iced19

COORDINATION IN DESIGN TEAMS AS A LENS TO IDENTIFYING TEAM ROLES

Collopy, Arianne X.; Li, Chengxi; Liu, Tianyi; Adar, Eytan; Papalambros, Panos Y.

University of Michigan

ABSTRACT

Coordination in system design requires an interplay between different roles. In this work, we identify five design team roles that pertain to the partitioning and coordination of distributed design team tasks. The proposed characterization is based on self-reported responsibilities and communication behaviors from 109 student designers in 22 teams at the conclusion of a semester-long design project. The self-reports capture both how team members viewed their own work as well as communication patterns between team members. We leverage two representations of this data. Through text analysis, we identify keywords describing team member roles and responsibilities. Social network analysis can further distinguish roles based on team communication behaviors. Cluster analysis on both types of data identifies groups of individuals with similar characteristics. The resulting five clusters capture common roles in system design teams that simultaneously capture the diverse responsibilities and communication patterns.

Keywords: Coordination, Communication, Teamwork, Human behaviour in design, Large-scale engineering systems

Contact:
Collopy, Arianne X.
University of Michigan
Integrative Systems + Design
United States of America
acollopy@umich.edu

1   INTRODUCTION

Decomposition-based design of engineered systems is achieved by partitioning into subsystems, assignment to individual engineers, and coordination. The goal of coordination is overall system compatibility when all subsystems are integrated. System design optimization algorithms (also referred to as multidzisciplinary design optimization) are well-established for performing portions of this process. Algorithms for decomposition and coordination of system design problems are given in Papalambros and Wilde (2017) and Martins and Lambe (2013). However, in actual system design, individuals are unlikely to coordinate by following algorithmic procedures. In a design organization, distributed tasks must be first partitioned so that they can be worked in parallel, and then coordinated so that the results can be joined together to effect the overall project goal. In this organizational context, coordination is primarily a communicative process focused on information sharing among parallel tasks. Critically, the roles of individual participants influence this coordination process.

Numerous sets of team roles have been identified in the literature, focused on team member behaviors (Aritzeta *et al.*, 2007), personality (Stewart *et al.*, 2005), and cognitive styles (Kress and Schar, 2012). Roles pertaining to coordination between groups have been identified (Sonnenwald, 1996; Sheard, 1996), but do not necessarily translate to coordination activity within groups or teams.

To better understand these roles, we analyze coordination in the controlled environment of student design projects (159 students in 32 teams). Our goal is to identify how team members coordinate their design work in a design task requiring partitioning. We use self-report surveys and communication as a proxy for coordination. Using network representations of the survey data, we identify structures and patterns in team communication. By looking at the correlation between network structures and self-reported team roles, responsibilities, and teamwork, we are able to identify what we term *coordination roles* comprising responsibilities and communication behaviors found in these design teams.

2   STUDY DESIGN: DATA COLLECTION AND ANALYSIS APPROACH

2.1 Data collection approach and context

Study participants were second year undergraduates at the University of Michigan enrolled in the course *Design and Manufacturing I* in Winter 2017. As part of the course, students work in teams of five to design and build robotic machine players (RMPs). In each class section, four teams design and build their own RMPs, which collaboratively compete as a squad. The squad's RMPs are remotely piloted by students with a goal of scoring the most points. An overhead view of an in-progress game is shown in Figure 1.
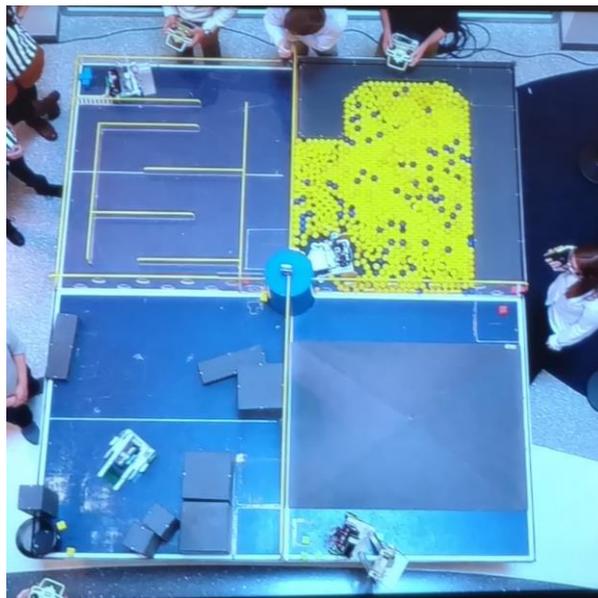

