms of evaluation must step in, such as journal articles. How long will it be before the practice of accepting journal articles turns to the practice of accepting online publishing? Probably not long. It is not that online venues don’t already exist if individuals chose to use them. What is at issue in a *community* is not that other forms of expression and products are not possible, but that only *certain* choices among them are deemed *acceptable*.

Knowledge Crowds

At the other end of the spectrum of contribution and participation is a mode I will refer to as *lightweight*. This is best demonstrated in the restricted contributory behavior associated with lean crowdsourcing projects. These forums exist to draw in contributions, responses and comments, but are configured by site owners to limit the types of input and the visibility of individual contributors and contributions. At their leanest, they ask for a contributor’s action, but not their opinions – e.g., as in clicking on surveys, identifying objects (NASA clickworkers), verifying spellings (Distributed Proofreaders: <http://www.pgdp.net/c/>) – and return only quantitative measures of participation or aggregate summaries of responses. In these highly lean forms, individuals interact with the computer, not with each other, and the site owners retain authority and control over the acceptance or rejection of submissions, and what is done with them.

But crowdsourcing also covers a spectrum. From impersonal lean data collection environments to ones that begin to take on the look and feel of communities. Crowdsourcing ranges from applications that appeal for isolated, minimal, discrete, objective and often anonymous contributions to versions that include more personal and social presence. Crowdsourced initiatives include opportunities to contribute in many forms, from the lean clickworkers contribution to the richness of Wikipedia. Through these initiatives, contributors can:

- Provide access to idle computer cycles (BOINC)
- Identify objects (NASA clickworkers; GalaxyZoo)
- Add annotations, and tags to others' content (The Commons on Flickr: <http://www.flickr.com/commons>, including photos from the Library of Congress²⁶; Steve: The Museum Social Tagging Project. http://steve.museum/index.php?option=com_frontpage&Itemid=1>)
- Contribute commentary for the owner of the site or poster of material (books put online for comment before publication; blog commentaries)
- Contribute corrections to others content (e.g., in Wikipedia, open source projects)
- Provide original data for inclusion in online applications (e.g., OpenStreetMap, Google Earth)
- Provide original comments, reviews of online content that addresses the owner/original poster and other participants (e.g., blog communities)
- Contribute original designs which are voted on for production by visitors to the site (t-shirts: <http://www.threadless.com/submit>; shoes: <http://ryzwear.com/>; comics: <http://zudacomics.com/>)¹
- Contribute original contributions for commentary and/or amendment by other visitors to the site (new entries in Wikipedia, original code for Open Source initiatives)
- Acknowledge and interact with others about the site, topic of interest, or collective of interest (e.g., on the Talk pages of a wiki)

At their richest, crowdsourced knowledge communities build on distributed, individually held knowledge and act as a portal through which distributed knowledge can be accumulated into a whole. As well as the well-known example of Wikipedia, there are sites for volunteered geographic information (VGI) where individuals contribute mapping information, corrections and images to such sites as OpenStreetMap,²⁷ and the collaborative world building that takes

¹ A more commercial crowdsourcing of design entails competitions with the company making the decision about what is the winning entry. This is indeed a crowdsourcing way of acquiring input, but one that follows long-standing ideas of open competitions. For example, see the call for redesign of everyday objects by Dwell Magazine: (Retrieved March 26, 2009 from <http://www.dwell.com/contests/innovate-it.html>). What at first looks like the crowdsourcing initiative – a call to anyone with access to the Internet to submit a design – reveals itself as a traditional sort of corporate competition. Note first that the review of the entries is not crowdsourced as in the examples given above: “Entries will be judged by a panel of Dwell editors, who will select what they believe to be the three most intriguing ideas for objects needing re-innovation.” Moreover, note that the ownership of the ideas submitted follows a traditional commercial model – ownership by the company: “By submitting an entry, you grant to Sponsors and their respective successors and assigns an unlimited, worldwide, perpetual license to publish, display, use, exploit, edit the text, adapt, modify, copy, disseminate, post, or dispose of the design, text and other submitted materials online, in print, film, television, or in any other media for editorial, advertising, promotional or other purposes without compensation or notification of any kind to you, except as prohibited by law.”

It would be unfortunate if these kinds of ‘old models now posted to the web’ were held up as great examples of web 2.0 practice. They are not. They have missed the opportunity to engage collaboratively with their participants and readers. They have missed the boat on creative commons licensing.

