

Observing network characteristics in mass collaboration design projects

Zachary Ball¹ and Kemper Lewis¹

¹ *Department of Mechanical and Aerospace Engineering, University at Buffalo, Buffalo, NY 14260, USA*

Abstract

Mass collaboration efforts can increase innovation and design possibilities. Incorporation of open innovation into the product development process allows for a vast array of unique perspectives and ideas. However, with the broad expansion of design possibilities, coordination of these development processes is paramount. To best make use of open innovation in product development, increased organizational efforts must be considered. The mass collaboration of individuals must account for individual intellectual abilities (competencies), working experience and even personality traits or idiosyncrasies. Approaches to this problem require the fusion of social network analysis with quantifiable design impacts. This work proposes a simulation framework that evaluates the design potential of a project team based on individual attributes and the team network structure. The overall contribution of this work comes from the exploration of team structure, focusing on network composition metrics such as centrality and network density, while attempting to understand the role of individual ability and positioning on the success of the design process. This work aims to garner a more thorough understanding of how the network structure of design teams correlates with their potential performance through a generalized simulation framework, applicable to future crowd and design initiatives.

Key words: mass collaboration, design projects, collaborative design

Received 31 March 2017
Revised 10 November 2017
Accepted 14 November 2017

Corresponding author

K. Lewis
kelewis@buffalo.edu

Published by Cambridge
University Press
© The Author(s) 2018
Distributed as Open Access under
a CC-BY 4.0 license
(<http://creativecommons.org/licenses/by/4.0/>)

Des. Sci., vol. 4, e4
journals.cambridge.org/dsj
DOI: 10.1017/dsj.2017.26



1. Introduction

Mass collaboration efforts have been applied using a variety of techniques. Crowdsourcing has been employed by numerous organizations as a method to garner ideas from the masses through the use of design challenges (Howe 2006). This method of mass collaboration generally places the organizational structure of the effort under the sponsoring company, only allowing for the crowd to generate potential solutions without complete engagement in the design process (Brabham 2008). Open-innovation methods have helped to increase the crowd's contribution by actively engaging individuals throughout the design process. These challenges are still carried out under the umbrella of the sponsoring organization; however, there is greater collaboration between the individuals within the organization and those supporting the open-innovation process (Shirky 2008). Continuing with increased individual engagement, open source projects allow for the greatest inclusion of individuals, as these projects allow for direct contribution to the design effort.

From crowdsourcing to open source projects, an increase of the labor force on a project opens many additional channels for unique perspectives. With the cost of communication drastically decreasing, individuals across the globe can virtually communicate instantaneously. The increase in communication possibilities has led to the promotion of the spread of knowledge and ideas where previously these initiatives were either too organizationally complex or cost prohibitive. The guiding idea of this phenomenon is that solutions gathered from the many increase innovation and variety when compared with solutions obtained from a single source (Surowiecki 2005; Tapscott & Williams 2008). This idea has been successful in many product design, innovation and social change initiatives (Change.org 2017; InnoCentive 2017; NASA NOIS 2017; OpenIdeo 2017).

The production of products or software can be generally categorized into two main approaches, namely traditional production and user production (Brabham 2013). Traditional production focuses on a top-down process contained within an organization in a hierarchical form. User production focuses on a bottom-up approach, allowing the community to generate content without much level of organization, in a somewhat self-governing manner (Panchal & Fathianathan 2008). Our work primarily focuses on user production while supplementing the organizational component of traditional production with social network analysis to allow for greater user production while also minimizing expensive managerial overhead (Hamel 2011).

The success of open-innovation efforts requires a shared design initiative, a pool of incentivized individuals and the organization of individual efforts (Chiu, Liang & Turban 2014). The shared design initiative can come from many sources such as consumers or even the crowd itself; however, the specifics of the project being developed and its inception are topics for future work. The individual labor pool must be composed of large masses who are willing to contribute, hinging on a provided incentive (Brabham 2010; Panchal 2015). For example, Amazon's Mechanical Turk boasts having over 500,000 workers, emphasizing the number of individuals willing to contribute to shared projects. While these projects are generally repetitive and simplistic crowdsourcing efforts, they do highlight the number of individuals willing to contribute.