Figure 1. Overhead photo of game showing four-zone board, each with its own remotely-piloted RMP. Each zone requires navigation of different obstacles: clockwise from top left obstacles are a maze, ball pit, pyramidal ramp, and heavy blocks. (photo taken April 2017)

RMPs are built from a standard kit though additional parts can be added or substituted. Typical features of an RMP include chassis, wheels, drivetrain, and a shovel or scoop to collect small cubes and move them to scoring zones. This course is an ideal situation to study coordination. Each team must coordinate its actions to design and manufacture an RMP and simultaneously coordinate with the other teams in their squad to implement a winning strategy. The course and typical project are described in more detail in Papalambros (2015).

We administered paper-form self-report surveys at the conclusion of the semester. The survey asked students to: reflect on the semester's project work; report their self-determined roles; identify tasks they were responsible for; describe the roles and responsibilities of their teammates; and recount the frequency and usefulness of task-focused communications. Existing survey instruments designed to conduct social network analysis within organizations (Cross *et al.*, 2002; Krebs, 2000) were adapted to create the survey instrument. The survey was voluntary but we received at least partial responses for 140 students in the course (88% of 159). In the study reported in this paper, our goal is to *characterize* the different approaches taken to coordinate distributed design work within teams, rather than to *evaluate* those approaches. Accordingly, we did not correlate responses to grades or game results.

2.2 Analysis approach

The survey asked each respondent to share both their own, and their teammates', main roles and responsibilities. These questions were open-response to avoid limiting the scope of described responsibilities offered by the students. The survey asked two sets of questions about inter-team communication, both focused on frequency and usefulness of communication with other team members. The questions differed in their focus on communication content. *Self* Questions and *Peer* Questions asked about communication pertaining to the respondent's own work and to their peers' work, respectively. Respondents were asked to indicate frequency of communication on a five-point scale ("4+ days per week", "2-3 times per week", "< 2 times per week", "Monthly", and "Never"), and usefulness of communication on a four-point scale ("N/A", "Not Needed", "Helpful", and "Essential"). Because each team member was asked for this information, each pairwise interaction should be described by two responses. In the case where one team member did not respond, there were still responses from their peers to at least partially represent their participation on the team. Responses were collected without names using alphanumeric labels for team members within each squad, then aggregated across all squads by assigning a single unique number to each respondent.

We processed the survey data using two different analyses, each applied to different data subsets. The first used text analysis (described in Section 2.2.1) to identify common keywords representative of the roles and responsibilities respondents reported within their team. This analysis allowed us to use participant language and terminology to describe types of tasks each undertook. We included all individuals that gave responses to questions about the role and/or responsibilities of themselves or their peers. This gave us data from 140 individuals across all 32 teams (with 19 non-respondents). However, because surveys also reflected peer roles and communications, we could infer details for 154 individuals. A second analysis focused on the communication network within the teams (detailed in Section 2.2.2). We modelled each team member as a node with an edge connecting two nodes (i.e., individuals) who reported communication in the survey. We excluded individuals with no responses (n = 19) or partial responses for the communication questions (n = 11). Entire teams were thus excluded if there were missing responses from two or more members out of five (n = 10 teams). The resultant networks covered 109 individuals from 22 teams (21 teams of five, and one team of four).

2.2.1 Keyword identification and analysis method

Keywords were identified from responses using the Term Frequency and Inverse Document Frequency (TF-IDF) weights (Salton, 1986). TF-IDF measures relevance and importance of words or terms found in a specific document. In our case, each student's response to a question was treated as a document. We used the TfidfVectorizer module (Scikit-Learn, 2018) implemented in the open source scikit-learn package to calculate TF-IDF scores. The term frequency, *tf*, for a word or phrase, *w*, is the raw count of that term in a given survey response, *r*. The inverse document frequency, *idf*, is calculated as a function of the total response count (*N*) and those responses *r* that contain a word *w*. Specifically:

