The Polymath Project: Lessons from a Successful Online Collaboration in Mathematics

Justin Cranshaw
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA, USA
jcransh@cs.cmu.edu

Aniket Kittur
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA, USA
nkittur@cs.cmu.edu

ABSTRACT

Although science is becoming increasingly collaborative, there are remarkably few success stories of online collaborations between professional scientists that actually result in real discoveries. A notable exception is the Polymath Project, a group of mathematicians who collaborate online to solve open mathematics problems. We provide an in-depth descriptive history of Polymath, using data analysis and visualization to elucidate the principles that led to its success, and the difficulties that must be addressed before the project can be scaled up. We find that although a small percentage of users created most of the content, almost all users nevertheless contributed some content that was highly influential to the task at hand. We also find that leadership played an important role in the success of the project. Based on our analysis, we present a set of design suggestions for how future collaborative mathematics sites can encourage and foster newcomer participation.

Author Keywords

large-scale collaboration, online collaborative mathematics, online collaborative science, online communities

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: Miscellaneous

General Terms

Human Factors

INTRODUCTION

The Internet has made creative collaboration possible on an unprecedented scale: tens of thousands of people work together to consolidate knowledge (e.g. Wikipedia), or to write software (e.g. Apache and Linux). However, there have been few examples of scientific knowledge creation being done by large online groups. Tasks such as scientific discovery are difficult to carry out on a large scale for a number of reasons: they are complex, they require significant expertise and prior knowledge, and they have high coordination demands. Furthermore, professional advancement in science is typically subject to a highly developed and closed-loop reward structure (e.g., publications, grants, tenure). These factors do not align well with the open, asynchronous, independent, peer production systems that have typified social collaboration online. However, one notable exception is the Polymath Project, a collection of mathematicians ranging in expertise from Fields Medal winners to high school mathematics teachers who collaborate online to prove unsolved mathematical conjectures.

There are many reasons to believe why mathematics might be well suited for large-scale online collaboration. Mathematicians, unlike most empirical scientists, require little experimental equipment to do their work, making the asynchronous nature of online collaborations less of a deterrent. In addition, mathematical ideas transfer between mathematicians, especially between those who specialize in the same sub-field, with remarkable speed and efficiency, which might help minimize coordination costs [30].

In many ways, the deductive nature of proofs and theorems is inherently collaborative. Although proofs themselves are typically produced by a small number of primary contributors, they often rely on a rich history of prior contributions to the problem made by past mathematicians. Gian-Carlo Rotta, in reference to Wiles's proof of Fermat's last theorem describes this notion of proof as "a triumph of collaboration across frontiers and centuries [28]." This pattern is a common one in mathematics. As Luca Trevisan explains, works of "genius" in mathematics typically occur when "a large collection of 'incremental' results and 'observations' has been made, and all the pieces are there, ready to be put together [31]." These series of incremental results are often the work of a great many mathematicians sometimes spanning decades or even centuries. The mathematical "genius" is then often one who excels at putting these pieces together. According to Trevisan, massively-collaborative mathematics could re-define this idea of genius. Instead of the few who are able to "put together the zeitgeist," genius could rather "be in the union of many minds, each doing nothing more than saying what is obvious to them [31]."

In this paper we combine social science theory, with an in-depth descriptive analysis of data gathered from the Polymath Project, in order to explicate the factors that contributed

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.
Copyright 2011 ACM 978-1-4503-0267-8/11/05...$10.00.

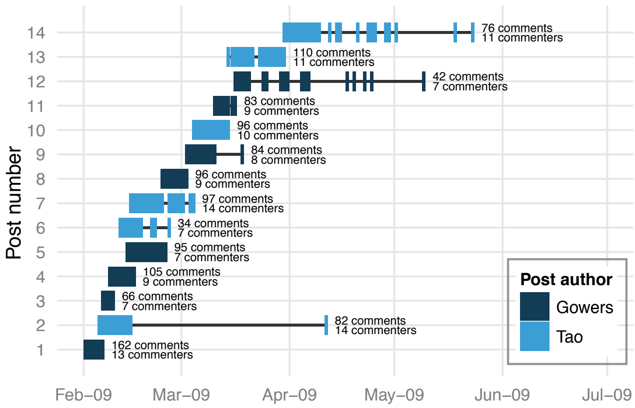

Figure 1. The time span of active comments for each blog post by blog author in the Polymath1 project. Between February and May of 2009, there were 14 blog posts by two authors, Timothy Gowers and Terence Tao. Vertical bars on the timeline indicate days of activity. During this time, a total of 39 contributors left a total of 1228 numbered comments.

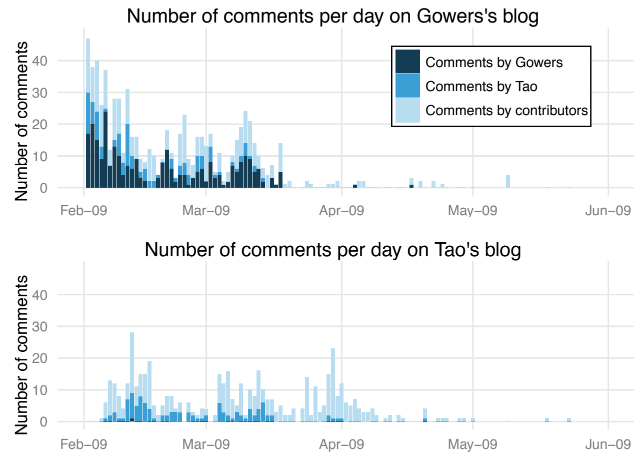

Figure 2. Number of comments per day on each blog, separated by comments from each of the two leaders, and the contributors of their blogs.

to its success. We also discuss design implications for technical tools and social structures that may promote large scale scientific collaborations in the future.