For a site and organization that “gets” the idea of crowdsourcing from a web 2.0 perspective, see www.burdastyle.com. On this “open sewing site” patterns have been made available online, copyright-free. For their take on “open source sewing” see <http://www.burdastyle.com/help/index/66#entry-2>. From the site: “You’ve arrived to your kind of ‘candy’ store. All patterns, accompanied by instructions and related creations are copyright-free and easy to download. If you have a pattern you are proud of, please share.” (<http://www.burdastyle.com/patterns>). And, “Our pattern catalogue is growing larger by the month. BurdaStyle and our members are diligently adding to the variety of trends, basics and accessories available on our site, all of which are open source (copyright-free).” (<http://www.burdastyle.com/help/index/60#entry-1>). Retrieved March 25, 2009.

place in Second Life. Contributors are distinguished by taking charge of different parts of a knowledge, geographical, or virtual world. Specialized or local knowledge that otherwise would be very difficult for an individual or an organization to gather is brought together in these crowdsourced initiatives.

Contrasting Crowds and Communities

By contrast with virtual communities, crowdsourcing enterprises are distinguished by low barriers to entry, low need for commitment, and an appeal to everyone to contribute what they can. For example, the Distributed Proofreaders site specifically asks for a little at a time:

Remember that there is no commitment expected on this site beyond the understanding that you do your best. Proofread as often or as seldom as you like, and as many or as few pages as you like. We encourage people to do 'a page a day', but it's entirely up to you! We hope you will join us in our mission of 'preserving the literary history of the world in a freely available form for everyone to use'.
(Retrieved March 25, 2009 from: <http://www.pgdp.net/c/>)

And Wikipedia takes a moment to assure the nervous contributor:

Be bold in editing, moving, and modifying articles. Although it should be the aim, perfection is not required. Do not worry about making mistakes. In most cases, all prior versions of articles are kept, so there is no way that you can accidentally damage Wikipedia or irretrievably destroy content. (Retrieved March 27, 2009 from: <http://en.wikipedia.org/wiki/Wikipedia:5P>)

In online communities, members often participate in ways that are similar to those for crowdsourcing, e.g., editing, annotating, adding commentary, entering content, uploading pictures, etc. The difference is that *in communities*, individuals contribute with *full attention to the other members of the community*, with full expectation of timely entry and response, commitment to a contribution that reflects on their status in the community, and a continued expectation of interaction with community members. They add to and engage in discussion that concerns the community, covering both the topic and the development and adherence to internal group practices.²⁸

Overall, the difference between a crowd and a community is not in what the collective does, but in how – or indeed whether – its participants need to pay attention to each other in order for the enterprise to succeed. *Knowledge creating communities* need to be designed to accommodate participants' attention to each other, since they are working through ideas, plans and outcomes in consultation and debate with others; *knowledge aggregating sites* do not need individuals to engage directly with each other, but do need adherence to site norms and practices, e.g., doing their best to contribute in a responsible and trustworthy way. Coorientation becomes important again, and perhaps more important for knowledge crowds. Where there is no interpersonal element as a reason for participation and as a sanction on non-conformity, commitment to the overall goal of the site may be the only motivator for contribution and the only control on appropriate contribution. Thus, an interest in astronomy, design, or software development, or a commitment to open source, open science or open access, may be key for understanding how and why crowds participate in lean initiatives.

Discussion lists, collaborative learning, and many game-oriented settings exemplify knowledge-creating communities, whether oriented to academic or serious leisure pursuits (e.g., Geo-caching: <http://www.geocaching.com/>; music: www.last.fm²⁹). Sites that accumulate contributions from many individuals or many individual points exemplify the knowledge aggregators, such as the clicks on pictures of Mars craters (NASA Clickworkers), pictures and other information on local places (OpenStreetMap, Google Earth), and individuals' genomic

information (23andMe³⁰). The collective becomes a knowledge resource by its accumulated, crowdsourced content. Note again how impersonal crowd behavior is, to the point where contribution can be reduced to actions (clicks, photos, GIS coordinates), not personalities. There are no network stars in a crowd, and indeed, the emergence and recognition of a star signifies a move to a community orientation. For example, Wikipedia contains an acknowledgement of the substantial contributions of Simon Pulsifer (http://en.wikipedia.org/wiki/Simon_Pulsifer). The posting of this as an entry on a site that otherwise does not attribute contributions to individuals (although this may be discoverable via the talk pages, it is not a feature of the encyclopedia entries), marks a change from topic goals to personalities, and thus to the human element in the production.