One of the less developed components of current mass collaboration efforts, which is needed to develop complex projects, is increased structure and organization, backed with rigorous network development. In order to take full advantage of all that mass collaboration has to offer, the organizational structure behind the crowd must be fully understood and managed. For that, this work explores social network analysis (SNA) metrics and their potential application toward design networks.

Social network analysis makes use of the analysis of network graphs, which consist of a combination of nodes and edges. The nodes can represent individuals, groups or locations while the edges represent flows of information such as ideas, concepts or even physical items such as packages. Edges within a network can also be directed and weighted depending on the network that is being analyzed. In directed networks, the direction of the information flow along edges is considered. In weighted networks, these edges can carry higher or lower importance depending on what information is being passed. The combination of nodes and edges also leads to network and individual metrics that describe the network and the roles of the individuals within it. The use of SNA to study

mass collaboration initiatives allows for the visualization and understanding of individual efforts on specific portions of the design initiative (Wu *et al.* 2016). Social network analysis also allows for the potential identification of cognitive biases. One potentially destructive area in group decision making is ‘*group think*’, which can corner a design team into one solution (Esser 1998). With this knowledge, efficiently placed information streams can be opened to increase the collaboration efforts and thus increase the design potential of the group.

This work evaluates simulated teams of individuals working on a shared design initiative, all while monitoring the network structure of each group. The teams comprise subsets of individuals within a crowd made up of members with unique experiential and educational backgrounds. The outcome of this work is demonstrated through the use of randomly developed networks, using three network generation models to simulate varying personality characteristics of crowd members. These networks are measured on the probability of product development success, while also being used to understand the effect of various social network metrics within each network.

The precise quantification of individual or collaborative design improvements is difficult to measure given the key performance indicators (KPIs) being studied. With this in mind, a predicted design score is evaluated based on the assumption that individuals with greater domain knowledge and experience will generally map to more effective design solutions. The total project success in this work is based on KPIs characterized within four fundamental domains: demand (sales), innovation, manufacturability and quality. Within this work, these four areas are of equal importance; however, they can be adjusted based on the project being proposed and the KPIs that are identified as having greater significance to the stakeholders.

The result of this work links team composition and network structure with predicted design success. The overall outcome of each project design team is compared against these network statistics to observe any correlations that could help to guide the future development of design teams. This work aims to simulate a design scenario that relies on the self-organization of openly formed mass collaboration efforts based on individual characteristics. This type of scenario would be similar to an open source software development project, but translated into an engineering design context. In this new context, CAD models and design variables become equivalent to the source code in an open source project. The results found within this work highlight the potential of this simulation framework, while any correlations found are specifically attributed to the group of individuals utilized within the analysis.

2. Motivating work

The primary motivation for this work is derived from the analysis of mass collaboration design projects in addition to the use of SNA and team formation concepts.

2.1. Mass collaboration

Mass collaboration can be found in many forms, from companies or governments utilizing crowdsourcing to individuals grouping together on open source software projects. Schenk and Guittard performed a review of mass collaboration

opportunities (Schenk & Guittard 2009), where they highlighted geographical mapping through OpenStreetMap, the digitization of archives through ReCaptcha, and content analysis through Amazon's Mechanical Turk. Each of these requires virtually no previous training or background knowledge, forcing the crowd participation to remain simplistic and the success of the projects to be based on the number of participants performing similar tasks. These tasks are considered to be micro tasks and are open to anyone who wishes to participate.

This concept is extended through the use of 'gamification'. Here, complex problems are presented to the crowd in the form of a game in which players are rewarded for better solutions, also considered as an application of 'crowd science' (Franzoni & Sauermaun 2014). One significant example is the game Foldit, where the crowd was tasked to develop protein structures. Launched in May 2008, the game had attracted 50,000 users by September 2008 and had already outperformed one of the top algorithms in the field (Cooper *et al.* 2010). The process of 'gamification' has even seen applications in vehicle design (Ren, Bayrak & Papalambros 2015). The application of these tasks requires a brute force approach that demands a large number of individuals to participate. Moreover, due to the lack of cross-communication between individuals, no organizational network needs to be established. This concept, however, does not lend itself well to complex system design where the required actions of each individual can take significant time and collaboration, all while incorporating specialized skill sets.