$$idf(w) = \log\left(\frac{1+N}{1+\sum_r w}\right) + 1 \tag{1}$$

TF favors frequent words within each document, while the IDF favors rare words across all documents. The final TF-IDF score is a product of these frequencies, in turn favoring words that have relatively high scores for both frequencies. This combination penalizes terms frequent within responses as well as between responses (e.g., 'a', 'the', 'we') as they may have high TF scores but should have low IDF scores. The TF-IDF score ranking identifies common jargon in our survey responses as well as specific terms that may only be used within one or two teams. We first calculated the TF-IDF scores of all words in the raw survey responses to identify and rank top words used in descriptions of team roles, responsibilities, and project tasks. We then reviewed the data iteratively to identify overlapping keywords such as 'manufacturing', a verb, and 'manufacturing plan', a noun. We also identified dissimilar words that can be interchanged with little loss of meaning such as 'initial design' and 'preliminary concept' or 'teammate' and 'team member'; these were combined under a single label for consistency. We manually consolidated terms into distinct keywords based on these rules, in effect transforming the open response data into closed form representations of respondent language. Table 1 gives examples of the final keyword set, categorized into six dimensions that reflect approaches to problem partitioning (aspect and object), approaches to project organization and coordination (team roles, management tasks, and reflection on how work was coordinated, e.g., split or shared), and deliverables or documentation (documents).

Table 1. Categories of keywords identified from survey responses, with examples

Aspect	Object	Documents	Team Role	Management	Coordination
CAD, design, manufacturing, assembly, brainstorm, analysis	Part, lift, system, scoop, chassis, drivetrain	Report, manufacturing plan, sketch, assignment	Leader, member, general engineer, specialist	Manage, organize, make sure, on track	Split work equally, everybody did a bit of everything, work together, help

The new set of transformed data used 471 distinct words across all responses. We recalculated TF-IDF scores using the transformed data. For this calculation we used as a single response unit all question responses about an individual, whether from that individual or their peers. The result was a vector of TF-IDF scores of length 471 for each individual, with a value of 0 indicating the term was unused in describing that individual's role. Each vector was then normalized by its Euclidean norm. We further removed stop words such as prepositions, conjunctions, and indefinite pronouns, leaving nouns, verbs, and some adjectives. We analyzed the top 50 remaining words (ranked by the average TF-IDF scores across all individuals). The result was a vector of length 50 with weighted values representing the dominance of a word in the responses about an individual's work. Individuals with the highest normalized TF-IDF scores for a word were those whose teammates described their role or responsibilities the same way, and for whom few other words were used to describe their role and responsibility.

2.2.2 Network representations and measures

To model the responses to communication questions, we represented the data as networks. The survey data was on a scale which we transformed to a binary edge indicator using a fixed threshold. A given threshold—or combination of frequency and usefulness levels—means we included data that met any combination of levels defined in the threshold. We compared several potential thresholds and ultimately selected a threshold of 4+ days per week under frequency, and 'helpful' or 'essential' under usefulness. This threshold yielded a dataset that includes 221 of 426 reported edges in the Self networks, and 190 of 430 reported edges in the Peer networks. This threshold was chosen over others as it presents connected team structures while preserving distinctive features. A lower threshold, including less frequent communication, forms team networks that are nearly all identical. A higher threshold, including only communication reported as essential, did not give connected teams in most cases. We also chose to separate the more objective frequency levels as opposed to the more subjective usefulness levels, including both 'essential' and 'helpful' in our chosen threshold. Under this threshold, only survey responses from respondent A about communication with peer B which follow the pattern below are counted as an edge:

- Self: A reports talking to B about A's work 4+ days per week and it is essential or helpful to A.
- Peer: A reports talking to B about B's work 4+ days per week and it is essential or helpful to A.

The Self and Peer network data cannot be directly superimposed as the edges in each network refer to communication about different things. To avoid oversimplification of our data, we created five networks, each telling us something different about the data. The *Self* and *Peer* networks were created directly from the survey responses to each communication question, the former about communication with peers about the respondent's own work and the latter about communication with peers about the peer's work. We then combined the results of these two questions in three different ways. The first is an *Initiator Network*, where there is an edge from A to B if respondent A reports talking to peer B about both A's own work as well as B's work. The second is a *Purpose-directed Network*, where there is an edge from A to B if either respondent A or B reports talking to the other about B's work. Finally, in an *Agreement Network*, there is an edge from A to B if both A and B respond that they talked to the other about B's work; in other words, the responses from both members agree. In each network, an edge is defined by the direction and content of communication, given that the edge reported meets the selected threshold of frequency and usefulness. The network definitions are summarized in Table 2.