While these examples refer to contribution of content, another layer of knowledge can also be crowdsourced – that of critique. An example of this is the accumulation of commentaries posted on Digg³¹ that provide a thumbs-up or thumbs-down rating of sites identified as of interest by other Digg members. Also, in the design examples given above, each case uses ratings from site visitors to evaluate entries for production. Open online publishing is ‘critiqued’ by the crowd as Google pagerank or YouTube ratings rise, or as academic articles appear on lists such as CitesULike. Contrast this with academic peer review – an invitation only, private and anonymous practice. An exception is – or was – Stevan Harnad’s open access journal *Psychology* that published articles with open peer commentary; however, publication of the journal is currently suspended.

In considering internal focus (and design for internal focus) as a distinguishing factor between crowd and community collectives, it is relevant to note that it is often hard to keep a crowd from pushing to become a community, and sometimes vice versa. For example, even the simplest count of computer cycles donated to a distributed computing effort have been used to generate competition across individuals, teams and regions of participants. Aggregator sites give totals across various distributed computing applications, providing a visible sign that has been adopted by contributors to view and engage with the contributions of others. As described on the DC-Vault site:

“The DC [Distributed Computing] Vault is the place to compare your team's performance against others, the place to look out for when you plan your next taunting fest, the place you can refer others to and brag about how devilishly high ranked your team is ... or not.” <http://www.dc-vault.com>, Retrieved March 25, 2009

Similarly, while it appears that membership in online communities sets individuals free from the constraints of synchronous, co-located interaction, opening the door for full participation from anywhere, anytime, such participation is often illusory. Individuals can participate in any of these venues in a lightweight manner, dropping in and out as lurkers, peripheral contributors, or intermittent participants. With the increase in types of online knowledge communities, this kind of partial, rotating commitment may be commonplace as a means of managing multiple memberships. Even without rotation, the communities themselves depend on a steady flow of new voices to maintain and promote community knowledge. Thus, existing virtual communities already manage lightweight contributors who are coming to know the topic and to join the community.

Summary

In this paper I have walked through some key distinguishing features of online crowds and communities, describing what differentiates these forms of organization in a knowledge context, with particular attention to the impact on academic knowledge communities. My

investigation of the differentiating factors of crowds and communities is a work in progress, developing as more sites and more kinds of online collectives appear and are identified. The current paper builds on my recent presentations and work articulating the dimensions of collaborative activity along which lightweight, crowdsourced and heavyweight, community-based collectives vary.³² Although it has not been possible within the scope of this paper to indicate all the background, this work also builds on the significant literature about the nature of community both online and off, social networks, epistemic communities, distributed knowledge, group processes, collaboration and collaborative learning, adult and expert learning, online learning, computer-mediated communication and the Internet.

While two forms of online knowledge organization – crowds and communities – have been highlighted in this discussion, the goal has been to articulate the underlying dimensions that define the continuum from crowds to communities rather than to classify any particular venture as one or the other. While individuals may consider themselves to be a strong community member of a knowledge crowd, or just a lurker in a virtual community, what differentiates these collectives is not individual's actions or perceptions, but rather the extent to which either environment provides the means for a more or less engaged knowledge collective. To summarize, three dimensions of activity emerge from this work that distinguish knowledge crowds and communities:

- ***Contribution type, granularity and authentication.*** At the leanest end of the lightweight collectives the contribution is straightforward, with easy to learn rules, coordinated by pooled interdependence of similar contributions; for heavyweight collectives, greater learning is required for contribution, with contributions evaluated by other participants in a peer review process, and the overall collective coordinated through reciprocal interdependence.
- ***Individual to group focus.*** Lightweight collectives are individually oriented, often anonymous, with no tie needed between contributors. Heavyweight collectives require contributors to pay attention to others for coordination, collaboration, and evaluation.
- ***Recognition, reputation, and reward.*** The previous two dimensions provide the framework for understanding individual motivation, reputation and reward systems, and thus also for design to support the aims of light and heavyweight initiatives. The leanest of lightweight collectives can assess contributions only through quantitative measures of contributions and statistical aggregation. Heavyweight collectives require qualitative judgements of contributions, contributors, and internal practices, increasing the attention and adherence to internal standards.

