When Distance Doesn't Really Matter: Effects of Geographic Dispersion on Participation in Online Enterprise Communities

Jeffrey Warshaw[1], Steve Whittaker[1], Tara Matthews[2*], Barton A. Smith[2]
[1]University of California at Santa Cruz, CA, USA, {jwarshaw, swhittak}@ucsc.edu
[2]IBM Research - Almaden, CA, USA, taramatthews@gmail.com, barton.smith@us.ibm.com

ABSTRACT
Research on small team collaboration repeatedly shows that "distance matters". More recent work has refined this concept of distance to develop *geographic dispersion measures* to explain the negative effects that team *configuration* has on productivity and interaction. Dispersion measures explain why teams distributed across multiple time zones, or across multiple sites, have more coordination difficulties than collocated teams with a single remote member. Although larger online communities are increasingly used in enterprises, few studies have examined the effects of dispersion on community behavior. We studied 1206 online enterprise communities (OECs), each using a set of collaborative tools. We present new data showing counter-intuitively that OEC dispersion does not affect content generation or contribution inequality, even when restricting community size to those that most resemble small teams (with 3-12 members). We found that previously documented negative effects of geographic dispersion seem to be reduced in enterprise communities regardless of size. Our results provide additional support to prior case studies suggesting that online communities can mitigate geographic dispersion by providing resources that support coordination and resource sharing.

Author Keywords
Online communities; distance; enterprise; geographic dispersion; measures; replication; workplace.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
Many studies show that for small teams "distance matters" [17, 32, 53]. Small teams are work groups with generally under a dozen people, with task and goal interdependency [9, 12]. They usually focus on shared goals with specific deadlines, and interdependent work tasks distributed between members [16,32]. This interdependency means that distributed small teams face more obstacles than collocated ones [51, 53, 61]. Distance between members is associated with delays, duplication of work, reduced ability to negotiate, and member conflict [3, 25, 26, 47, 53]. These findings have led organizations to make strategic decisions about the structure and geographic dispersion of teams [59, 67], and the deployment of CMC technologies [59].

Early work on 'distance effects' drew simple contrasts between collocated versus distributed teams [26, 47, 53], but more recent research has drawn distinctions between different *configurations* of distributed teams [3, 34, 51]. Predominantly collocated teams with a single remote member operate very differently from teams with members spread across multiple locations [51]. These complex team configurations reduce real-time collaboration and access to shared resources [34]. And absolute distance may be less important than time zone differences when teams need to collaborate in real-time [61, 66].

Thus prior work has developed *dispersion measures* that characterize team configurations as well as time zone differences [52]. Dispersion measures have been very successful in explaining the consequences of team configuration for multiple aspects of work productivity [3, 51, 61]. For example, teams with greater geographic dispersion have much higher overall levels of online activity because more interactions are needed to achieve coordination [61]. And teams that are unequally dispersed across different locations show unequal levels of contribution to common group tasks [51]. Finally, isolated members contribute more as they need to speak up in order to be heard over more integrated members [34, 51].

Work on dispersion has largely focused on small teams, and less attention has been paid to the effects of distance and dispersion on online enterprise communities (OECs) who may also experience collaboration challenges. OECs are organizationally-sponsored online communities that serve many important workplace functions. OECs support collaboration, knowledge sharing, reuse of resources, expertise location, innovation, organizational change management, and social networking [43, 45, 49, 62]. OECs have a broader set of goals than internet

CSCW '16, February 27 – March 2, 2016, San Francisco, CA

* Current affiliation: Google Inc.

communities which tend to focus on knowledge sharing, support, and social networking [10, 37].

Dispersion measures offer a new lens for analyzing collaboration in larger groups such as OECs. Prior literature argues that online communities are effective in supporting communication and collaboration across geographic boundaries [48, 68]. Nevertheless we would expect them to be affected by dispersion. As with small teams, we would predict collaboration to become harder, and knowledge sharing to be diluted when OECs disperse across multiple locations and time zones. For example, greater dispersion should induce higher coordination costs leading to lower levels of interaction. Also, unequal numbers of members at different locations should lead to inequalities of contribution, with isolated members in particular contributing more.

Although OECs share the overall workplace context of small teams, their membership, goals, and functions in other ways resemble social media communities (SMCs). SMCs are explicitly-created groups on social media such as Facebook Groups, where participants share information and organize events based around common interests. SMCs are used to distribute news, share materials or organize virtual or physical events such as meet-ups. Compared to small teams, OECs and SMCs tend to be larger and more interpersonally focused. Recent studies have shown few distance effects for participation in SMCs [39, 42].

We use quantitative methods to evaluate opposing hypotheses about dispersion effects in OECs. On the one hand, OECs resemble small teams in their workplace context and focus on active collaboration. OECs might therefore show negative dispersion effects similar to those seen in small teams. Conversely, OECs also resemble SMCs with their larger size, interpersonal focus, and use of social tools. OECs might therefore be protected from negative dispersion effects. Specifically, we test three predictions about dispersion effects in OECs derived from the small teams literature. In each case we expected that dispersion would lead OECs like small teams to suffer dispersion effects. Failing to replicate these effects would imply that OECs like SMCs are resistant to dispersion. These predictions mirror prior dispersion findings:

1. Spatial distance should be positively correlated with overall activity of an OEC, as participants try to address coordination challenges.
2. Imbalance of OEC members across sites should be correlated with greater inequality of contributions by members.
3. The proportion of isolated members in an OEC should be negatively correlated with OEC activity, as isolates find it difficult to understand and contribute to common objectives.