Table 2. Description of five network representations of thresholded survey data

Network	Edge from A→B in this network if:
Self	A reports talking to B about A's work
Peer	A reports talking to B about B's work
Initiator	A reports talking to B about A's work AND A reports talking to B about B's work
Purpose-directed	A reports talking to B about B's work OR B reports talking to A about B's work
Agreement	A reports talking to B about B's work AND B reports talking to A about B's work

In the Self, Peer, and Initiator Networks, a node's out-edges represent their communication with someone else. In contrast, in the Purpose-directed and Agreement Networks, a node's out-edges can be interpreted as that person having information about what their peers at the other end of those edges are working on – the direction of the edge indicating the topic of conversation. We characterize the data by calculating several measures from each network. These measures look at common features of social networks: degree, centrality, and clustering coefficient (Newman, 2018). We use node-level measures rather than network-level measures as the current analysis is focused on individuals rather than the team as a whole.

Degree is a count of the number of edges inbound to or outbound from a node. *Out-degree* is the number of people a respondent is communicating with or whose work they are receiving information about. *In-degree* in the number of people a respondent is on the receiving end of communication from. In the Initiator Network, out-degree carries a special meaning we call *initiator-ness*: the behavior of communicating with others about both one's own work and their peers. Those with maximum initiator out-degree we call *top initiators*. We find that most teams have one top initiator and rarely have more than two. Overlaying the edges in the Agreement Network with the Purpose-directed Network, we can also calculate what percentage of directed edges are reported by both respondents.

Centrality measures of nodes in a network quantify the importance of a node. *Betweenness centrality* is a measure of how many shortest paths between pairs of nodes a certain node lies on. *Path length* is the number of edges in the network traversed to get from one node to another. In a directed network, the paths are also directed. In networks where edges indicate information flow, a person with high betweenness centrality indicates they are receiving a lot of information and are also serving as a bridge or source from which others might be able to access that information. *Closeness centrality* measures how closely connected a node is to all others in the network. *Harmonic closeness centrality* is the ratio of the number of peer nodes in a network to the minimum path length required to reach all nodes. The highest value of harmonic closeness centrality is achieved for a person who reported communicated with all other members of their team.

Finally, the *clustering coefficient* measures the tendency of nodes to be closely linked, or clustered (Watts and Strogatz, 1998). Here we used the local clustering coefficient measure defined for the ego-network of a node: a person and their immediate connections. The local clustering coefficient is the ratio of links that exist within a node's ego-network to the number of links that could exist. In small teams, the local clustering coefficient is strongly correlated to overall team density.

The node-level measures we calculated from network representations of our data are summarized in Table 3. We used these measures as inputs for clustering analysis to identify groups of similar nodes

based on the structural features of their team networks. Some detail on how we conducted our clustering analysis is provided in Section 2.2.3 and the results of the analysis are in Section 3.

Table 3. Measures calculated for each node from network representations of survey data

Network	Measures calculated
Self	In-degree, out-degree, betweenness centrality, harmonic closeness centrality
Peer	In-degree, out-degree, betweenness centrality, harmonic closeness centrality
Initiator	Out-degree (initiator-ness), out-degree = max possible? (binary)
Purpose-directed	In-degree, out-degree, total degree, number agreement edges out of total degree, betweenness centrality, harmonic closeness centrality, local clustering coefficient

2.2.3 Clustering analysis

The text and network-based analyses result in a numeric vector that represents every individual in the analysis. We concatenate the weighted TF-IDF scores for each of the top-50 keywords and the 17 network measures calculated for each individual. This results in a vector of 67 numbers representing task responsibilities and communication behaviors for each of 109 individuals in 22 teams as reported by themselves and their peers. We use these vectors as inputs to clustering analysis.

Since this clustering is exploratory, we selected an unsupervised clustering algorithm. We used hierarchical clustering based on the Euclidean distance calculated between each pair of vectors. In hierarchical clustering, each individual starts out in her own cluster. Then clusters are joined one at a time to their most similar neighbor based on the contents of their vector representation. Once each cluster has more than one individual, new calculations are measured relative to the cluster centroid. This proceeds until all individuals are joined into one final cluster. Recording the within-cluster similarity at each step allows the creation of a plot of similarity versus number of clusters. A sharp drop in within-cluster similarity as the number of clusters decreases suggests a good stopping point for clustering; continuing to join dissimilar clusters yields less interpretable and less meaningful results (Thorndike, 1953). We used this "knee" or "elbow" found from hierarchical clustering as an input for k-means clustering. In k-means clustering, clusters are calculated the same way as in hierarchical clustering except individuals can be moved from one cluster to another as clustering proceeds until each individual is closer to its cluster centroid than any others. The results were obtained from clustering with k=5 clusters. We describe the network measures and keyword statistics for each cluster in the following section.