While tasks such as these represent the majority of current crowdsourced projects due to their ease of implementation, more complex tasks are also currently being performed. Focus has also been placed on tasks that require some previous knowledge or a specific skill (Brabham 2008). This work studies previous examples of successful crowdsourced designs such as t-shirt design on *Threadless*, image capture through *iStockphoto* and more complex design tasks through *InnoCentive*, which is more closely described as open innovation. Crowdsourcing has even seen applications within the realm of public policy generation and infrastructure design (Brabham 2009; Aitamurto & Landemore 2015; Certoma, Corsini & Rizzi 2015). These projects require more industry-specific knowledge to complete; however, they still remain open to anyone who would like to participate. The concept of open access is paramount to the mass collaboration process. While it may be beneficial to have experience in photography to submit photos to *iStockphoto*, it is not required. By not limiting those who can contribute, these projects all open the door for potentially very unique and innovative solutions (Surowiecki 2005). These examples highlight the potential for individual competency requirements on participants; however, since they still primarily rely on crowdsourcing, the network of individuals does not need to be established.

The quantification of individual competency has been approached by analyzing current crowd development platforms. Recent effort has been made to understand which ideas/users should carry higher weight and which ideas/users should be given less consideration. Burnap *et al.* examined this idea by attempting to identify the experts within a heterogeneous crowd (Burnap *et al.* 2015). They made the underlying assumption that low-expertise members are more likely to guess toward a solution while expert members will be much more consistent in their evaluations. While a cluster of experts was identified, multiple other clusters were identified which skewed the results. Due to the difficulty of assigning ability based on previous projects, our work assumes that individuals within the

crowd have known levels of experience and ability. In this work, decomposition of abilities is applied in a manner similar to the process used in Takai (2010). Here, Takai estimated the complementary abilities of two individuals by decomposing their individual skills into vectors composed of knowledge domains. The degree of orthogonality between the two competency vectors was calculated to determine the degree to which they complemented each other's domain-specific knowledge. By evaluating the collaborative abilities of all members within a group, the combined ability of large groups could be estimated as a culmination of the individual skills of each member, and the combined abilities of each collaboration. This concept leads to the evaluation of the overall ability of a network of individuals based on their individual competencies and network ties.

2.2. Network analysis

Current network analysis research primarily focuses on analyzing developed networks to discover what characteristics are observed and then attempting to explain the causes of these characteristics and their subsequent outcomes. Borgatti and Foster performed a comprehensive review of current network structures from an organizational standpoint (Borgatti & Foster 2003). Some key aspects of their review highlight research carried out in social capital theory and the effects of relating individual connections with overall organizational performance. This idea is further exemplified in their discussion of organizational networks and the idea of what industry conditions can lead to stronger connections.

In subsequent work, Borgatti continued on to emphasize the flow characteristics as they relate to various network centrality measures (Borgatti 2005). The flow characteristics of information or ideas are very different from the flow of physical items such as packages. Certain centrality metrics may be better suited based on the flow characteristics being observed. This idea is addressed during the discussion of the centrality measures applied in our work.

Application of the idea of network analysis to a crowdsourcing network has been performed on an already successful crowdsourcing platform, OpenIDEO (Fuge & Agogino 2014; Fuge *et al.* 2014). These works evaluate the community structure of multiple OpenIDEO projects by looking into the network composition, determining the overall structure of these networks and how they compare with other social networks. However, members within these networks do not specify individual competency levels.