These dimensions help distinguish among current forms of organizing, both in terms of site and project design, and of collective knowledge processes. Attention to these dimensions may help in the design of new knowledge projects, as well as in the appropriate matching of incentives to organizational form.

Conclusion

This paper has described in brief some basic differences in structures and individual motivations for engaging in knowledge crowds and communities. The Internet has opened up a vast new space for the formation of knowledge collectives. To date, the emphasis has been on online communities, virtual associations of like-minded individuals with common interests in a knowledge domain. However, recent successful ventures, from NASA's clickworkers to Wikipedia have shown important new ways of harnessing the knowledge of individuals into a

crowdsourced whole. As we look to future knowledge collectives, both lean lightweight collectives and richly nuanced heavyweight communities will offer new and different means and options for knowledge aggregation, evaluation, collaboration and dissemination.

Acknowledgements

Earlier versions of ideas about the organization of crowd and community based enterprises were presented at the The Annual Conference for Print Internet and Community, Tel Aviv, Israel, June 2006; the Southern Illinois-Edwardsville University Library and Information Services Spring Symposium, Edwardsville, IL, April 2008; and at the Hawaii International Conference on System Sciences, Big Island, HI, January 2009.

References

- American Historical Association (2001). *Suggested Guidelines for Evaluating Digital Media Activities in Tenure, Review, and Promotion ~ An AAHC Document*, American Historical Association. Retrieved March 24, 2009.
<http://www.historians.org/perspectives/issues/2001/0110/0110pro1.cfm>
- Andrews, R. & Haythornthwaite, C. (2007). *Handbook of E-learning Research*. London: Sage.
- Baym, N. & Ledbetter, A. (2009). Tunes that bind?: predicting friendship strength in a music-based social network. *Information, Communication and Society*, 12(3).
- Baym, N. K. (2000). *Tune in, log on: Soaps, fandom and online community*. Thousand Oaks, CA: Sage.
- Budhathoki, N.R., Bruce, B.C., and Nedovic-Budic, Z. (2008). Reconceptualizing the role of the user of spatial data infrastructure, *GeoJournal*, 72(3-4), 149-160.
- Crane, D. (1972). *Invisible colleges: Diffusion of knowledge in scientific communities*. Chicago: University of Chicago Press.
- Crane, G. (2006). What do you do with a million books? *D-Lib Magazine*, 12(3). Retrieved March 27, 2009. <http://www.dlib.org/dlib/march06/crane/03crane.html>
- Daft, R. L. & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management Science*, 32(5), 554-571.
- Davenport, T. H. & Beck, J. C. (2001). *The attention economy: Understanding the new currency of business*. Harvard Business School Press.
- DeSanctis, G. and Poole, M. S. (1994). Capturing the complexity in advanced technology use: Adaptive structuration theory. *Organization Science*, 5(2), 121-47.
- Estabrook, L. with B. Warner (2003). *The Book as the gold standard for tenure and promotion in the humanistic disciplines*. Committee for Institutional Cooperation, IL. Retrieved March 24, 2009. http://www.cic.net/Libraries/Reports/ScholarlyCommunicationsSummitReport_Dec03.sflb
- Goldhaber, M. H. (1997a). The attention economy and the net. *First Monday*, 2(4). http://www.firstmonday.dk/issues/issue2_4/goldhaber/
- Goldhaber, M. H. (1997b). *The Attention Economy*.
- Goodchild, M.F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69, 211-221.
- Haklay, M., and Weber, P. (2008). OpenStreetMap--User-Generated Street Map. *IEEE Pervasive Computing*, 7(4), 12-18.
- Halavais, A. (2009). Do dugg diggers digg diligently? feedback as motivation in collaborative moderation systems. *Information, Communication and Society*, 12(3).