3 CLUSTERING RESULTS AND INTERPRETATION

Clustering individuals based on combining network measures and keywords appears to give a balance between clusterings obtained individually. We present here only the results of this composite clustering based on the highly representative keywords and typical values of network measures. Section 4 discusses further the composite cluster characteristics.

3.1 Clusters as described by keywords

The top 20 keywords in each cluster are given in Figure 2. Several keywords are highly ranked in all clusters: CAD, manufacturing, design, machining, manufacturing plans, team leader, and shared responsibilities. This finding highlights core aspects of coursework. Project work involves designing parts in CAD, putting together documentation for those parts, and manufacturing or machining them. It therefore makes sense that these terms would be widely used. Team leaders also were not performing just dedicated roles but were also engaged in many other tasks. It is then also reasonable that those serving as team lead may be grouped in various clusters with others serving different primary functions on the team. Finally, shared work is common, likely due to the inclusion in analysis of the responses to a survey question specifically asking respondents to indicate which of a peer's roles overlap with their own. In most cases, the students reported working together or having shared responsibilities. Common to four of the five clusters is the keyword "split the work". Combined with "shared responsibilities", this may suggest a variety of working patterns within each cluster. We briefly discuss each cluster in turn.

- *Leaders*: Six of the seven individuals in this group had responsibilities described as a combination of being team lead, managing work, organizing tasks or schedules, or leading subsystem design work. While only three members of this group were described using the term "team lead", those

three were given this title by their teammates often enough to give an average TF-IDF score in the group of 0.16.

- *Supporters*: This group is small (3 members). While in other clusters it is more common for individuals to refer to themselves as a "team member", in this group it is more common for individuals to refer to themselves as "general engineers". "Engineer" receives an average TF-IDF score of 0.21 in this group. They emphasized working together, evenly splitting work, and sharing responsibilities. Also emphasized are initial design and brainstorming, typically collaborative tasks.

- *Makers*: Almost everyone in this group was described as working on manufacturing, design, CAD, and generic parts. Specific subsystems were separately labelled in the keyword analysis. Chassis is emphasized for six members of this group, enough to bring it into the top 20 keywords in this group with an average TF-IDF score of 0.06.

- *Documenters*: As with other groups, making parts is emphasized. We call them documenters, however, due to the frequent mention that they were working on documentation such as manufacturing plans, engineering drawings, and writing reports. These terms apply to 19 members (70 percent) of this cluster.

- *Generalists*: This cluster is defined by the diversity of keywords in the top 20 of this group. Overall the keywords are similar to the group of Makers. Some individuals are also team leaders and organizers, others are working on, and several specifically mentioned working on various supports to help integrate subsystems together. 18 members (55 percent) of this group were perceived as working on "everything" by themselves or their peers.

Cluster (Group size)	Keywords
Leaders (n = 7)	**Team leader**, **scoop**, **mill**, **manage**, **CAD**, **video**, **analysis**, **machining**, **assembly**, **shared responsibilities**, **organize**, **help**, **design**, **manufacturing plans**, **everything**, **build**, **report**, **manufacturing**, **team**, **lead**
Supporters (n = 3)	**Engineer**, **manufacturing**, **work together**, **part**, **analysis**, **design**, **manufacturing plans**, **split the work**, **drivetrain**, **CAD**, **initial design**, **team leader**, **engineering drawings**, **work**, **assembly**, **complete**, **writing**, **assignments**, **shared responsibilities**, **machining**
Makers (n = 39)	**Manufacturing**, **team member**, **design**, **CAD**, **part**, **assembly**, **lathe**, **team leader**, **manufacturing plans**, **create**, **everything**, **machining**, **shared responsibilities**, **split the work**, **chassis**, **help**, **work**, **no specific role**, **mill**, **everybody did everything**
Documenters (n = 27)	**CAD**, **manufacturing**, **design**, **shared responsibilities**, **part**, **mill**, **team leader**, **engineering drawings**, **lathe**, **manufacturing plans**, **engineer**, **report**, **help**, **work**, **writing**, **work together**, **machining**, **team**, **create**, **split the work**
Generalists (n = 33)	**CAD**, **manufacturing**, **team member**, **design**, **team leader**, **everything**, **machining**, **part**, **assembly**, **manufacturing plans**, **engineering drawings**, **help**, **mill**, **create**, **lathe**, **supports**, **shared responsibilities**, **split the work**, **organize**, **no specific role**

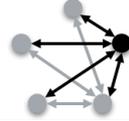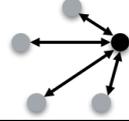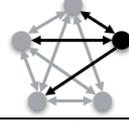0.25   0.20   0.15   TF-IDF score   0.10   0.05   0.00

Figure 2. Top 20 keywords in each cluster in order based on decreasing average TF-IDF score within cluster. The color of each word indicates the average TF-IDF score.