The main contributions of this work are threefold. First, we find that although a large percentage of the Polymath content is produced by small number of individuals, most participants, even those only who only contribute a few comments, nevertheless produce high quality content relevant to the ongoing task. Additionally, we examine the role of leadership in Polymath, both qualitatively and quantitatively, and find that comments from leaders have a greater effect at spurring new discussion than non-leader comments. Finally, we look at the burden that newcomers have in getting acclimated, and provide a set of design suggestions that might lessen this burden and help increase the number of sustained contributors to the project.

RELATED WORK

Although scientists have worked together for centuries [11], science is becoming progressively more collaborative, with teams increasingly dominating the production of knowledge [33]. These teams are also becoming increasingly distributed, with labs spread across continents and spanning multiple institutions [13]. However, despite the push by researchers and funding agencies to share resources and expertise across geographic and institutional boundaries [10], distributed collaborations impose high coordination costs that can have a significant negative effect on the likelihood of success [7, 8]. These costs are rooted not only in geographical distance and time differences but also in institutional culture and norms, and differences in reward structure [8].

There have been a number of notable success stories of using the Internet to crowdsource parts of the scientific research process. For example, NASAs ClickWorkers program had hundreds of thousands of volunteers classifying craters on Mars' surface, resulting in work virtually indistinguishable from expert geologists [18]. Other examples of citizen science include eBirds, a volunteer bird observation network [29]; GalaxyZoo, which has provided visual classifications for nearly one million galaxies [22]; and FoldIt, in which volunteers play an online game help scientists to predict protein folding [5]. Even science-based R&D is being crowdsourced by companies such as Innocentive, who tap into a network of more than 120,000 scientists around the world to solve problems too difficult or labor intensive for internal R&D labs; over 400 problems have been solved to date [21].

In contrast to these successes, large scale online collaboration between established scientists has been rare. The Internet has made it easier to span geographical and time boundaries, with technologies such as email, blogs, and wikis supporting asynchronous communication around the globe. However, this has led to large scale collaborative scientific successes. There are some notable exceptions, including the MIT Registry of Standard Biological Parts, which has spurred multiple scientific papers [24]; and the Gene Ontology project, which aims to provide a dynamic controlled vocabulary for biology [1]. These exceptions, however, focus on sharing and standardizing resources for individual researchers rather than supporting large scale collaborations between researchers.

In this context Polymath is truly remarkable in that it has succeeded at large scale collaboration between scientists. Understanding the principles underlying its success and how they might be improved and applied more generally is an important area of research. While there have been a few studies qualitatively describing the Polymath Project to date [2, 15]), to our knowledge ours is the first study to take a quantitative approach to investigating the project.

BACKGROUND OF THE POLYMATH PROJECT

The Polymath Project began when mathematician Timothy Gowers published a blog post with the provocative title "*Is massively collaborative mathematics possible?*" [14]. Of course, as Gowers pointed out, collaboration on a large scale is not a new phenomenon to mathematics. Hundreds of mathematicians produced over 500 publications in the second half of the 20th century in a successful collaborative attempt to the classify all the finite simple groups in abstract algebra. Although this was indeed a massive collaboration, the nature of the collaboration was quite different from what Gowers was proposing, since in the case of finite simple group classification, the larger task being solved had hundreds if not thousands of natural and predefined subtasks whose solutions were largely independent of the solutions to the others subtasks. Gowers was instead proposing to collaboratively solve a single problem that "does not naturally split up into a vast number of subtasks."

Gowers outlines three *potential* advantages for such large-scale collaborations in mathematics. First there is the role of chance in problem solving. Having more people work on a problem increases the odds that one person will get lucky and discover a great insight. Second, different mathematicians have different areas of expertise, so having a large number of contributors work on a problem creates a collective expertise that cannot be achieved by small groups of contributors. Finally, Gowers argues that different people have different characteristics, and will thus assume different roles in the problem solving process. Some contributors will be proficient at coming up with new ideas or approaches, others will perform complex calculations, still others will "fact check" the work of their piers. In short, having a large pool of people working on the problem allows for a greater opportunities for people to specialize in the problem solving process. These advantages are succinctly summarized by Gowers:

> If a large group of mathematicians could connect their brains efficiently, they could perhaps solve problems very efficiently as well. [14]

Of course there are risks to this approach too, most notably the possibility that an individual might use material being developed on the blog to individually publish a great breakthrough.

Organization of the Polymath Project

Following Gowers's initial problem proposal, the community rapidly organized itself in order to begin work on solving the problem. Contributors made use of numerous and distinct web-tools available to them to fulfill their needs. The resulting suite of tools that they assembled was highly decentralized and spread across at least 5 Internet domains. We describe these components of the Polymath Project below.

Blogs of Noted Mathematicians

The primary venue for participants to work on the Polymath problems is within the posts and comments of the personal blogs of noted mathematicians. Initial work began on Gowers's blog just days after his initial post. On February 1st, mathematician Terence Tao wrote a general post discussing Gower's initiative, and shortly after that Tao started an official post working on Polymath from his blog as well. Since then other mathematicians have also on occasion hosted Polymath work on their blogs.

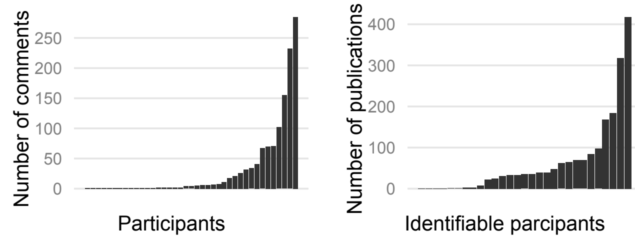

Figure 3. Right: Number of Google Scholar publications of each of 29 the users who tied their blog identity to their real names, sorted from lowest to highest by number of publications. Left: Count of numbered-comments left by the 39 users, sorted from lowest to highest by comment count.