- Hauben, M. & Hauben, R. (1997). *Netizens: On the history and impact of Usenet and the Internet*. Los Alamitos, CA: IEEE.
- Haythornthwaite, C. & Bregman, A. (2004). Affordances of persistent conversation: Promoting communities that work. In C. Haythornthwaite & M. M. Kazmer (Eds.), *Learning, Culture and Community in Online Education: Research and Practice* (pp. 129-143). NY: Peter Lang.
- Haythornthwaite, C. (2006). Articulating divides in distributed knowledge practice. *Information, Communication & Society*, 9(6), 761-780.
- Haythornthwaite, C. (2007). Social networks and online community. In A. Joinson, K. McKenna, U. Reips & T. Postmes (Eds.), *Oxford handbook of Internet psychology* (pp. 121-136). Oxford, UK: Oxford University Press.
- Haythornthwaite, C. (2009). Crowds and communities: Light and heavyweight models of peer production. *Proceedings of the 42nd Hawaii International Conference on System Sciences*. Los Alamitos, CA: IEEE Computer Society. Available online at: <https://www.ideals.uiuc.edu/handle/2142/9457>
- Haythornthwaite, C., Kazmer, M.M., Robins, J. & Shoemaker, S. (2000). Community development among distance learners: Temporal and technological dimensions. *Journal of Computer-Mediated Communication*, 6(1). <http://jcmc.indiana.edu/vol6/issue1/haythornthwaite.html>
- Haythornthwaite, C., Lunsford, K.J., Bowker, G.C., & Bruce, B. (2006). Challenges for research and practice in distributed, interdisciplinary, collaboration. In C. Hine (Ed.), *New Infrastructures for Science Knowledge Production* (pp. 143-166). Hershey, PA: Idea Group.
- Hjørland, B. & Albrechtsen, H. (1995). Toward a new horizon in information science: Domain analysis. *Journal of the American Society for Information Science*, 1995, 46(6), 400-425.
- Howe, J. (2006). The rise of crowdsourcing. *Wired* 14.06 (June 2006). Retrieved March 25, 2009. <http://www.wired.com/wired/archive/14.06/crowds.html>
- Inside Higher Ed (December 30, 2005). Radical change for tenure. Retrieved March 24, 2009. <http://www.insidehighered.com/news/2005/12/30/tenure>
- Jenkins, H., with Clinton, K., Purushotma, R. Robinson, A. J., & Weigel, M. (2006). *Confronting the Challenges of Participatory Culture: Media Education for the 21st Century*. Chicago, IL: MacArthur Foundation. http://digitalllearning.macfound.org/atf/cf/%7B7E45C7E0-A3E0-4B89-AC9C-E807E1B0AE4E%7D/JENKINS_WHITE_PAPER.PDF
- Kanfer, A., Haythornthwaite, C., Bruce, B. C., Bowker, G., Burbules, N., Porac, J., & Wade, J. (2000). Modeling distributed knowledge processes in next generation multidisciplinary alliances. *Information Systems Frontiers*, 2(3/4), 317-331.
- Knorr-Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge*. Cambridge, MA: Harvard Univ. Press
- Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Philadelphia, PA: Open University Press.
- Lessig, L. (1999). *Code: and other laws of cyberspace*. NY: Basic Books.
- Lessig, L. (2006). *Code: Version 2.0*. NY: Basic Books.
- Monge, P. R. & Contractor, N. S. (2003). *Theories of Communication Networks*. Oxford, UK: Oxford University Press.
- Orlikowski, W. J. (2002). Knowing in practice: Enacting a collective capability in distributed organizing. *Organization Science*, 13(3), 249-273.
- Perrow, C. (1970). *Organizational Analysis: A Sociological View*. Monterey, CA: Wadsworth.