3.2 Clusters as described by network measures

We identify typical network measures within each cluster. Select measures calculated from within each cluster along with a representative network diagram are shown in Table 5. Although these clusters are the same as discussed in the previous section, we describe them here in terms of their local network characteristics. We discuss the labels in combination in Section 4.

There are 39 *top initiators* out of 43 individuals in the groups *Connectors*, *Hubs*, and *Low-density Hubs* due to their maximum out-degree in the Initiator network. The remaining four individuals in this group have high Initiator out-degree. The three clusters are divided by their purpose-directed betweenness centrality values, reflecting the number of other communication edges in the team network. The remaining two groups of Communicators and Receivers are not initiators, typically having zero Initiator out-degree. A brief summary of each cluster is below.

Table 5. *Individual cluster descriptions with selected network measures and example team network from purpose-directed network with representative node highlighted*

Cluster label (group size)	Select network measures: mean (standard deviation)							Example network, purpose directed network
	Self out-degree	Self in-degree	Peer out-degree	Peer in-degree	Initiator out-degree	Purpose directed total degree	Purpose directed betweenness centrality	
Connectors (n = 33)	3.8 (0.6)	2.3 (0.8)	3.9 (0.2)	1.9 (0.7)	3.8 (0.6)	7.8 (0.6)	0.9 (0.8)	
Hubs (n = 7)	3.9 (0.4)	1.0 (0.6)	3.9 (0.4)	0.6 (0.5)	3.9 (0.4)	7.7 (0.8)	3.7 (1.1)	
Low-density Hubs (n = 3)	4 (0)	0.7 (0.6)	4 (0)	0 (0)	4 (0)	8 (0)	10 (1.7)	
Communicators (n = 39)	1.5 (1.8)	2.5 (0.9)	0.5 (1.2)	2.4 (0.8)	0.1 (0.3)	5.9 (1.0)	0.0 (0.2)	
Receivers (n = 27)	0 (0)	1.4 (0.9)	0 (0)	1.1 (0.8)	0 (0)	3.2 (0.8)	0 (0)	
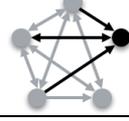
- *Connectors*: These individuals are top initiators in medium to high density teams as indicated by their low purpose-directed betweenness centrality values of 0-3. They tend to be one of multiple connectors on their team.
- *Hubs*: They are top initiators but in contrast to the Connectors, they have low in-degree in the Self and Peer networks. This means that their teammates are communicating with them less often, making Hubs central within their team.
- *Low-density Hubs*: This group is particularly identified by high purpose-directed betweenness centrality values, ranging from 9-12. They are top initiators in the sparsest of the 22 teams.
- *Communicators*: They are particularly defined by their total degree in purpose-directed network. The average degree of 5.9 indicates that they are well-connected to most of their teammates. Their Self and Peer network in-degree measures are similar to Connectors and their out-degree in these networks is similar to Receivers. This combination suggests these individuals are situationally engaged in discussion with their teammates about either their own work or their peers' work, and rarely both.
- *Receivers*: They are always on the receiving end of communication from their peers, with zero out-degree and typically non-zero in-degree in both Self and Peer networks. Several in this group were completely independent of the rest of their team, meaning no communication with their peers that met our frequency and usefulness threshold for analysis, resulting in network measures of zero.

4 DISCUSSION: COORDINATION ROLES

Our analysis identified five clusters of individuals, each representing a different combination of project tasks and communication behaviors. Thus far we have considered these roles on their own. Since coordination is a system-level outcome, it is important to also examine how these roles are distributed within teams. Connectors-Generalists tend to be on teams with Communicators-Makers: 59% of the teams in this analysis are comprised of only these two roles. The other 41% of teams have at least one Receiver-Documenter. These individuals did not report frequent and useful communication with their peers, and are typically in teams with a Hub-Leader or Hub-Supporter. The tasks these Hub roles engaged in are different,

but they serve a similar communicative role of keeping the team connected. Their presence alongside Receiver-Documenter roles (and never alongside Communicators-Makers) suggests that team roles are in part a reflection of the team composition. This is consistent with observations that individuals take on different roles as the situation demands (Senior, 1997, Stewart, 2005, Adams *et al.*, 2009).