The Polymath Wiki

A wiki for the project was created and hosted by physicist Michael Nielsen[1]. The wiki primarily serves as a clearinghouse for background material, summaries, and preliminary write-ups relevant to work being done on the blogs. Each problem the community decides to work on has its own devoted page on the Polymath Wiki page where material related to that problem is posted.

The Polymath Blog

A dedicated blog was created primarily for administrative discussions, and for high level summaries of the progress being made on the other blogs[2]. Additionally, some of the discussions about what problems the community wishes to work on are carried out on the Polymath Blog.

The Ground Rules for Participation

In many ways Gowers's proposal was an experiment. For the experiment to be a success, it would not be enough to simply prove a theorem via blog comments. It is instead, the way the problem is ultimately solved that is essential to the experiment's success. For instance, if one contributor decided to work independently from the group continuously for days and solved the problem on his own, this would clearly not be an example of massively collaborative mathematics. To minimize the chances for this type of failure, and to generally set the tone of the project, Gowers proposed a set of 12 ground rules. Many were intended to guide the nature of the comments that participants produce to maximize the chance that the experiment succeeds. Several rules explicitly discourage collaborators from working extensively on their own on the problem without discussing their progress on the blog. Additionally, there are rules that are designed to make sure the discourse remains polite and focused, while also not discouraging comments that may not be fully thought out. Finally the last ground rule specifies that any publications that result from work done on the blog will be published under a pseudonym, and will contain links to the full discourse for readers to examine.

[1] http://michaelnielsen.org/polymath1
[2] http://polymathprojects.org/

The Polymath Projects

Currently there are five official Polymath Projects (denoted by Polymath1 through Polymath5). Each of these projects focuses on a significant unsolved problem in mathematics. These five projects make up the core of the serious work being done by the group. In addition, there are also two Mini-Polymath Projects, which are intended mainly as puzzle exercises and are not serious open problems, and a wiki page dedicated to the discussion of the purported proof by Deolalikar that $\mathbf{P} \neq \mathbf{NP}$ [9]. To date, on the 5 Polymath Projects and 2 Mini-Polymath Projects, there have been over 275 unique commenters, who have left over 5000 comments containing well over 20000 equations.

POLYMATH1

In this work, our analysis will focus almost entirely on the Polymath1 project, which was the first, and to date most successful of the Polymath Projects. At the time of this writing, Polymath1 was the only project that was complete, making it a particularly appealing artifact to study in its entirety as an example of successful collaboration.

The target problem for Polymath1 was to find a combinatorial proof of the density version of the Hales-Jewett theorem, a theorem in combinatorics. Although proofs exist, they have relied on techniques from ergodic theory [12] and are non-combinatorial. There is often great value in mathematics to finding multiple proofs for one theorem. In particular, elementary proofs (ones that rely only on first principals) such as what Gowers was proposing, are often most valued for their instructiveness. As G. H. Hardy said, "the history of mathematics shows conclusively that mathematicians do not evacuate permanently ground which they have conquered once [17]."

Success from a mathematical point of view in Gowers's words would be "anything that could count as genuine progress towards an understanding of the problem." In this strict sense, Polymath1 was successful well beyond Gowers's expectations. To date there are two journal articles that have been submitted [26, 25]. Additionally, at least two new proofs of the theorem, plus new bounds on Hales-Jewett numbers were discovered. In this work we will examine whether and to what extent Polymath1 was successful in the larger sense, that is as a collaborative framework for doing higher mathematics on a large scale.

Participation and demographics

Between February 1 and May 23, 2009, there were 1555 comments written on the 14 blog posts of Polymath1. These 14 blog posts were distributed across two mathematics blogs; 8 of which were written by Timothy Gowers[3] and 6 were written Terence Tao[4]. Of the comments, 1228 were assigned a sequential and unique "comment number" by the participants so that they could be easily and succinctly referred to by subsequent comments. The comments that received a number typically were direct contributions to the proof attempt, whereas comments that did not receive a number were typically administrative in nature, commenting on the process, rather directly contributing to the proof. We call these two types of comments *numbered comments* and *meta-comments* respectively.

[3]http://gowers.wordpress.com

[4]http://terrytao.wordpress.com

Previous day:	b	β	p-value
Comments by contributors	0.24	0.24	0.044
Comments by leaders	0.35	0.41	< 0.001

Table 1. The extent to which comments on the previous day spur contributor comments on the current day. Raw b-coefficient, standardized β-coefficient estimates and p-values are shown from a multiple linear regression. The two independent variables are the number of comments by the two leaders and the number of comments by the contributors on the previous day. The dependent variable is the number of comments by the contributors on the current day.

Since we wish to study the act of proving a theorem collaboratively, we will primarily focus on numbered comments and the participants who write these numbered comments. This is not to say the meta-comments are without value. On the contrary, some meta-comments played a critical role in the organization of the discourse. Rather, we will primarily treat numbered-comments as quantitative data, and meta-comments as qualitative data.

In total 39 distinct users contributed at least one numbered comment to Polymath1. Although, this number seems small compared to the thousands of users who actively edit popular Wikipedia articles, it is large in comparison to the number of authors who traditionally collaborate on the proof of a theorem. The number of comments per user is shown in Figure 3 (left). Similar to other online content producing communities, a majority of the content is produced by a very small number of users. Indeed the top three commenters produced 55% of the comments, and the top ten commenters produced nearly 90% of the comments.