- Poole, M. S. & DeSanctis, G. (1990). Understanding the use of group decision support systems: The theory of adaptive structuration. In J. Fulk & C. W. Steinfield. (Eds.), *Organizations and Communication Technology* (pp. 173-193). Newbury Park, CA: Sage.
- Preece, J. & Maloney-Krichmar, D. (Eds.) (2005). Online Communities, *JCMC*, 10(4). Articles 1-10.
- Preece, J. (2000). *Online communities: Designing usability and supporting sociability*. New York: John Wiley & Sons.
- Preston, C. J. (2008). Braided learning: an emerging process observed in e-communities of practice. *International Journal of Web Based Communities*, 4(2), 220 – 243.
- Raymond, M. (2008). My friend Flickr: A match made in photo heaven. *Library of Congress Blog*. Posted on: January 16th, 2008. Retrieved March 25, 2009
<http://www.loc.gov/blog/?p=233>.
- Renninger, A. & Shumar, W. (Eds.) (2002). *Building virtual communities: Learning and change in cyberspace*. Cambridge, UK: Cambridge University Press.
- Rheingold, H. (2000). *The Virtual Community: Homesteading on the Electronic Frontier* (revised edition)., Cambridge, MA: MIT Press.
- Scola, N. (2009). Crowdsourcing the Genome. *Seed*, February 27, 2009. Retrieved March 24, 2009. http://seedmagazine.com/content/article/crowdsourcing_the_genome
- Scott, W. R. (1992). *Organizations: rational, natural, and open systems*. Toronto: Prentice-Hall.
- Shapin, S. (1995). Here and everywhere: Sociology of scientific knowledge, *Annual Review of Sociology*, 21, 289-321.
- Shapin, S. (2008). The scientist in 2008. *Seed*. November 20, 2008. Retrieved March 25, 2009.
http://seedmagazine.com/stateofscience/sos_feature_shapin_p1.html
- Sproull, L. & Kiesler, S. (1991). *Connections: New ways of working in the networked organization*. Cambridge, MA: MIT Press.
- Wellman, B. & Haythornthwaite, C. (Eds.) (2002). *The Internet in Everyday Life*. Oxford, UK: Blackwell Publishers.
- Wellman, B. (2001). Physical place and cyber place: The rise of personal networking. *International Journal of Urban and Regional Planning*, 25(2), 227-252.
- Wellman, B. (Ed.) (1999). *Networks in the Global Village*. Boulder, CO: Westview Press.
- Wellman, B., Salaff, J., Dimitrova, D., Garton, L., Gulia, M., & Haythornthwaite, C. (1996). Computer networks as social networks: Collaborative work, telework, and virtual community. *Annual Review of Sociology*, 22, 213-238.
- Willinsky, J. (2005a). The unacknowledged convergence of open source, open access, and open science. *First Monday*, 10(8). http://firstmonday.org/issues/issue10_8/willinsky/index.html.
- Willinsky, J. (2005b). The access principle: The case for open access to research and scholarship. Cambridge, MA: MIT Press.
- Willinsky, J. (2009). Toward the design of an open monograph press.
<http://pkp.sfu.ca/node/1895>.

Biography

Caroline Haythornthwaite is a Professor at the Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign. Her research examines how the Internet and computer media support and affect work, learning and social interaction, with a focus on how information and knowledge are shared through social networks and how collaborative practices are facilitated through information technologies. Her studies have examined social networks of work and media use, the development and nature of community online, distributed knowledge processes, and transformative effects of the Internet and web 2.0 technologies on learning and collaborative practice. Current work focuses on identification of characteristics of online communities, and automated processes to support comparison across online environments. Major publications include *The Internet in Everyday Life* (Blackwells, 2002, with Barry Wellman), *Learning, Culture and Community in Online Education* (Peter Lang publishers, 2004, with Michelle M. Kazmer), and the *Handbook of E-learning Research* (Sage, 2007, with Richard Andrews). For further information see:
<http://people.lis.illinois.edu/~haythorn/> or <http://haythorn.wordpress.com/>

¹ Lee Sproull and Sara Kiesler (1991). *Connections: New ways of working in the networked organization*. Cambridge, MA: MIT Press.

² Steve Shapin (1995). Here and everywhere: Sociology of scientific knowledge, *Annual Review of Sociology*, 21, 289-321.; John Willinsky (2005a). The unacknowledged convergence of open source, open access, and open science. *First Monday*, 10(8).
http://firstmonday.org/issues/issue10_8/willinsky/index.html.

³ Diane Crane (1972). *Invisible colleges: Diffusion of knowledge in scientific communities*. Chicago: University of Chicago Press.

⁴ Inside Higher Ed (December 30, 2005). Radical change for tenure. Retrieved March 24, 2009. <http://www.insidehighered.com/news/2005/12/30/tenure>; American Historical Association (2001). *Suggested Guidelines for Evaluating Digital Media Activities in Tenure, Review, and Promotion ~ An AAHC Document*, American Historical Association. Retrieved March 24, 2009. <http://www.historians.org/perspectives/issues/2001/0110/0110pro1.cfm>

⁵ Leigh Estabrook, with B. Warner (2003). *The Book as the gold standard for tenure and promotion in the humanistic disciplines*. Committee for Institutional Cooperation, IL. Retrieved March 24, 2009.
http://www.cic.net/Libraries/Reports/ScholarlyCommunicationsSummitReport_Dec03.sflb

⁶ For a review, see Caroline Haythornthwaite (2007). Social networks and online community. In A. Joinson, K. McKenna, U. Reips and T. Postmes (Eds.), *Oxford handbook of Internet psychology* (pp. 121-136). Oxford, UK: Oxford University Press.