The application of SNA to design initiatives must also be complemented with an understanding of team formation concepts. The team formation problem has been addressed most frequently in operational and managerial research and has been proven to be NP-hard (Lappas, Liu & Terzi 2009). These problems are best suited to a more heuristic approach such as genetic algorithms or agent based modeling (Panchal 2009). In addition, Farasat and Nikolaev presented a mathematical framework for the optimization of the team formation problem with the inclusion of social network theory (Farasat & Nikolaev 2016). This work highlighted some of the key components in team development such as social exchange theory (Wasserman & Faust 1994; Contractor, Wasserman & Faust 2006) and homophily (McPherson, Smith-Lovin & Cook 2001).

Team formation based on previous knowledge and collaboration potential has also seen recent development (Fitzpatrick & Askin 2005; Hahn *et al.* 2008; Wi *et al.* 2009; Dorn & Dustdar 2010; Feng *et al.* 2010). However, these works

do not approach the problem from a mass collaboration perspective. One notable development in this field is the skill extraction method (Dorn & Dustdar 2010). In this work, expert behavior was observed from online posts to determine which skills each member possessed, which then could be used for competency assignment. The matching of individual skills to project difficulty has also been explored to build near-optimal teams based on skill coverage and team connectivity (Zhu, Huang & Contractor 2013).

The product side of engineering design has also seen applications of SNA to determine consumer–product relations (Wang *et al.* 2015), mine product features (Tuarob & Tucker 2015) and model distributed designs (Cormier, Litterman & Lewis 2011). Our work applies a basic product model to assess the design ability of each network; however, further development of this approach could support the addition of a network driven product model resulting in a multi-dimensional network including both the connected product components and individuals.

The contribution of this work lies in the development of mass collaboration initiatives driven by the network structure of individuals with known abilities and experience. The combination of these components allows for a comprehensive network evaluation for open and distributed design efforts.

3. Simulation framework for organizing and quantifying design efforts

3.1. Network development

Before we can begin to analyze team network structures, a pool of individuals must be available with known background characteristics such as educational attainment, previous work experience or individual characteristics. These characteristics help to describe the potential for value added work in each design initiative. Each individual must have a corresponding skill set to denote their abilities as they relate to team functions. The set of individual abilities is defined as the individual's competency vector, to be introduced in Section 3.3.

3.1.1. Definitions and notation

First, we must define the individuals and their corresponding skill sets. Let S be a set of all individuals i , where $i = 1, 2, \dots, N$. Here, N represents the total number of individuals within the data set and we define M as the total number of individuals on each team. Let C represent a matrix of corresponding skills, where c_i is the competency vector of individual i . Each individual also has a corresponding trait vector T_i to be utilized in the development of communication links between individuals. Each design team is represented as a network graph $G(V, E)$, where V represents the set of individuals and E represents the set of edges between individuals for each design team.

Graph theory allows for the observation of pairwise connections between various components (Wilson 1996). In an unweighted graph, a connection between node v_i and node v_j is represented by $e_{ij} = 1$ and an absence of a connection is given by $e_{ij} = 0$. Throughout this work, all graphs represented are unweighted. Weighted graphs apply a stronger or weaker connection between nodes for instances where the amount of information traveling among the edges depends on which nodes they pass between. However, since this work supports the development of newly formed design teams, previous communication history is

disregarded. Throughout this work, nodes represent individual actors within the networks and the edges between them represent information flows, such as shared design variables or design components.

When an edge exists between two nodes, those nodes are considered to be adjacent. Each graph has an associated adjacency matrix represented by a binary $n \times n$ matrix A (Borgatti & Everett 2006). Within the adjacency matrix node, v_i is adjacent to node v_j if both $a_{ij} = 1$ and $a_{ji} = 1$. For undirected graphs, the adjacency matrix is symmetric. Graphs can be either directed or undirected; however, for this work, undirected graphs will be utilized, as the information flowing between two members is assumed to be mutually available.

Non-adjacent nodes can pass information between one another by traversing through intermediary nodes. This information can follow a walk, a path or a trail. A walk from node v_i to node v_j follows a sequence of adjacent nodes, beginning with node v_i and ending on node v_j . A trail is similar to a walk; however, each edge can only be used once; a path is similar to a trail; however, each node and edge cannot be repeated (Borgatti & Everett 2006). A geodesic represents a walk between two nodes in which the length of the walk is the shortest possible between these two nodes. Different modes of travel are beneficial for specific pieces of information and measures of centrality (Borgatti 2005).