Of these 39 users, 29 linked their online profiles to their true identity (the remaining 10 were either anonymous, or we were unable to map their account to a true identity). Their range of backgrounds varies widely, from Fields Medalists winners to high school mathematics teachers, however the distribution of their backgrounds is relatively homogeneous; almost all of the participants are either university professors or graduate students in mathematics or computer science and almost all of them are male (28 male, 1 female).

What does vary fairly drastically among these 29 is their level of seniority in the higher-mathematics community as measured by number of academic publications. To measure this, we use the Google Scholar author search restricted to mathematics, computer science, engineering, and physics articles. Figure 3 (right) shows the number of Google Scholar publications for each of the 29 identifiable participants. The median number of publications is 35, and the mean is 65.2. It is also notable that there are 4 participants with 0 publications, 6 with 1 publication or less, and 8 with 5 publications or less, showing that there is a mix of well established mathematicians, and those who are early in their career.

THE ROLE OF LEADERSHIP

Research in online content producing communities has shown that leadership can be essential to the production of high quality content [27, 23]. We find that this is also true of the Polymath1 project.

From its inception, leadership was essential to the Polymath Project. The idea was initially conceived and implemented by Timothy Gowers, a well know mathematician and Fields Medal winner. It is natural then that Gowers would take on a number of leadership duties in the Polymath1 project. In addition to setting the guidelines for participation, and selecting the problem for Polymath1, it was Gowers's intention to host the work in the comments section of his mathematics blog. This made him best positioned to be the person that guides the progress on the problem, if for any other reason than he has administrative privileges on the blog and others do not. It was Gower's role to summarize past progress and guide future progress with each new blog post.

Very soon after Gower's initial post, fellow Fields medalist Terance Tao wrote a blog post discussing Gowers's project that quickly would become a second thread to Polymath1, running in parallel to Gowers's first post (see Figure 1). Like Gowers, this naturally positioned Tao to take various leadership roles, both on his blog, and on Polymath1 in general.

Though there were certainly other participants that at various times showed leadership, Gowers and Tao were leaders in Polymath1 throughout. For this reason, we will refer to Gowers and Tao as the *leaders* of Polymath1, and we refer to the collective readership of their blogs as the *contributors*.

Blog posts summarizing progress

In a meta-comment near comment number 350, Tao proposes that every post should be limited to 100 comments. Although this proposal was out of necessity, as it allowed threads to run in parallel while still being numbered sequentially (since the numbers for a thread could be allocated in advance), there were important unforeseen consequences. Ending a thread after 100 comments allowed for the leaders to make fairly frequent summary posts, highlighting what they thought was relevant in the previous thread and guiding the contributors on the next thread. This proved to be one of the essential roles of the leaders throughout the project.

Leader comments versus contributor comments

Aside from this very important role of summarizing the progress, the leaders also made direct contributions to the proofs by posting comments. On their own Gowers and Tao were the top two commenters on Polymath1. Gowers contributed 285 numbered comments, and Tao contributed 232 numbered comments, thus in total there were then 517 leader comments, which was 42% of all numbered commented (readers contributed a total of 711 numbered comments).

Judging by volume alone, it is clear that Gowers and Tao made significant contributions to the work of the proof. However we hypothesize that their comments were themselves very important in guiding and driving the work of the contributors. To address this hypothesis, we examined whether or not the leader comments on any given day actually spurred increased contributor activity on the next day. To control for the fact that comments in general on any given day (both from leaders and contributors) could spur increased contributor activity on the next day, we compare the effects of leader comments to contributor comments. Table 1 shows the results of a multiple regression, where the two independent variables are the number of comments by the leaders and the number of comments by the contributors on the previous day. The dependent variable is the number of comments by the contributors on the current day. This shows that contributor and leader comments are both significant factors in spurring activity on the next day. However leader comments have a much greater effect on spurring new comments ($\beta = 0.41$ for leader comments, compared to $\beta = 0.24$ for contributors).

COORDINATION AND THREADING

Collaboration among many people on a task with a potentially large set of interdependencies can require a great deal of coordination. Here we look for evidence of coordination and interdependencies by examining the attempts of users to parallelize their efforts.

Threading

The decision to have a fixed number of comments (roughly 100) on any single blog post was consequential in opening the possibility of parallel blog threads (i.e. two or more blog posts with active comments). Setting a fixed number of comments per post allowed the the space of comment numbers for any post to be allocated in advance, thus maintaining a sequential numbering system on parallel posts.

At the same time, very early in the comments of the first post, parallel discussions began to develop on different possible proof approaches. Participants attempted to use the limited tools made available to them by the blog site to organize these discussions. They annotated the subjects of comments with either a thread name, or a list of relevant comment numbers. It was clear very early that there was certainly potential for some parallelism in discussing distinct approaches to the problem.

Dual threads: Gowers and Tao

This potential for parallel discussions was partially realized when Tao and the readers of his blog began working on Polymath1 in parallel to Gowers. In Figure 1, note the high degree of temporal overlap between work being done on Gowers's weblog and that being done on Tao's weblog.

Despite this overlap, it is difficult to claim that this is evidence of coordination and organization, since the work being done by Tao and his readers was related but distinctly different from the work being done on Gower's blog. Tao and his readers were trying to compute exact bounds on Hales - Jewett numbers for small dimensions, while Gowers and his readers were attempting to prove the general theorem.

In general there was little evidence of any coordination constraints between work being done across the two blogs, which allowed them to operate almost indendependently.

However, aside from the parallelism of the dual Gowers / Tao posts, there is very little evidence of any successful parallel threads despite explicit attempts to parallelize both blogs. Parallelism was attempted at least once by Gowers. He intended that post number 4 proceed in parallel to post number 3 on a separate but related topic. However, note in Figure 1 that work on post 3 ends soon after the creation of post 4, and well before the allocated 100 comments for that post. In a meta-comment on the success of threading, Gowers remarks

> Im not completely convinced that the multiple-threads idea has worked. Or rather I think it has partially worked. It seems that the upper-and-lower bounds thread was sufficiently separate that it has thrived on its own. But the experience of the other two threads was that once the 400 thread was opened the 300 thread died pretty rapidly.