We apply the dispersion measures developed by O'Leary and Cummings [52] to 1206 OECs. Our analyses showed that, unlike small teams, dispersion measures do not predict participation in enterprise communities. Rather, it seems, OECs reduce the negative effects of geographic dispersion, more closely resembling their SMC counterparts. We find that the biggest factor explaining community participation is *size*, not geographic dispersion. These findings provide empirical support for prior work that theorized that dispersion effects might be reduced in OECs due to their greater focus on interpersonal interaction and use of social tools [48, 68]

RELATED WORK

As we have seen, online enterprise communities share features and goals with two different types of online groups: small work teams and SMCs. In this section, we describe prior research on dispersion effects in small teams and contrast that work with the large body of research finding simple distance effects in small teams. We also review distance effects in SMCs and OECs.

Dispersion in Small Teams

O'Leary and Cummings refined prior work that relied on simple measures of distance (e.g., proportion of remote members, collocation as a binary variable) to develop new *dispersion measures* that describe spatial distance, time zone separation, number of sites, imbalance of members across sites, and isolation of individual members [52]. These dispersion measures have strong predictive power for the performance of small teams in multiple settings, and the effects of each measure are distinguishable from each other [13, 19, 20, 22, 24, 34, 51]. Next, we describe each dispersion measure.

Spatial distance is the average pairwise distance between contributors. As spatial distance increases, face-to-face communication is reduced [35], reliance on CMC increases [53], and perceptions of communication quality and team performance decrease [63].

Time separation is the average pairwise difference in time zone hours between contributors. Time separation is associated with increased inability to adjust when problems arise [19, 20]. Teams with fewer overlapping work hours are more prone to delays from communication breakdowns, need for clarification, and rework [19, 20].

Number of sites: As the number of sites increases, knowledge sharing decreases [13, 22], and costs of maintaining control over team goals increases [24].

Site imbalance describes how the number of contributors varies across sites. Imbalanced teams (e.g., 4 members at one site and 2 at another) generally show different levels of conflict and trust compared to balanced teams (e.g., 3-3), though the direction of this effect is inconsistent between studies [51, 58].

Isolation is the percentage of singleton contributors with no other team members at their work site. Isolation effects differ from imbalance effects. Isolated team members are not members of a local subgroup, and therefore the team is motivated to more frequently communicate with the isolate [51]. As a result, isolates have more opportunities to provide a divergent perspective that increases the overall team effectiveness [34].

Distance in Small Teams

The large number of global enterprises and perceived cost reductions to outsourcing have made distributed enterprise collaboration commonplace [2, 3]. However, there are consistent issues associated with distance in small distributed teams compared to small collocated teams. Distributed teams show coordination difficulties because interactions are less frequent without the opportunity for spontaneous face-to-face interactions [19, 53]; decreases in work quality despite greater speed of work [61]; and project management problems as a result of differences in work culture and practices [47]. There are also interpersonal costs, with distributed teams experiencing more conflict [26, 58] and challenges with communication and trust [47, 58] than non-distributed teams.

One potential reason for these difficulties is that geographic distance is a highly salient grouping characteristic, or *faultline* [40]. Research on faultlines in teams has shown that team members make more negative attributions about members on the other side of the faultline [40, 41], including when the faultlines are geographic [58]. These difficulties are not insurmountable but often require costly preparations and efforts at team integration to prevent their negative effects on team performance [47, 54].

Distance in Social Media Communities

By contrast, research on social media communities (SMCs) has claimed few negative effects of distance. SMCs are group spaces that social media users create to interact with others over time, with the group serving to facilitate socialization, organize events, discuss common issues, and share materials among the group [56]. One reason for the lack of distance effects in SMCs may be that SMCs can act as extensions of local, offline communities [18] helping users to maintain relationships they have already established [27]. Prior social history can reduce misunderstandings between group members, enable easier trust formation [11] and reduce conflict [26]. Related work that analyzes Twitter interactions has also focused on simple distance rather than the more detailed dispersion measures [39, 42]. However, an important limitation of most SMC research is that it does not involve comparisons between differently-dispersed communities, but instead draws analytical contrasts between different parts of the same online community.

Features of Online Enterprise Communities

Online enterprise communities (OECs) are different from small teams as well as from SMCs. Key differences between OECs and small teams, are OEC's *larger size*, focus on *interpersonal interaction* in communities of practice (CoPs), and use of newer *social tools* such as blogs and wikis [15, 37, 68]. However OECs are also different from SMCs. In addition to being comprised of an employee population, OECs aim to create collective resources to support work collaboration, and serve greater expertise-finding and coordination functions than SMCs [15, 43, 49]. Although OECs' structure and interpersonal focus overlap somewhat with their social media counterparts, their membership and goals remain qualitatively different from SMCs. As a result, we explore here whether the effects of dispersion on OECs are more similar to small teams or SMCs.

METHOD

We present a quantitative analysis of 1206 workplace OECs. Here we describe the research context and dataset used.

Research Context

This research was conducted in a global enterprise offering technology products and services to businesses. The company widely encouraged employee leadership of, and participation in, internal online communities. The commercial community space technology was available to all employees. All participants we studied used this technology, which enabled communities to easily create an online space combining various social media tools, including forums, blogs, wikis, bookmarks, feeds, file sharing and task management (activities). As a result, there was a proliferation of communities and widespread membership throughout the organization. Prior work in the same organization and using the same tool found that over 75% of employees had joined at least one community over the first four years in which the service had operated [49]. Communities ranged in size from a few contributors to thousands. Many employees were members of multiple communities.

The communities utilizing this tool were diverse in purpose. Prior work examining the goals of communities in the same organization and using the same tool found that communities were primarily focused on member learning, reuse of resources, and collaboration [43]. About half were self-described Communities of Practice (CoPs): learning environments where employees with a shared interest communicate, build relationships, and share resources to assist with work [49,69]. The remainder of communities consisted of self-described teams, which we call community-based teams. These community-based teams overlap with the small teams in prior work in that they collaborate toward a shared deliverable, however they also share overlapping goals with traditional CoPs, including social learning and networking, and they use social tools aimed at communities [43, 48, 49]. Throughout the rest of the paper we will refer to the groups that use this tool as simply *communities*.