Within a crowd based design project, relations between members are very ill-defined as the crowd is theoretically composed of thousands of individuals who have most likely not worked together on previous design projects. Because of this, communication among members of the crowd must be simulated to develop network groupings. For this work, random intersection models are employed.

3.1.2. *Random intersection model*

A random intersection graph is utilized to develop modes of communication between members (Karonski, Scheinerman & Singer-Choen 2013). As the members of an open-innovation project are unknown to one another at initial conception, this method of network generation is used to generate connections based on matching personality traits of individuals within the crowd. Random intersection graphs can provide purely random connections between members, as well as guided connections based on individual preferences or previously known characteristics (Deijfen & Kets 2009). This concept is explored by developing communication flows between members based on three network development techniques.

First, a random network is developed, where each individual is randomly assigned a set of 12 possible traits. These traits are represented by binary values, with one indicating that an individual exhibits the specific trait and zero indicating that they do not possess that specific trait. Twelve traits are used as the network pool for this work consists of six unique disciplines, thus allowing for a mapping of two traits to each discipline. Additional traits allow for a greater decomposition of each individual personality, which is a topic of future work. Second, a guided network is developed, where there exists a higher probability for individuals of similar backgrounds to share similar traits. Finally, the third developed network consists of connections where individuals only communicate with others who have identical assigned skill vectors.

It is important to note that the random nature of the developed networks is a result of the trait assignment and not edge allocation. The edge placement between

Table 1. Sample trait matrix T for the full random intersection model

Individual	Trait Number											
	1	2	3	4	5	6	7	8	9	10	11	12
1	0	0	1	0	1	1	1	0	1	0	1	0
2	1	0	1	0	0	1	1	0	0	0	0	0
3	1	1	0	1	1	0	1	0	1	0	1	1
4	1	0	1	1	0	0	0	1	0	1	0	1
5	0	0	0	1	1	0	1	0	1	0	0	1

nodes is a result of the similarity between members, which, in the context of open design, supports the notion that two individuals with common interests are more likely to work with one another. In alternative random network generation methods, such as the Erdős–Rényi model, the random nature of the network comes directly from the assignment of edges (Erdős & Rényi 1960).

All three models work on the principle of developing information flows between members with similar interests; however, the difference lies in the assignment of those interests. Once the interests have been individually assigned, we look for commonality of interests between members to determine whether a connection should be made between these two individuals based on a threshold of common interests. Threshold values represent how many traits individuals must share in order for a link to be drawn between them. The effects of different threshold values are studied further in Section 4.2. The specific degree of commonality is variable, allowing for multiple combinations of network connections. The explicit development of the three models is described below.

Full random model. The full random model assigns 12 discrete traits for each individual, with equal probability of that individual having a specific trait. This process ensures complete randomness, with respect to their design abilities, when considering which individuals are able to communicate with one another. During this process, all traits are treated equally, with no explicit regard to their corresponding meaning. The result of this model when considering five individuals with 12 traits leads to the creation of Table 1 which exhibits no obvious pattern between individuals and their corresponding traits.

From Table 1, it is observed that individuals 1 and 2 share three trait assignments. Assuming a threshold of three, a communication link would be generated in this case. Individuals 4 and 5 only share two common traits, indicating that with a threshold of three, a communication link would not be generated.

Probabilistically guided model. The probabilistically guided model allows for members of similar skill sets to have a higher probability of similar traits. For this, each trait must exhibit a correlation between the individual’s skill set and their specific traits. As this work focuses on six areas of background knowledge, further described in Section 4, traits will be mapped to the following: development, marketing, production, quality management, research and sales. While these areas are not all inclusive to a design process, they do provide a broad look at the functional components of a design team. Further decomposition of the functional processes of a team and trait mapping can be performed for networks of different skill compositions. The mapping of these traits is shown in Table 2.