Tao had some success in parallelizing his blog, as post 6 was active in parallel to each of post 2 and post 7. However, the nature of post 6 was quite different than the others, as it was designed to be an interactive "reading group" to discuss background material. Aside from this one instance, all other posts comprised one linear thread.

Although parallel threading was not as successful as the participants had hoped, there were unintended benefits to this approach. Gowers, best summarizes this: "An advantage of the multiple threads was that we had multiple summaries: I think it would be quite good to be forced to summarize more often." In this sense having multiple threads was a success, though not in the way originally envisioned.

The comment reference graph

To examine the nature of dependencies in the comments, we constructed a directed graph based on comment references. Nodes of the graph are numbered comments, and directed edges are placed from node b to node a if comment b explicitly mentions the comment number for a. There were 655 edges, 702 non-isolated vertices, and 135 weakly connected components of the comment reference graph. The mean component size was 5.2, the median was 2, and the standard deviation was 25.5. The largest component contained 299 comments. The second largest had 15 comments. This large gap in size suggests the largest component might have comprised the "main discussion" and the remaining components were side discussions.

Figure 4 shows a visualization of the structure of the comment references with respect to the 14 blog posts. The columns indicate the post number of the originating comment, and rows indicate the post number of the referenced comment. Thus the diagonal shows the number of intra-post comment references, and off the diagonal shows inter-post comment references. The size of each bar is proportional to the number of such references, and the color of the bar is according to the originating post author. Figure 4 illustrates that there was a fair amount of comment referencing, highlighting the importance and usefulness of the sequential comment numbering scheme. Although inter-post comment references did exist, the majority of comment references were intra-post, suggesting that the discussion was largely localized and focused. This is further quantitative evidence of the linear nature of the work, perhaps suggesting that comment references might have been constrained by the linearity of the blog.

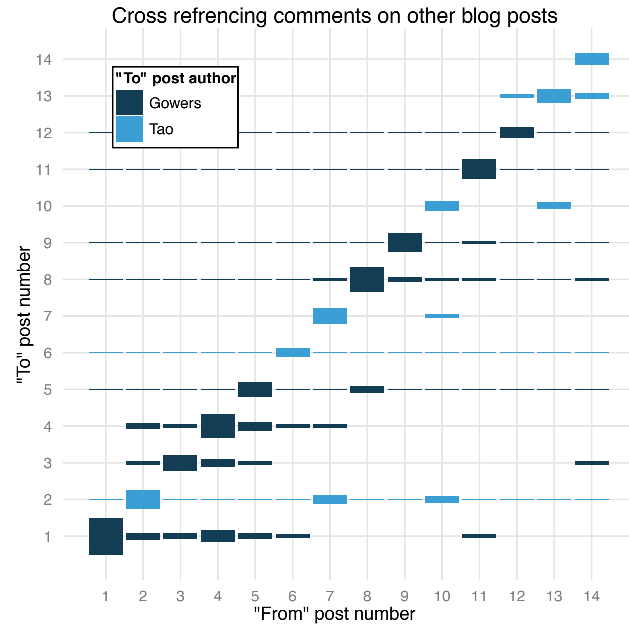

Figure 4. A visualization of the structure of the comment reference graph with respect to inter- and intra-post comment references. Along the diagonal shows the number of intra-post comment references. Off of the diagonal shows inter-post references. The size of the bars is proportional to the number of such comments.

THE BURDEN TO NEWCOMERS

It has been shown that socializing newcomers is an essential component to sustaining a healthy and productive online community [4, 3]. The level of technical sophistication of the Polymath content and the pace at which progress on problems moves means that it is necessary to have measures in place to allow newcomers to quickly and efficiently get up to speed with the material.

Figure 5 shows the cumulative number of contributors to Polymath1 over time, split by the number of comments the users would ultimately contribute to the project. This figure shows that the core set of contributors started contributing to the project very early. By the end of the first week of the project, over 2/3 of the contributors who would go on to produce at least 10 comments had already joined. This could suggest that once the project started, the technical nature of the generated content made it difficult for newcomers to join. This hypothesis is supported by the meta-comments. Gowers, at one point remarks

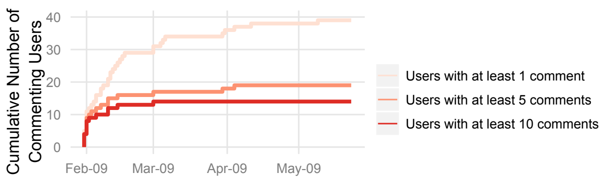

Figure 5. Cumulative number of users who join the project grouped by total number of comments.

> A number of people have commented that it is difficult to catch up with the discussion if you come to it late. We should try to start each thread in a way that makes it as self-contained as possible.

There is additional evidence that the work done on Gower's blog required a significant amount of background knowledge in combinatorics in order to participate, in contrast to the problem on Tao's blog, which was more elementary. On this issue, one participant remarked:

> Regarding the amount of background knowledge required, while on the thread here it did stray through intimidating territory, [...] on Terry's blog the arguments [...] were relatively elementary.

This burden may have discouraged a number of qualified individuals who were interested in the project from actually participating. This also may explain why there were more individuals contributing to Tao's blog. Over the course of the project 32 people contributed to Tao's blog, 20 contributed to Gowers's blog, with 14 contributing to both.