Community Data and Measures

We selected a sample of all the communities using the online community space technology, two years after its deployment in the company. We applied a number of filters to ensure that communities were active, had enough members, and where we had accurate location information for the majority of participants.

We wanted to ensure that the communities we analyzed were active, thus we included only communities that had been updated in the last month and included a total of at least 50 posts. This identified a total of 1402 communities, for whom we collected the following data concerning participation:

Posts: We measured the overall productivity of each community by collecting *overall number of posts*, defined as any data added to the community, including posts or comments in the forum, blog, wiki, files, activities, bookmarks, and RSS feeds. The amount of communication within a community has been commonly used in prior work to measure success [10, 16, 29, 30]. Number of posts ranged from 50 to 2842, M = 160.02, SD = 199.70.

Contribution inequality: Additionally, we wanted a measure of the balance of contributions, which captured how posts were distributed across people who contributed to the community. We used *contribution inequality,* measured by the Gini coefficient [29, 31, 36]. A Gini coefficient closer to zero denotes relatively equal contribution by many contributors, whereas a coefficient closer to one indicates a large proportion of posts are generated by a small minority. Contribution inequality is prevalent in online communities, with a small number of active participants dominating discussions [49, 70]. Evidence from online production communities such as wikis suggests communities with greater contribution inequality create higher quality work due to the increased concentration of work by a core group of contributors [5, 33, 55], however work teams perform better when contribution is more equal [21, 38]. Community inequality within this sample ranged from .20 to .81, M = .56, SD = .10.

We used number of posts and contribution inequality as our measures of workplace community success. In addition, we included a single control variable due to its expected relationship to the number of posts and contribution inequality:

Community size: One critical factor in explaining community interaction is the size of the community [10, 60, 69, 70]. We wanted to control for this so we therefore gathered data about the *size,* operationalized as the number of contributors. Prior work [50] proposes different definitions of belonging to a community depending on whether people actively post. That work distinguishes passive from active contributors. Passive contributors may sign up for the community but restrict themselves to lurking. Active contributors are people who make one or more posts. For our analyses, we used active contributors as they are likely to be more engaged in the community. Most prior work on communities has focused on active contributors [6]. We refer to active contributors as members. Community size ranged from 3 to 379, M = 26.06, SD = 35.91.

Geographic Dispersion Measures

Here we operationally define the 5 geographic dispersion variables developed by O'Leary and Cummings [52], which we applied to the sampled OECs. We also describe related research applying and evaluating each individual measure. With one exception, these measures all scale in relation to group size, suggesting they are appropriate for use despite the larger group sizes in our sample. *Number of sites*, which correlates highly with group size, was only included in analyses of groups similar in size to that described in prior research (i.e., < 12 members).

To derive the geolocation of members of the community, we used the corporate directory to determine each member's work location and the street address associated with that location. For employees who worked from home, this returned the closest work location to their home address, and that location was used for the following analyses. The percentage of users for whom we could determine geolocation differed by community, ranging from 2% to 100%. In this paper, we include analyses conducted on communities in which we knew geolocational data for at least 50% of the members, leaving 1206 of the original 1402 communities sampled. Increasing that threshold did not change the significance or predictive power for the following analyses, so we retained the communities for which we had incomplete geolocational data to evaluate a larger proportion of communities.

Spatial distance. The average pairwise great-circle distance between each member of a community, in kilometers. This ranged from 0 km indicating full collocation to 9880 km indicating a community whose four members spanned three continents (M = 3551.53 km, SD = 2596.39 km).

Time separation. The number of time zones between each pair of work sites represented in a community, weighted by the number of members present at each site. This ranged from 0 indicating all of a community's members occupied the same time zone; to 9 within a community where half of its members worked at UTC-8, with the other half at UTC+9 or UTC+10 (M = .64, SD = .93).

	Model 1 (size control)		Model 2 (adding dispersion measures)	
Variable	β	η^2	β	η^2
Size: # of Contributors	.63**	.40	.65**	.29
Spatial Distance			-.05	< .01
Time Separation			.07	< .01
Isolation			-.08	< .01
Site imbalance			-.06	< .01
Overall R^2	.40		.40	
F_{change} for R^2	787.22**		3.25*	

Table 1. *Size* rather than dispersion predicts *posts*. Summary of hierarchical regression analysis for community *size* and geographic dispersion measures predicting number of *posts* (N = 1203). *$p < .05$. **$p < .001$.

Number of sites. The raw numerical count of unique work sites represented in a community, ranging from 1, meaning all community members work at the same work site, to 140 in a community of 379 members (M = 12.64, SD = 15.22). Due to collinearity between the number of members and number of sites variables, we only include the number of sites as a predictor variable in selected analyses below where the number of members is constrained.

Site imbalance. The standard deviation of the count of members at each work site, divided by the total number of members in that community. This ranged from 0 in communities where each work site represented had the same number of community members in that community (e.g., 2-2-2) to .64 in a community where 18/19 members worked at the same site (M = .09, SD = .12).

Isolation. The number of community members who were the only person from their work site represented in a community, divided by the total number of members in that community. This ranged from 0 in communities where no member was the only one from their work site in a community, to 1 in communities where each of its members worked at a different location (M = .53, SD = .32).