The dichotomy between teams that include Communicators-Makers and Connectors-Generalists and those that are some combination of all roles except Connectors-Generalists may also be an indication of how work was partitioned among the team members. The central Hub role, either Supporter or Leader, is in teams with Communicators-Makers and/or Documenters-Receivers. This combination suggests highly partitioned work, where the Communicators-Makers and Documenters-Receivers all work on their own parts. In a hierarchical partitioning strategy, subsystems are divided and their coordination is left to a central overseer, in this case, a Hub role. This strategy is consistent with the observed low communication reported from the remaining team members. In contrast, the teams with Communicators-Makers and Connectors-Generalists tended to form highly connected team networks with frequent communication among all members. This strategy is consistent with non-hierarchical partitioning, where there is no central coordinator and all parties manage their own interactions.

We also note that while we have a group we call 'Leaders', all clusters included a few team leaders. According to Ehrlich and Cataldo (2014), teams with technical leaders that have high centrality in their team and high communication relative to others in the team are more likely to deliver a higher quality product. Again, we did not evaluate the performance of the teams' output, but the Hubs-Leaders grouping fits the profile of high-centrality leaders. Shared leadership is also common in student project teams (Novoselich and Knight, 2018), which may look like a relative lack of centrality for those doing leadership tasks (Ehrlich and Cataldo, 2014). Team leaders not in Hub clusters may be an example of instances where leadership tasks were shared by multiple team members. In addition, different leadership styles have different associated communication styles (de Vries *et al.*, 2010). Therefore, these teams may exhibit a combination of shared leadership and varied leadership styles.

5   CONCLUSIONS AND FUTURE WORK

This study is an early step in understanding how coordination is practiced in actual system design. We used a combination of network analysis and text analysis to identify five coordination roles adopted in distributed project work. This methodology allowed identification of cluster descriptions and roles that are tailored to the design task. We see some indication that teams seemed to adopt hierarchical or non-hierarchical partitioning strategies based on the combination of roles within each team. Future work could look into additional factors impacting successful teamwork such as those identified by Maier *et al.* (2008) and Crabtree *et al.* (1997) to more fully characterize coordination roles.

In addition, we acknowledge several limitations that should be addressed in future iterations of this study. The first is that novice designers are not necessarily the best proxy for understanding of organizational processes. Novice and expert designers approach problem-solving differently (Cross, 2004; Smith and Leong, 1998). Novice designers also tend to focus on subsystem level optimization rather than system optimization (Austin-Breneman *et al.*, 2012). This suggests professionals may have different coordination approaches than the student teams recruited for this study, indicating validation from industry practice is warranted.

Furthermore, self-report surveys suffer from recency bias and reporting bias. Deploying our survey at the end of the design project means end-of-term activities are likely overrepresented in reported project responsibilities. Questions about communication frequency and usefulness may be inaccurate, and tasks reported for each team member may be incomplete. In subsequent studies, these biases in survey responses could be mitigated by additional data sources, such as more frequent surveys, review of documentation, or communication observed via project management software.

REFERENCES

Adams, R., Mann, L., Jordan, S. and Daly, S. (2009), "Exploring the Boundaries: Language, Roles, and Structures in Cross-Disciplinary Design Teams", In McDonnell, J. and Lloyd, P. (eds.), *About: Designing - Analysing Design Meetings*, CRC Press, London, 339–358

Aritzeta, A., Swailes, S. and Senior, B. (2007), "Belbin's Team Role Model: Development, Validity and Applications for Team Building", *Journal of Management Studies*, Vol. 44 No. 1, pp. 96–118. http://doi.org/10.1111/j.1467-6486.2007.00666.x

Austin-Breneman, J., Honda, T. and Yang, M.C. (2012), "A Study of Student Design Team Behaviors in Complex System Design", *Journal of Mechanical Design*, Vol. 134 No. 12, p. 124504. http://doi.org/10.1115/1.4007840

Crabtree, R.A., Fox, M.S. and Baid, N.K. (1997), "Case studies of coordination activities and problems in collaborative design", *Research in Engineering Design*, Vol. 9 No. 2, pp. 70–84. https://doi.org/10.1007/BF01596483

Cross, N. (2004), "Expertise in design: an overview", *Design Studies*, Vol. 25 No. 5, pp. 427–441.