In addition to the burden of keeping up with the fast pace, and the burden of mastering the required background knowledge, an additional barrier to participation emerged from the meta-comments. One participant was discouraged by the large volume of comments that were not fully "thought out" that nevertheless required a significant amount of time to validate. Despite making some important contributions early on, he concluded his time was "better spent on problems that do not get as much attention."

All of this evidence suggests that the project could have benefited from targeted mechanisms to aid and encourage newcomers to join and continue to participate.

DOES EVERYONE REALLY CONTRIBUTE?

As we noted, there were many users who only contributed a few comments to the project. Indeed, there were 13 participants who only contributed 1 comment, and 21 participants contributed 5 comments or less. This begs the questions, did they all really contribute to solving the problem? Can one or two comments from a user who never comments to the project again really have any impact on the final solution when there are over 1200 comments in total? Perhaps all but the top few contributors were irrelevant?

To investigate these questions, we first turn to the official Polymath1 timeline[5], which was created on the wiki to highlight the comments that were important milestones to the proof. We will use the timeline to to better understand which comments were important as judged by the participants themselves. In total there were 215 numbered comments listed in the timeline, from 22 distinct users (18% of all numbered comments, 56% of all users). Out of the users who only commented 5 times or less, there were 6 comments listed in the timeline from 5 distinct users (17% of the 36 comments from these users, 23% of these users). Thus it was not uncommon for users who commented very infrequently to nevertheless have a large impact on the proof.

Comment graph centrality

Of course a comment does not need to be a milestone in the proof for it to be a meaningful contribution. For instance, imagine if someone leaves a comment that by itself is not a milestone, but it links together several other contributions in an important way. Or perhaps, the user leaves a comment that directly contributes to the the ideas of a subsequent milestone comment. To look for evidence of such contributions, we can examine the structure of the comment reference graph.

We first compute a series of graph centrality measures on each node of the comment reference graph. The centrality measures we use are: eigenvector centrality, betweenness centrality, authority score, Bonpow centrality, closeness centrality, alpha centrality, page rank, in degree, and total degree. We next compute the ranking on the vertices that is induced by each centrality measure. Finally, we aggregate the rankings by averaging them to produce a single ranking of the 1228 comments.

Figure 6 is a visualization of three variables that are essential to understanding the nature of participation and contribution in Polymath1. Each of the identifiable users are plotted according to the number of comments they contributed to to the project, and the maximum ranking of all of their comments. Additionally, the size of each point is proportional to the number of publications of that user according to Google Scholar.

There are a number of observations that one can immediately make from this Figure. First, the leadership of Gowers and Tao can be seen instantly. Not only did Gowers and Tao have the highest, and second highest number of comments respectively, they produced the second highest and third highest ranking comments respectively under our ranking method. Additionally, it can be seen that there is a third potential leader, that had the highest ranked comment, and the third highest number of comments overall. Like Gowers and Tao, in addition to being a prolific contributor, this individual took on various administrative roles in the project, including being in charge of one of the final write-ups.

In addition to observing leadership in the upper right quadrant, we can see that there are a number of contributors in the bottom right quadrant. These are individuals who were

[5]http://michaelnielsen.org/polymath1/index.php?title=Timeline

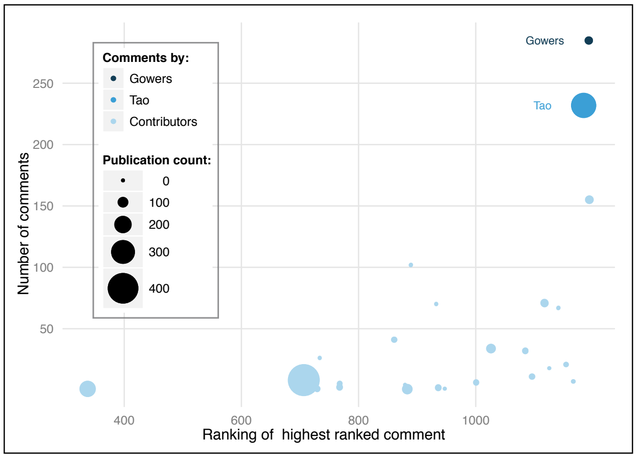

Figure 6. Here we show the ranking of the highest ranked comment for each user together with the total number of comments from that user. Points are sized proportional to the number of scholarly publications of that user (as given by Google Scholar).

not prolific contributors, but who nevertheless made significant, highly ranking comments. Finally, using the number of Google Scholar publication as a proxy for seniority, one can see that the level of seniority of the participants varies greatly. Furthermore, if you discount Tao and Gowers, it does not seem to correlate to volume or importance of contribution.

To put this finding in more precise terms, the 662 comments that received highest ranking (those above the median) were produced by 28 users. Conversely, there were only 11 users whose comments all ranked less than the median ranking. We feel this result is quite surprising given the volume of comments written by the top contributors (recall 90% of the comments were written by the top contributors).

DISCUSSION

Was Polymath1 a success? From a mathematical point of view, the answer is unequivocally, yes. In just over 6 weeks a collective group of distributed individuals accomplished some highly non-trivial mathematical feats. Whether or not the project was a success as an experiment in "massive-scale" collaboration, requires a more nuanced discussion.

Certainly, the scale of the project could hardly be considered "massive," falling short of the vision originally put forth by Gowers [15]. However, it is clear that the scale was large compared to most collaborations in higher math. It is relatively uncommon for a mathematics publication to have more than 3 authors, let alone 40. Additionally, although the core set of contributors by volume was small, we have shown that several users were still able to make important contributions to the work, despite producing a small percentage of the comments.