RESULTS

We expected differences in community performance based on dispersion measures from [52], but this was not the case. Contrary to our expectations, when we controlled for community size, dispersion measures explained a negligible proportion of the variance in both total posts and contribution inequality. Although contrary to the literature on dispersion effects in small teams [3, 52], this finding supports one of the proposed benefits of online communities—that they overcome the negative effects of geographic dispersion [3, 52]—which has not previously been systematically tested with quantitative OEC data. For each of the following analyses, we note where influential multivariate outliers were removed. Outlier communities were identified by disconnected, large standardized residuals, Mahalanobis distance, and Cook's distance [65].

Finding 1. Dispersion is not associated with the number of *posts* or *contribution inequality* in a community.

Posts. We used an ordinary least squares (OLS) regression to explore the predictive power of the dispersion measures on the number of *posts* per community. To control for the effects of community *size* on *posts*, predictor variables were included in two sequential blocks. Block 1 contained *size* (log-transformed to reduce positive skew). Block 2 introduced the dispersion measures of *spatial distance*, *isolation*, site *imbalance*, and *time separation*. Because larger communities tended to include more separated work sites, we observed high collinearity between the number of *contributors* and *work sites* in each community (r = .81), which was not reduced by centering. As a result, we did not include *work sites* as an independent variable in this analysis. *Posts* was log-transformed to correct for positive skew. From the original 1206 communities, 3 influential multivariate outliers were removed, N = 1203, using methods advocated in [65].

Because the predictor variables we included are still moderately correlated, we report beta (β) and partial eta-squared (η^2) rather than r and r^2 for individual predictors. Beta represents the strength of the correlation between the predictor and dependent variables not accounted for by other predictor variables. Partial eta-squared represents the proportion of variance in the dependent variable explained by the predictor variable that is not explained by other predictor variables in the model.

The results are shown in Table 1. We found a strongly predictive model in Block 1 with *size* as control variable, $F(1,1202) = 787.22$, $p < .001$, adj. $R^2 = .40$. As expected,

	Model 1 (size control)		Model 2 (adding dispersion measures)	
Variable	β	η^2	β	η^2
Number of Contributors	.32**	.10	.35**	.07
Spatial Distance			.02	< .01
Time Separation			.02	< .01
Isolation			.17**	.02
Site imbalance			.14**	.01
Overall R^2	.10		.12	
F_{change} for R^2	138.44**		7.71**	

Table 2. *Size* rather than dispersion predicts *contribution inequality*. Summary of hierarchical regression analysis for community *size* and geographic dispersion predicting *contribution inequality* (N = 1199). *$p < .05$. **$p < .01$.

larger communities tended to have many more posts. However, after controlling for the positive relationship between *size* and *posts*, the four dispersion measures in Block 2 did not meaningfully improve the model, $F_{change}(4,1198) = 3.25$, p < .05, adj. $R^2 = .40$. In this second model, none of the dispersion measures are associated with variation in *posts* (all η^2s < .01).

Contribution inequality. We next examined whether dispersion explained differences in *contribution inequality*. We repeated the OLS hierarchical regression procedure above, with *contribution inequality* as the dependent variable. As before, Block 1 consisted of *size* as a control variable. Block 2 contained *spatial distance*, *time separation*, *isolation*, and site *imbalance*. In this analysis, seven influential multivariate outliers were removed from the original 1206 communities, N = 1199.

Table 2 shows the results of the two-step regression. A tenth of the variance in communities' *contribution inequality* can be explained by their *size* (Block 1, adj. $R^2 = .10$, $F(1,1197) = 138.44$, $p < .001$), suggesting that this is a reasonable explanatory model. As expected, communities that were larger in *size* tended to show greater *contribution inequality*, a finding consistent with previous literature [49]. Again, however, the four dispersion measures only slightly improved the predictive power of Block 2 over Block 1, adj. $R^2 = .12$, $F_{\text{change}}(4,1193) = 7.71$, $p < .001$.

Because the calculations for the dispersion measures are related to community size, we were interested in the explanatory power of the dispersion measures independent of size. This is measured by η^2. In this model, size explains the largest proportion of non-shared variance in Gini ($\eta^2 = .07$), with isolation ($\eta^2 = .02$) and site imbalance ($\eta^2 = .01$) only providing marginally more predictive power. When size is controlled, communities with more isolated contributors and greater imbalance between contributors at different work sites show very small increases in contribution inequality.

These results show a distinct lack of predictive power for dispersion measures, but we wanted to be sure that this was not due to our sampling strategy. One clear divergence from prior work is that we explore large communities. With few exceptions [61], prior research using these dispersion measures has usually focused on teams of fewer than a dozen members. In the next section, we test the effects of evaluating a subset of communities of smaller size to see whether this affects our results.

Finding 2: Even when *size* is restricted to small communities, dispersion measures still do not predict the number of *posts* or *contribution inequality*.

Posts. To reduce the collinearity between community *size* and number of *work sites*, we conducted separate analyses featuring just communities between 3 and 12 contributors (N = 572). We chose these community sizes because they are comparable to the small teams for which the dispersion measures were developed [52]. Other research on OECs [43, 49] shows that smaller enterprise communities are often used by teams with deliverables.

We again conducted an OLS hierarchical regression on the log-transformed number of total posts to control for *size* as in the previous analyses. Again Block 1 contained *size*. Block 2 contained *spatial distance*, *time separation*, *isolation*, site *imbalance*, and, due to less collinearity between *work sites* and *size* than in the prior analyses, we were also able to include *work sites* (log-transformed to correct for positive skew). Three influential multivariate outliers were removed, N = 569. Again, however, neither *size* (adj. $R^2 < .01$, $F(1,567) = 4.37$, $p < .05$) nor the full model containing *size* and all five dispersion measures (adj. $R^2 < .01$, $F(6,562) = 1.49$, $p > .05$) predicted *posts*.