⁷ For example, Barry Wellman, Janet Salaff, Dima Dimitrova, Laura Garton, Milena Gulia and Caroline Haythornthwaite (1996). Computer networks as social networks: Collaborative work, telework, and virtual community. *Annual Review of Sociology*, 22, 213-238. ; Barry Wellman (Ed.) (1999). *Networks in the Global Village*. Boulder, CO: Westview Press. ; Jenny Preece (2000). *Online communities: Designing usability and supporting sociability*. New York: John Wiley & Sons.; Alaina Kanfer, Caroline Haythornthwaite, Betram (Chip) Bruce, Geoffrey Bowker, Nicholas Burbules, Joseph Porac, and James Wade (2000). Modeling distributed knowledge processes in next generation multidisciplinary alliances. *Information Systems Frontiers*, 2(3/4), 317-331. ; Caroline Haythornthwaite, Michele M. Kazmer,

-
- Jennifer Robins and Susan Shoemaker (2000). Community development among distance learners: Temporal and technological dimensions. *Journal of Computer-Mediated Communication*, 6(1). <http://jcmc.indiana.edu/vol6/issue1/haythornthwaite.html>; Barry Wellman (2001). Physical place and cyber place: The rise of personal networking. *International Journal of Urban and Regional Planning*, 25(2), 227-252. ; Wanda J. Orlikowski (2002). Knowing in practice: Enacting a collective capability in distributed organizing. *Organization Science*, 13(3), 249-273. ; Ann Renninger and Wes Shumar (Eds.) (2002). *Building virtual communities: Learning and change in cyberspace*. Cambridge, UK: Cambridge University Press. ; Barry Wellman and Caroline Haythornthwaite (Eds.) (2002). *The Internet in Everyday Life*. Oxford, UK: Blackwell Publishers.; Caroline Haythornthwaite and Alvan Bregman (2004). Affordances of persistent conversation: Promoting communities that work. In C. Haythornthwaite and M. M. Kazmer (Eds.), *Learning, Culture and Community in Online Education: Research and Practice* (pp. 129-143). NY: Peter Lang.; Jenny PreeceJ. and D. Maloney-Krichmar (Eds.) (2005). Online Communities, *JCMC*, 10(4). Articles 1-10.; Richard Andrews and Caroline Haythornthwaite. (2007). *Handbook of E-learning Research*. London: Sage.
- ⁸ Howard Rheingold. (2000). *The Virtual Community: Homesteading on the Electronic Frontier* (revised edition)., Cambridge, MA: MIT Press.; Nancy Baym (2000). *Tune in, log on: Soaps, fandom and online community*. Thousand Oaks, CA: Sage.; Christina Preston. (2008). Braided learning: an emerging process observed in e-communities of practice. *International Journal of Web Based Communities*, 4(2), 220 – 243.
- ⁹ Jeff Howe (2006). The rise of crowdsourcing. *Wired* 14.06 (June 2006). Retrieved March 25, 2009. <http://www.wired.com/wired/archive/14.06/crowds.html>
- ¹⁰ Michael Hauben and Rhonda Hauben (1997). *Netizens: On the history and impact of Usenet and the Internet*. Los Alamitos, CA: IEEE.
- ¹¹ Henry Jenkins, with Clinton, K., Purushotma, R. Robinson, A. J., and Weigel, M. (2006). *Confronting the Challenges of Participatory Culture: Media Education for the 21st Century*. Chicago, IL: MacArthur Foundation. http://digitalllearning.macfound.org/atf/cf/%7B7E45C7E0-A3E0-4B89-AC9C-E807E1B0AE4E%7D/JENKINS_WHITE_PAPER.PDF
- ¹² Hauben and Hauben, *Netizens*, 1997
- ¹³ Richard L. Daft and R.H. Lengel (1986). Organizational information requirements, media richness and structural design. *Management Science*, 32(5), 554-571.
- ¹⁴ Charles Perrow (1970). *Organizational Analysis: A Sociological View*. Monterey, CA: Wadsworth.; Daft and Lengel, *Management Science*.
- ¹⁵ Caroline Haythornthwaite (2009). Crowds and communities: Light and heavyweight models of peer production. *Proceedings of the 42nd Hawaii International Conference on System Sciences*. Los Alamitos, CA: IEEE Computer Society. Available online at: <https://www.ideals.uiuc.edu/handle/2142/9457>
- ¹⁶ Bruno Latour (1987). *Science in action: How to follow scientists and engineers through society*. Philadelphia, PA: Open University Press.; Karin Knorr-Cetina (1999). *Epistemic cultures: How the sciences make knowledge*. Cambridge, MA: Harvard University Press.; Birjer Hjørland and H. Albrechtsen (1995). Toward a new horizon in information science: Domain analysis. *Journal of the American Society for Information Science*, 1995, 46(6), 400-425.;