The ambiguity regarding the success of Polymath1 is amplified by the roles played by the two leaders. Not only were Gowers and Tao by far the top contributors to the project in terms of number of comments, but they also each undertook a significant number of administrative duties in order to guide the project to its mathematical success. Additionally, it is evident from the number of popular press articles written about Polymath, that their collective celebrity had a large role in both the publicity the project received and in the recruitment of contributors. Their involvement was so significant all throughout, that it is questionable whether the Polymath project would have succeeded without their efforts. Whether or not large-scale collaborative mathematics can succeed without such high-profile leaders is of course an open question.

Incentives for participation

In order to maintain a diverse community, incentives for participation should be such that all levels of mathematicians have incentive to participate. This may be difficult to maintain in practice unless the individuals are able to get proper academic attributions for the work they do in Polymath. There is a potentially large misalignment in incentives between the demands of a professional academic career in mathematics, and long-term involvement in the Polymath Project. This misalignment is best expressed by an anecdote told by a Polymath contributor during meta-discussion. While having a conversation about the Polymath with an offline colleague, the colleague asked "how will participants put their grant acknowledgment lines" in the final document?

This begs several other questions regarding the sustainability of long-term involvement. How do the universities that employ the participants receive attribution? Could someone justify to their university that they deserve tenure if all of their work is published collectively? Unlike work in open source software, where there are significant economic incentives at play at the heart of these communities, it's unclear that the economics of truly large scale collaborative mathematics are scalable absent the right incentive structure.

Design recommendations for encouraging newcomers

We have discussed several barriers to entry that might discourage newcomers from participating. Here we focus on three challenges: identifying the important content to read, identifying tasks to work on, and learning the required background material.

Identifying important content

We have shown that one potential problem discouraging newcomers is the large volume of content that is generated by existing users, together with the lack of any mechanisms that sort out important content from unimportant content. Although it is inevitable that newcomers must read some existing content in order to become acclimated, when so much of the work done in mathematics is validating ideas that may or may not work, participants may be discouraged by reading content that has already been invalidated by others. Newcomers may feel that they are simply replicating work that has already been done by the existing users in order to get acclimated to the project. Thus in addition to the cost associated to with reading a potentially large volume of irrelevant content, there is the added cost associated with the redundancy of this task.

We have two proposals to address this challenge. First, there should be better explicit linkage among the comments utilizing the explicit comment referencing done by the participants. For any comment, there should be automated tools that generate and display all the comments that refer to it. This might be accompanied through a visualization of the comment reference graph. This mechanism will allow newcomers to see where each comment fits into the larger body of work. Second, there should be mechanisms for users to rate the importance of each comment. Ideally this would include both a numeric rating of importance, such as a "star" mechanism, as well as a categorical importance rating where old-timers could, for example designate a given comment as "Important for getting up to speed" or "A core contribution." One might also pair socially generated ratings, with algorithmically generated ratings, such as the graph-based comment ranking that we employed in this work.

Identifying tasks to work on

Similar structural problems make it difficult for newcomers to identify tasks to work on. Research on task identification in Wikipedia suggests that automated mechanisms can significantly improve productivity [6]. The manifestation of this challenge in Polymath was outlined in Gowers & Nielsen

> A significant barrier to entry was the linear narrative style of the blog. This made it difficult for late entrants to identify problems to which their talents could be applied. There was also a natural fear that they might have missed an earlier discussion and that any contribution they made would be redundant. [15]

They suggest using a task management system similar to those in use by open source software communities. To some extent the participants attempted to use the wiki for these purposes. However, without the explicit assignments of tasks to people, this likely had little affect. We agree that a task management system would be of use here, however we strongly suggest that such a system be designed to have explicit tasks for acclimating newcomers. These tasks need not be content producing tasks, but could be administrative tasks, or even tasks designed to get the newcomer up to speed. We suspect that task assignment, or automated task suggestion might not be useful for experienced Polymath contributors, and could even discourage them, since they are more familiar with the content and their own abilities, and thus may not work well with the rigidity of a task assignment system.

Dealing with background material

Finally, it is natural that many Polymath projects will require a significant amount of background material to participate, as is often the case in communities of practice. However, this should not be so prohibitive that it precludes someone with sufficient technical abilities from participating. The community has attempted to address this issue by suggesting longer "lead times" for the start of new projects, for potential participants to research the background material. Additionally, they experimented with "reading group" blog threads. These are both temporary solutions. Ideally there would be some structural mechanism in the design of the collaborative website that deals with this issue. One approach would be to have a forum section of the website which is designed to discuss and ask questions related to background material.

This would ideally be integrated so that primary content could link and refer back to the forum in a structured way. This would help with the problem of identifying what background material should be studied to understand a given piece of primary content. For the most part, the wiki was useful as a clearinghouse for background material, however the lack of integration between the wiki and the primary content being produced on the blogs was likely prohibitive. Like comments, different pieces of background material have differing and evolving levels of importance. Tools that connect the primary content back to the secondary content and visa versa, could significantly reduce the burden of newcomers by helping them focus on what background material they need to understand most to contribute to any given area.

CONCLUSION

In this work, we examine the Polymath Project, an experiment in large scale collaborative mathematics, to better understand its successes and shortcomings as a online community and tool for open collaboration. We find that a large percentage of the content is produced by small number of individuals. However, many individuals that produce a low volume of content are still producing high quality, influential content. We also examined the role of leadership, and the burden to newcomers. Based on our analysis, we gave a set of design recommendations for encouraging newcomers.

REFERENCES

1. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. Gene ontology: tool for the unification of biology. the gene ontology consortium. *Nature genetics*, 25(1):25–29, May 2000.

2. M. J. Barany. '[b]ut this is blog maths and we're free to make up conventions as we go along': Polymath1 and the modalities of 'massively collaborative mathematics'. In *WikiSym '10: Proceedings of the 6th International Symposium on Wikis and Open Collaboration*, pages 1–9, New York, NY, USA, 2010. ACM.