Posts analyzed by tool. To check whether the impact of dispersion on *posts* may have been masked by aggregating all types of *posts* for each tool together, we ran separate regressions predicting the log-transformed *posts* for separate tool types—wikis, forums, and blogs—based on the five dispersion measures, for both the full set of communities and the subset of small communities. All of these models had adj. $R^2 < .14$, with none of the dispersion measures explaining more than 1% of the difference in *posts* for any tool (all βs < .15, all η^2s < .02). For this collection of small communities, dispersion did not affect the amount of content different communities produced in the individual collaborative tools.

Contribution inequality. Using the same method of restricting to communities between 3 and 12 contributors, we ran an OLS hierarchical regression with *contribution inequality* as the dependent variable. Block 1 contained community *size*. Block 2 contained the five dispersion measures: *site index*, *spatial distance*, *time separation*, *isolation*, site *imbalance*, and (again, unlike the large community analyses) *work sites.* Two influential multivariate outliers were removed, N = 570.

Size still predicted *contribution inequality*, adj. $R^2 = .27$, $F(1,568) = 209.70$, $p < .001$. The dispersion measures in Block 2 improved slightly on this model, adj. $R^2 = .29$, $F_{change}(5,563) = 3.47$, $p < .01$. Here, *site index* showed a faint correlation with *contribution inequality* ($\beta = .19$, $t(563) = 2.19$, $p = .05$). However, *site index* explains a low proportion of the non-shared variance in *contribution inequality* ($\eta^2 < .01$), due to its high overlap with *size* ($r = .60$). In addition, none of the other dispersion measures explained more than 1% of the variance in *contribution inequality* for the size-restricted sample. This sample of smaller communities echoes our larger sample results, showing that communities demonstrate greater *contribution inequality* when they are larger in *size,* but do not appear affected by dispersion.

DISCUSSION

We extend research on geographical dispersion [13, 19, 20, 22, 24, 34, 51], providing evidence for claims that online communities can mitigate the impact of geographic barriers [48, 68]. Despite prior evidence of strong negative dispersion effects in small teams, we found no appreciable dispersion effects for enterprise online communities. We examined five dispersion measures deployed in prior work: *spatial distance*, *time separation*, *isolation,* and site *imbalance.* None showed major effects on *contribution inequality* or content *posting*.

At the same time, we found that the *size* of communities did impact participation performance. Larger communities generated more posts but struggled to reduce contribution inequality, compared with smaller communities. These results suggest that communities that desire equal input from all participants should be kept small.

When we restricted our sample to only those communities who most closely resemble the small teams in previous research ($3 \leq$ *size* ≤ 12), we still did not find compelling evidence of dispersion effects. This contrasts with a wide body of work showing negative effects on small teams [13, 19, 20, 22, 24, 34, 51]. Instead our results provide empirical support to prior work hypothesizing that online enterprise communities can reduce the negative impacts of geographic barriers [48, 68]. We describe differences between teams and OECs that may be responsible for the divergent ways that dispersion affects each, and suggest reasons for OECs' lack of dispersion effects.

Communities are Larger, Have Different Goals, and Curate More Content than Teams

Size and Goals. Over half of the communities in our initial sample had 13 or more members. In prior literature, dispersion effects are most commonly reported for teams of six to twelve members [51, 61, 63]. Furthermore, our sample included *communities*, whose goals focused on social learning and building interpersonal relationships [43]. This difference in goals between communities and teams has several implications for how *size* might differentially affect teams and communities. Members of small teams often have interlocking goals with strong dependencies between members' goals [14]. Distributed teams may therefore incur coordination costs as they grow larger.

On the other hand, communities may actually *benefit* as they become larger. Their goals are less tightly intertwined (e.g., on a single shared deliverable), so coordination is less critical than in a team setting [43]. Adding members to a community also provides more potential sources of expertise, increasing its overall value [1]. In addition, having more members means an increased capacity for picking up the slack when others are committed elsewhere. These characteristics have strong implications in terms of anticipated dispersion effects. Dispersion (specifically, *time separation* and *work sites)* exacerbates the coordination costs inherent to dispersed team work [3, 20, 66], whereas communities may lack these strong coordination demands.

Content curation. Another potential explanation for these results may lie in some modern enterprise communities' greater emphasis on *content curation* over *content creation* [44]. OECs benefit not only from their own members' novel contributions, but also by organizing (*curating)* material generated by themselves and other communities [44, 46]. Compared to OECs, the tasks that small teams enact are more likely to focus on *content creation* (e.g., documents, lines of code [61]). Because coordination is more pertinent to content creation compared to curation, dispersion may have less impact on communities focused more on curation than creation. Dispersion Size would continue to be an advantage for communities in this regard. The more members a community has, the greater the number of external resources they should have knowledge of and be able to link to.

Coordination and Resource-Sharing Tools May Reduce Dispersion Effects for Communities

In contrast to many of the small teams studied in prior research, all of the communities in our sample had access to multiple collaborative tools, including tools focused on discussion (i.e., forums, blogs), resource-sharing (i.e., files), task coordination (i.e., activities), and curation (i.e., wikis, RSS feeds, bookmarks). This set of multiple integrated tools may allow them to target specific difficulties they might otherwise face (e.g., coordination and resource-sharing) buffering them against vulnerabilities that dispersion typically causes [8]. For example, using wikis to keep track of project assignments may prevent misunderstandings about individual members' responsibilities. This may be important as teams increasingly shift to using multiple social media tools.

Changing Practices for Facilitating Distributed Work

Our work echoes recent literature emphasizing that dispersion is not insurmountable. Distributed teams that collaboratively set ground rules on team interaction can perform as effectively as face-to-face teams [23]. Team familiarity and increased one-on-one communication can also reduce dispersion effects [21, 34, 66]. The semantic web practice of linking open data to improve reuse of internet resources [4, 7] has been applied successfully in work settings as well [28, 57, 64]. Enabling easier reuse of existing enterprise data has a number of benefits for dispersed collaboration, as it reduces the amount of repeated work and provides greater context for work done at different work sites. Our findings echo this prior research, showing that work practices have, in some ways, adapted to dispersion as a norm.