-
- ¹⁷ Caroline Haythornthwaite (2006). Articulating divides in distributed knowledge practice. *Information, Communication & Society*, 9(6), 761-780.
- ¹⁸ Willinsky, *First Monday*.
- ¹⁹ See also John Willinsky (2005b). The access principle: The case for open access to research and scholarship. Cambridge, MA: MIT Press.; John Willinsky (2009). *Toward the design of an open monograph press*. <http://pkp.sfu.ca/node/1895>.
- ²⁰ Steve Shapin (2008). The scientist in 2008. Seed. November 20, 2008. Retrieved March 25, 2009. http://seedmagazine.com/stateofscience/sos_feature_shapin_p1.html
- ²¹ Lawrence Lessig (1999). *Code: and other laws of cyberspace*. NY: Basic Books.; Lawrence Lessig (2006). *Code: Version 2.0*. NY: Basic Books.
- ²² Haythornthwaite, *Information, Communication & Society*; Caroline Haythornthwaite, Karen Lunsford, Geoffrey Bowker and Bertram Bruce (2006). Challenges for research and practice in distributed, interdisciplinary, collaboration. In Christine Hine (Ed.), *New Infrastructures for Science Knowledge Production* (pp. 143-166). Hershey, PA: Idea Group.
- ²³ W. Richard Scott (1992). Organizations: rational, natural, and open systems. Toronto: Prentice-Hall.; Goldhaber, M. H. (1997a). The attention economy and the net. *First Monday*, 2(4). http://www.firstmonday.dk/issues/issue2_4/goldhaber/; Goldhaber, M. H. (1997b). *The Attention Economy*.; Thomas H. Davenport and J.C. Beck. (2001). *The attention economy: Understanding the new currency of business*. Harvard Business School Press.
- ²⁴ Peter R. Monge and Noshir S. Contractor (2003). *Theories of Communication Networks*. Oxford, UK: Oxford University Press.
- ²⁵ Gregory Crane (2006). What do you do with a million books? *D-Lib Magazine*, 12(3). Retrieved March 27, 2009. <http://www.dlib.org/dlib/march06/crane/03crane.html>
- ²⁶ Raymond, M. (2008). My friend Flickr: A match made in photo heaven. *Library of Congress Blog*. Posted on: January 16th, 2008. Retrieved March 25, 2009 <http://www.loc.gov/blog/?p=233>.
- ²⁷ Goodchild, M.F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69, 211-221.; Haklay, M., and Weber, P. (2008). OpenStreetMap--User-Generated Street Map. *IEEE Pervasive Computing*, 7(4), 12-18.; Nama R. Budhathoki, Bertram Bruce, and Zorica Nedovic-Budic (2008). Reconceptualizing the role of the user of spatial data infrastructure, *GeoJournal*, 72(3-4), 149-160.
- ²⁸ Marshall Scott Poole and Gerardine DeSanctis (1990). Understanding the use of group decision support systems: The theory of adaptive structuration. In Janet Fulk and Charles W. Steinfield. (Eds.), *Organizations and Communication Technology* (pp. 173-193). Newbury Park, CA: Sage.; Gerardine DeSanctis and Marshall Scott Poole (1994). Capturing the complexity in advanced technology use: Adaptive structuration theory. *Organization Science*, 5(2), 121-47.
- ²⁹ Nancy Baym and Andrew Ledbetter (2009). Tunes that bind?: predicting friendship strength in a music-based social network. *Information, Communication and Society*, 12(3).
- ³⁰ Nancy Scola (2009). Crowdsourcing the Genome. *Seed*, February 27, 2009. Retrieved March 24, 2009. http://seedmagazine.com/content/article/crowdsourcing_the_genome
- ³¹ Alex Halavais (2009). Do dugg diggers digg diligently? feedback as motivation in collaborative moderation systems. *Information, Communication and Society*, 12(3).
- ³² Haythornthwaite, *HICSS*