3. M. Burke, R. Kraut, and E. Joyce. Membership Claims and Requests: Conversation-Level Newcomer Socialization Strategies in Online Groups. *Small Group Research*, 41(1):4–40, 2010.

4. M. Burke, C. Marlow, and T. Lento. Feed me: motivating newcomer contribution in social network sites. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09, pages 945–954, New York, NY, USA, 2009. ACM.

5. S. Cooper, F. Khatib, A. Treuille, J. Barbero, J. Lee, M. Beenen, A. Leaver-Fay, D. Baker, Z. Popovic, and F. Players. Predicting protein structures with a multiplayer online game. *Nature*, 466(7307):756–760, August 2010.

6. D. Cosley, D. Frankowski, L. Terveen, and J. Riedl. Suggestbot: using intelligent task routing to help people find work in wikipedia. In *Proceedings of the 12th international conference on Intelligent user interfaces*, IUI '07, pages 32–41, New York, NY, USA, 2007. ACM.

7. J. N. Cummings and S. Kiesler. Collaborative research across disciplinary and organizational boundaries. *Social Studies of Science*, (35), 2005.

8. J. N. Cummings and S. Kiesler. Coordination costs and project outcomes in multi-university collaborations. *Research Policy*, 36(10):1620–1634, 2007.

9. V. Deolalikar. P neq np, unverified claim, 2010.

10. T. A. Finholt. Collaboratories. *Annual Review of Information Science and Technology (ARIST)*, 36:73–107, 00 2002.

11. T. A. Finholt and G. M. Olson. From laboratories to collaboratories: a new organizational form for scientific collaboration. *Psychological Science*, (8):28–35, 1997.

12. H. Furstenberg and Y. Kaznelson. A density version of the hales-jewett theorem. *J. Anal. Math*, (57), 1991.

13. O. Gassmann and M. von Zedtwitz. New concepts and trends in international randd organization. *Research Policy*, 28(2-3):231 – 250, 1999.

14. T. Gowers. Is massively collaborative mathematics possible? http://gowers.wordpress.com/2009/01/27/is-massively-collaborative-mathematics-possible/.

15. T. Gowers and M. Nielsen. Massively collaborative mathematics. *Nature*, 461(7266):879–881, 10 2009.

16. C. Gutwin, R. Penner, and K. Schneider. Group awareness in distributed software development. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, CSCW '04, pages 72–81, New York, NY, USA, 2004. ACM.

17. G. H. Hardy. Mathematical proof. *Mind*, 38(149):1–25, 1929.

18. B. Kanefsky, N. G. Barlow, and V. C. Gulick. Can Distributed Volunteers Accomplish Massive Data Analysis Tasks? In *Lunar and Planetary Institute Science Conference Abstracts*, volume 32, page 1272, Mar. 2001.

19. A. Kittur and R. E. Kraut. Harnessing the wisdom of crowds in wikipedia: quality through coordination. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, CSCW '08, pages 37–46, New York, NY, USA, 2008. ACM.

20. A. Kittur and R. E. Kraut. Beyond wikipedia: coordination and conflict in online production groups. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, CSCW '10, pages 215–224, New York, NY, USA, 2010. ACM.

21. K. R. Lakhani and J. A. Panetta. The principles of distributed innovation. *Innovations: Technology, Governance, Globalization*, 2(3):97–112, July 2007.

22. C. Lintott, K. Schawinski, S. Bamford, A. Slosar, K. Land, D. Thomas, E. Edmondson, K. Masters, R. Nichol, J. Raddick, A. Szalay, D. Andreescu, P. Murray, and J. Vandenberg. Galaxy zoo 1 : Data release of morphological classifications for nearly 900,000 galaxies. Jul 2010.

23. K. Luther and A. Bruckman. Leadership in online creative collaboration. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, CSCW '08, pages 343–352, New York, NY, USA, 2008. ACM.

24. J. Peccoud, M. F. Blauvelt, Y. Cai, K. L. Cooper, O. Crasta, E. C. DeLalla, C. Evans, O. Folkerts, B. M. Lyons, S. P. Mane, R. Shelton, M. A. Sweede, and S. A. Waldon. Targeted development of registries of biological parts. *PLoS ONE*, 3(7):e2671+, July 2008.

25. D. H. J. Polymath. Density hales-jewett and moser numbers, preprint, available at arxiv.org/abs/1002.0374, 2009.

26. D. H. J. Polymath. A new proof of the density hales-jewett theorem, preprint, available at arxiv.org/abs/0910.3926, 2009.

27. J. M. Reagle, Jr. Do as i do:: authorial leadership in wikipedia. In *WikiSym '07: Proceedings of the 2007 international symposium on Wikis*, pages 143–156, New York, NY, USA, 2007. ACM.

28. G. C. Rota. The phenomenology of mathematical proof. *Synthese*, 111:183–196, 1997. 10.1023/A:1004974521326.

29. B. L. Sullivan, C. L. Wood, M. J. Iliff, R. E. Bonney, D. Fink, and S. Kelling. ebird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, 142(10):2282 – 2292, 2009.

30. W. Thurston. On proof and progress in mathematics. In R. Hersh, editor, *18 Unconventional Essays on the Nature of Mathematics*, pages 37–55. Springer New York, 2006. 10.1007/0-387-29831-2_3.

31. L. Trevisan. A people's history of mathematics. http://lucatrevisan.wordpress.com/2009/02/01/a-peoples-history-of-mathematics/.

32. H. Wickham. *ggplot2: elegant graphics for data analysis*. Springer New York, 2009.

33. S. Wuchty, B. F. Jones, and B. Uzzi. The Increasing Dominance of Teams in Production of Knowledge. *Science*, 316(5827):1036–1039, 2007.