Limitations

Although our findings describe posting behavior in online communities, our focus here was on asynchronous, community-wide communications. It is possible that synchronous forms of communication (e.g., video/audio conferencing or IM) would show greater effects of dispersion, especially *time separation*. Additionally, in common with many studies of online communities, our data only include activity that took place within the online communities themselves, but it is likely that OEC members interact in other ways (e.g., email, phone, face-to-face). As a result, our data speaks to the habits of the OECs themselves but dispersion may affect interactions members have in other media.

Despite our finding reduced effects of dispersion on post frequency and inequality, these analyses did not capture how dispersion affects other community success metrics important to organizations, such as member satisfaction and organizational impact. Prior work suggests that membership size and number of posts do not predict member satisfaction [43]. Although we were unable to perform qualitative analyses in the current study due to the retrospective nature of this work, a mixed-methods approach similar to [43] could establish dispersion's effects on other aspects of OECs in future work.

Finally, we explored these effects in a single large organization, which limits generalizability. The high level of support for, and widespread proliferation of, OECs within this organization may not be present in other settings. That said, the communities we sampled were highly diverse and representative of communities reported elsewhere. Prior work details the goals and activities of communities in the same organization and using the same tool, finding that communities were primarily focused on common community goals like member learning, reuse of resources, and collaboration [43]. Our study results may best generalize to other organizations with communities that focus on similar goals. Although these communities were drawn from a single large organization, similar dispersion issues face teams and communities on the internet, and who work across organizations. Dispersed work is common, and increasingly, communities and teams are adopting the social media tools observed in this study.

CONCLUSION

Our results suggest that geographic dispersion is not an insurmountable barrier to participation in enterprise online communities. This sets online communities apart from small teams, for which prior research on geographic dispersion has shown mostly negative effects [13, 19, 20, 22, 24, 34, 51]. However our results support case studies that have claimed online communities can break down geographic barriers [48, 68]. Based on our results, we advocate the use of community approaches and collaborative tools to facilitate global communication in enterprise settings.

Our results show that online enterprise communities that use multiple tools to support asynchronous collaboration seem able to counteract the effects of dispersion on participation. We add to a growing body of work suggesting that dispersion can be compensated for by deploying collaborative technologies that facilitate coordination and resource sharing [15, 23, 28].

Our findings, which show few dispersion effects for participation in online enterprise communities, emphasize the importance of replication in CSCW research as technologies evolve and new populations of interest arise. Although earlier work has described large, negative effects for geographic dispersion in small teams, those effects did not materialize for a large sample of online enterprise communities that included several hundred small work teams using community tools. Work culture adaptations that more readily accommodate spatial distance and time separation, the increasing prevalence of online multiple-tool communities, and greater awareness of how to design for global work all affect the ways that dispersion manifests in online enterprise settings. Although distance assuredly still matters, we might be getting to the point of designing around it rather than succumbing to it.

REFERENCES

1. Mark S. Ackerman, Juri Dachtera, Volkmar Pipek and Volker Wulf. 2013. Sharing knowledge and expertise: The CSCW view of knowledge management. *CSCW 22*, 4-6, 531-573.
2. Nader Ale Ebrahim, Shamsuddin Ahmed and Zahari Taha. 2009. Virtual R & D teams in small and medium enterprises: A literature review. *Scientific Research and Essays 4*, 13, 1575-1590.
3. Nguyen-Duc Anh, Daniela S. Cruzes and Reidar Conradi. 2012. Dispersion, coordination and performance in global software teams: A systematic review. In *Proceedings of the International Symposium on Empirical Software Engineering and Measurement* (ESEM '12), 129-138.
4. Anupriya Ankolekar, Markus Krötzsch, Thanh Tran and Denny Vrandecic. 2007. The two cultures: Mashing up web 2.0 and the semantic web. In *Proceedings of the International Conference on World Wide Web* (WWW '07), 825-834.
5. Ofer Arazy and Oded Nov. 2010. Determinants of wikipedia quality: the roles of global and local contribution inequality. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (CSCW '10), 233-236.
6. Gerard Beenen, Kimberly Ling, Xiaoqing Wang, Klarissa Chang, Dan Frankowski, Paul Resnick and Robert E. Kraut. 2004. Using social psychology to motivate contributions to online communities. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (CSCW '04).
7. U Bojārs, John G. Breslin, Aidan Finn and Stefan Decker. 2008. Using the semantic web for linking and reusing data across web 2.0 communities. *Journal of Web Semantics 6*, 1, 21-28.