Cross, R., Borgatti, S.P. and Parker, A. (2002), "Making Invisible Work Visible: Using Social Network Analysis to Support Strategic Collaboration", *California Management Review*, Vol. 44 No. 2, pp. 25–46. https://doi.org/10.2307/41166121

de Vries, R.E., Bakker-Pieper, A. and Oostenveld, W. (2010), "Leadership = Communication? The Relations of Leaders' Communication Styles with Leadership Styles, Knowledge Sharing and Leadership Outcomes", *Journal of Business and Psychology*, Vol. 25 No. 3, pp. 367–380. https://doi.org/10.1007/s10869-009-9140-2

Ehrlich, K. and Cataldo, M. (2014), "The Communication Patterns of Technical Leaders: Impact on Product Development Team Performance", *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ACM, New York, NY, USA, pp. 733–744. https://doi.org/10.1145/2531602.2531671

Krebs, V. (2000), "InFlow Survey", available at: https://web.archive.org/web/20070418102219/; http://www.orgnet.com/INSNA/survey.html (accessed 4 April 2017).

Kress, G.L. and Schar, M. (2012), "Teamology - The Art and Science of Design Team Formation", in Plattner, H., Meinel, C. and Leifer, L. (Eds.), *Design Thinking Research: Studying Co-Creation in Practice*, Springer, pp. 189–209. https://doi.org/10.1007/978-3-642-21643-5_11

Maier, A.M., Kreimeyer, M., Hepperle, C., Eckert, C.M., Lindemann, U. and Clarkson, P.J. (2008), "Exploration of Correlations between Factors Influencing Communication in Complex Product Development", *Concurrent Engineering*, Vol. 16 No. 1, pp. 37–59. https://doi.org/10.1177/1063293X07084638

Martins, J.R.R.A. and Lambe, A.B. (2013), "Multidisciplinary Design Optimization: A Survey of Architectures", *AIAA Journal*, Vol. 51 No. 9, pp. 2049–2075. https://doi.org/10.2514/1.J051895

Newman, M.E.J. (2018), *Networks: An Introduction*, 2nd ed., Oxford University Press, New York, NY.

Novoselich, B.J. and Knight, D.B. (2018), "Shared Leadership in Capstone Design Teams: Social Network Analysis", *Journal of Professional Issues in Engineering Education and Practice*, Vol. 144 No. 4, p. 04018006. https://doi.org/10.1061/(ASCE)EI.1943-5541.0000376

Papalambros, P.Y. (2015), "Design Science in Design Education", XI Mudd Design Workshop on Design Thinking in Design Education, Claremont, CA.

Papalambros, P.Y. and Wilde, D.J. (2017), *Principles of Optimal Design: Modeling and Computation*, 3rd ed., Cambridge University Press. https://doi.org/10.1017/9781316451038

Salton, G. and McGill, M.J. (1986), *Introduction to Modern Information Retrieval*, McGraw-Hill, Inc., New York, NY, USA.

Scikit-Learn Developers. (2018). "TfidfVectorizer", available at: https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html (accessed April 2018).

Senior, B. (1997), "Team roles and team performance: Is there 'really' a link?", *Journal of Occupational and Organizational Psychology*, Vol. 70 No. 3, pp. 241–258. https://doi.org/10.1111/j.2044-8325.1997.tb00646.x

Sheard, S.A. (1996), "Twelve systems engineering roles", *INCOSE International Symposium*, Vol. 6, pp. 478–485. https://doi.org/10.1002/j.2334-5837.1996.tb02042.x

Smith, R.P. and Leong, A. (1998), "An Observational Study of Design Team Process: A Comparison of Student and Professional Engineers", *Journal of Mechanical Design*, Vol. 120 No. 4, p. 636. https://doi.org/10.1115/1.2829326

Sonnenwald, D.H. (1996), "Communication roles that support collaboration during the design process", *Design Studies*, Vol. 17 No. 3, pp. 277–301. https://doi.org/10.1016/0142-694X(96)00002-6

Stewart, G.L., Fulmer, I.S. and Barrick, M.R. (2005), "An Exploration of Member Roles as a Multilevel Linking Mechanism for Individual Traits and Team Outcomes", *Personnel Psychology*, Vol. 58 No. 2, pp. 343–365. https://doi.org/10.1111/j.1744-6570.2005.00480.x

Thorndike, R.L. (1953), "Who belongs in the family?", *Psychometrika*, Vol. 18 No. 4, pp. 267–276. https://doi.org/10.1007/BF02289263

Watts, D.J. and Strogatz, S.H. (1998), "Collective dynamics of 'small-world' networks", *Nature*, Vol. 393, pp. 440–442. https://doi.org/10.1038/30918

ACKNOWLEDGEMENTS