8. Roman Boutellier, Oliver Gassmann, Holger Macho and Manfred Roux. 1998. Management of dispersed product development teams: The role of information technologies. *R&D Management 28*, 1, 13-25.
9. Erin Bradner, Gloria Mark and Tammie D. Hertel. 2003. Effects of team size on participation, awareness, and technology choice in geographically distributed teams. In *Proceedings of the Hawaii International Conference on System Sciences* (HICSS '03).
10. Brian S Butler. 2001. Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Information Systems Research 12*, 4, 346-362.
11. Brian Butler, Lee Sproull, Sara Kiesler and Robert Kraut. 2007. Community effort in online groups: Who does the work and why? *Leadership at a Distance 11*, 171-194.
12. Michael A Campion, Gina J. Medsker and A. Catherine Higgs. 1993. Relations between work group characteristics and effectiveness: Implications for designing effective work groups. *Personnel Psychology 46*, 4, 823-850.
13. Paul S. Chinowsky and John E Taylor. 2011. Distance Matters: A Social Network Analysis of Geographic Dispersion in Engineering Organizations. In *Proceedings of the Engineering Project Organization Conference* (EPOC '11).
14. Susan G. Cohen and Diane E. Bailey. 1997. What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of Management 23*, 3, 239-290.
15. Gregorio Convertino, Sanjay Kairam, H. Chi, Antonietta Grasso, Peter Pirolli, Tanya Stricker and Eduardo Bascaran. 2010. Designing for learning communities in a large enterprise. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (CSCW '10).
16. Jonathon N. Cummings. 2001. Work groups and knowledge sharing in a global organization. In *Proceedings of the Academy of Management* (AOM '01), D1-D6.
17. Jonathon N. Cummings. 2011. Geography is alive and well in virtual teams. *Communications of the ACM 54*, 8, 24-26.
18. Nicole B Ellison, Charles Steinfield and Cliff Lampe. 2007. The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. *Journal of Computer‐Mediated Communication 12*, 4, 1143-1168.
19. J Alberto Espinosa and Erran Carmel. 2003. The impact of time separation on coordination in global software teams: a conceptual foundation. *Software Process: Improvement and Practice 8*, 4, 249-266.
20. J Alberto Espinosa and Cynthia Pickering. 2006. The effect of time separation on coordination processes and outcomes: A case study. In *Proceedings of the Hawaii International Conference on System Sciences* (HICSS '06).
21. J. Alberto Espinosa, Sandra A. Slaughter, Robert E. Kraut and James D. Herbsleb. 2007. Familiarity, complexity, and team performance in geographically distributed software development. *Organization Science 18*, 4, 613-630.
22. Cristina B. Gibson and Jennifer L. Gibbs. 2006. Unpacking the concept of virtuality: The effects of geographic dispersion, electronic dependence, dynamic structure, and national diversity on team innovation. *Administrative Science Quarterly 51*, 3, 451-495.
23. Zixiu Guo, John D'ambra, Tim Turner and Huiying Zhang. 2009. Improving the effectiveness of virtual teams: A comparison of video-conferencing and face-to-face communication in China. *IEEE Transactions on Professional Communication 52*, 1, 1-16.
24. Sean M. Handley and W.C. Benton Jr. 2013. The influence of task-and location-specific complexity on the control and coordination costs in global outsourcing relationships. *Journal of Operations Management 31*, 3, 109-128.
25. James D. Herbsleb, Audris Mockus, Thomas A. Finholt and Rebecca E Grinter. 2000. Distance, dependencies, and delay in a global collaboration. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (CSCW '00), 319-328.
26. Pamela J. Hinds and Mark Mortensen. 2005. Understanding conflict in geographically distributed teams: The moderating effects of shared identity, shared context, and spontaneous communication. *Organization Science 16*, 3, 290-307.
27. Mohammad Delwar Hossain and Aaron S Veenstra. 2013. Online maintenance of life domains: Uses of social network sites during graduate education among the US and international students. *Computers in Human Behavior 29*, 6, 2697-2702.
28. Bo Hu and Glenn Svensson. 2010. A case study of linked enterprise data. in *The Semantic Web–ISWC 2010*, Springer, 129-144.
29. David A Huffaker. 2011. The impact of group attributes on communication activity and shared language in online communities. *First Monday 16*, 4.
30. Alicia Iriberri and Gondy Leroy. 2009. A life-cycle perspective on online community success. *ACM Computing Surveys (CSUR) 41*, 2, 11.
31. James K Jackson. 2012. *Outsourcing and insourcing jobs in the US economy: Evidence based on foreign investment data*, Washington, DC: Congressional Research Service.
32. Sara Kiesler and Jonathon N. Cummings. 2002. What do we know about proximity and distance in work groups? A legacy of research. *Distributed Work 1*, 57-80.
33. Aniket Kittur and Robert E. Kraut. 2010. Beyond Wikipedia: Coordination and conflict in online

production groups. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (CSCW ’10), 215.
34. Benjamin Koehne, Patrick C. Shih and Judith S. Olson. 2012. Remote and alone: Coping with being the remote member on the team. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (CSCW ’12), 1257-1266.
35. Robert E. Kraut, Susan R, Fussell, Susan E. Brennan and Jane Siegel. 2002. Understanding effects of proximity on collaboration: Implications for technologies to support remote collaborative work. *Distributed Work*, 137-162.
36. Robert Kraut, Carmen Egido and Jolene Galegher. 1988. Patterns of contact and communication in scientific research collaboration. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (CSCW ’88), 1-12.
37. Robert E. Kraut and Paul Resnick. 2012. *Building Successful Online Communities*. The MIT Press.
38. Robert Kraut, Susan Fussell, F. Lerch and Alberto Espinosa. 2005. Coordination in teams: Evidence from a simulated management game. *Human-Computer Interaction Institute*.
39. Juhi Kulshrestha, Farshad Kooti, Ashkan Nikravesh and P. Krishna Gummadi. 2012. Geographic Dissection of the Twitter Network. In *Proceedings of the International AAAI Conference on Web and Social Media* (ICWSM ’12).
40. Dora C. Lau and J. Keith Murnighan. 1998. Demographic diversity and faultlines: The compositional dynamics of organizational groups. *Academy of Management Review 23*, 2, 325-340.
41. Dora C. Lau and J. Keith Murnighan. 2005. Interactions within groups and subgroups: The effects of demographic faultlines. *Academy of Management Journal 48*, 4, 645-659.
42. Kalev Leetaru, Shaowen Wang, Guofeng Cao, Anand Padmanabhan and Eric Shook. 2013. Mapping the global Twitter heartbeat: The geography of Twitter. *First Monday 18*, 5.
43. Tara Matthews, Jilin Chen, Steve Whittaker, Aditya Pal, Haiyi Zhu, Hernan Badenes and Barton Smith. 2014. Goals and perceived success of online enterprise communities: what is important to leaders & members? In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI ’14), 291-300.
44. Tara Matthews, Steve Whittaker, Hernan Badenes and Barton A. Smith. 2014. Beyond End User Content to Collaborative Knowledge Mapping: Interrelations Among Community Social Tools. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (CSCW ’14), 900-910.
45. Tara Matthews, Steve Whittaker, Hernan Badenes, Barton A. Smith, Michael Muller, Kate Ehrlich, Michelle X. Zhou and Tessa Lau. 2013. Community insights: Helping community leaders enhance the value of enterprise online communities. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI ’13), 513-522.
46. Tara Matthews, Steve Whittaker, Thomas P. Moran, Sandra Y. Helsley and Tejinder K. Judge. 2012. Productive interrelationships between collaborative groups ease the challenges of dynamic and multi-teaming. *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (CSCW ’12).
47. Edward F. Mcdonough, Kenneth B. Kahnb and Gloria Barczaka. 2001. An investigation of the use of global, virtual, and colocated new product development teams. *Journal of Product Innovation Management 18*, 2, 110-120.
48. David R. Millen, Michael A. Fontaine and Michael J. Muller. 2002. Understanding the benefit and costs of communities of practice. *Communications of the ACM 45*, 4, 69-73.
49. Michael Muller, Kate Ehrlich, Tara Matthews, Adam Perer, Inbal Ronen and Ido Guy. 2012. Diversity among enterprise online communities: Collaborating, teaming, and innovating through social media. In *HFCS*, 2815-2824.
50. Blair Nonnecke and Jenny Preece. 2000. Lurker demographics: Counting the silent. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI ’00), 73-80.
51. Michael Boyer O'Leary and Mark Mortensen. 2010. Go (con) figure: Subgroups, imbalance, and isolates in geographically dispersed teams. *Organization Science 21*, 1, 115-131.
52. Michael B O’Leary and Jonathon N. Cummings. 2007. The spatial, temporal, and configurational characteristics of geographic dispersion in teams. *MIS Quarterly 31*, 3, 433-452.
53. Gary M. Olson and Judith S. Olson. 2000. Distance matters. *Human-Computer Interaction 15*, 2, 139-178.
54. Gary Olson and Judith Olson. 2003. Mitigating the effects of distance on collaborative intellectual work. *Economics of Innovation and New Technology 12*, 1, 27-42.
55. Felipe Ortega, Jesus M. Gonzalez-Barahona and Gregorio Robles. 2008. On the inequality of contributions to Wikipedia. In *Proceedings of the Hawaii International Conference on System Sciences* (HICSS ’08), 304-304.
56. Namsu Park, Kerk F. Kee and Sebastián Valenzuela. 2009. Being immersed in social networking environment: Facebook groups, uses and gratifications, and social outcomes. *CyberPsychology & Behavior 12*, 6, 729-733.
57. Alexandre Passant, Philippe Laublet, John G. Breslin and Stefan Decker. 2010. Enhancing enterprise 2.0 ecosystems using semantic web and linked data technologies: The SemSLATES approach. in *Linking Enterprise Data*, Springer, 79-102.

58. Jeffrey T. Polzer, C. Brad Crisp, Sirkka L. Jarvenpaa and Jerry W. Kim. 2006. Extending the faultline model to geographically dispersed teams: How colocated subgroups can impair group functioning. *Academy of Management Journal 49*, 4, 679-692.
59. Malin Pongolini, Johan Lundin and Lars Svensson. 2011. Global online meetings in virtual teams: from media choice to interaction negotiation. In *Proceedings of the International Conference on Communities and Technologies* (C&T '11), 108-117.
60. Daphne R. Raban, Mihai Moldovan and Quentin Jones. 2010. An empirical study of critical mass and online community survival. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (CSCW '10), 71.
61. Narayan Ramasubbu, Marcelo Cataldo, Rajesh Krishna Balan and James D. Herbsleb. 2011. Configuring global software teams: a multi-company analysis of project productivity, quality, and profits. In *Proceedings of the ACM/IEEE International Conference on Software Engineering* (ICSE '11), 261-270.
62. Matthew Rowe, Miriam Fernandez, Harith Alani, Inbal Ronen, Conor Hayes and Marcel Karnstedt. 2012. Behaviour analysis across different types of enterprise online communities. In *Proceedings of the ACM Web Science Conference* (WebSci '12) 255-264.
63. Linda Schweitzer and Linda Duxbury. 2010. Conceptualizing and measuring the virtuality of teams. *Information Systems Journal 20*, 3, 267-295.
64. François Paul Servant. 2008. Linking Enterprise Data. In *Proceedings of the Linked Data on the Web Workshop* (LDOW '08).
65. James P. Stevens. 1984. Outliers and influential data points in regression analysis.*Psychological Bulletin*, *95*, 2, 334-344.
66. John C. Tang, Chen Zhao, Xiang Cao and Kori Inkpen. 2011. Your time zone or mine?: A study of globally time zone-shifted collaboration. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (CSCW '11), 235-244.
67. James J Treinen and Susan L Miller-Frost. 2006. Following the sun: Case studies in global software development. *IBM Systems Journal 45*, 4, 773-783.
68. Etienne Wenger, Richard Arnold Macdermott and William M Snyder. 2002. *Cultivating Communities of Practice: A Guide to Managing Knowledge*. Harvard Business Press.
69. Steve Whittaker. 1996. Talking to strangers: An evaluation of the factors affecting electronic collaboration. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (CSCW '96), 409-418.
70. Steve Whittaker, Loren Terveen, Will Hill and Lynn Cherny. 1998. The dynamics of mass interaction. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (CSCW '98), 257-264.