Table 2. Mapping traits to background characteristics

Background	Traits
Development	1, 2
Marketing	3, 4
Production	5, 6
Quality	7, 8
Research	9, 10
Sales	11, 12

Directed model. The directed model creates communication connections between members of identical disciplines. This model leads to the network with the greatest segregation as only members with identical background knowledge communicate with one another.

Following the completion of the trait matrix, individual communication links can be formed based on the similarity of traits between individuals. These links are shown through the development of the adjacency matrix.

3.1.3. Forming the adjacency matrix

Each trait matrix only gives insight into which traits each individual possesses. To develop connections within the network, individuals with similar interests and traits are connected with respect to an overall threshold of similar interests. As this threshold increases, fewer connections become present, as individuals must be extremely similar to one another in order to develop communications. With a lower threshold, networks begin to increase in density, as the required trait similarity between individuals looking to form a connection decreases. All three aforementioned models of trait distribution follow the same process of adjacency matrix formation shown below, with the only difference being that the directed model must have a threshold of two, as each level of background information is only mapped onto two traits.

The application of required trait similarity variances is used to observe the effects of homophily. The principle of homophily is based on the idea that individuals with similar characteristics are more likely to develop network connections (McPherson *et al.* 2001). Individual characteristics such as age, religion, education or occupation have all been observed to impact the dynamics of networks. For example, the directed model exhibits high levels of occupational homophily, as individuals are connected solely on their product development abilities. As the number of required trait similarities increases, the homophily within the network also increases.

Upon creation of the adjacency matrix for each network formation mode, we can visually observe the formed team. The sample network, shown in Figure 1, is developed using the probabilistically guided network, in which individuals with similar competencies have a higher probability of sharing similar traits. We observe clusters of individuals indicated by their node color; however, unlike the directed network, we also observe cross-departmental network communications.

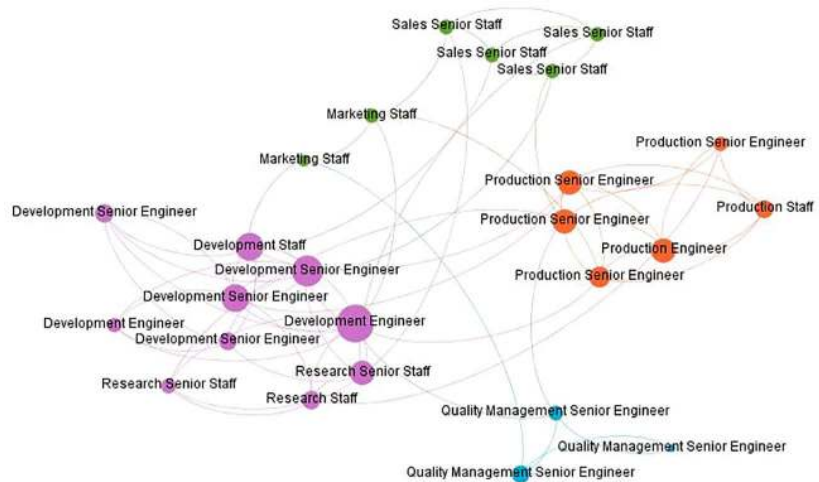


Figure 1. Sample network graph.

3.2. Network parameters

Upon completion of the random team formation, each team is analyzed for key network attributes including closeness, betweenness, eigenvector centrality, diameter and density. The degree of each member is taken into account, with specific consideration to understand which group members communicate with one another. Each of these metrics is averaged across an entire design team to review any correlations between network composition and team performance. The top performing and worst performing teams are given extra consideration to understand the individual characteristics of those network graphs. The applicable network parameters are summarized in Table 3.

3.3. Predicted design score

The calculated design score represents the potential design ability of each team. The values from the design score are purely for comparative purposes and do not have a real world correlation. These values serve as a surrogate for design success, and while they present a greatly simplified estimation, they address core elements of a design process for comparative purposes. The intent is for high design scores to represent a strong development team with the incorporation of four major KPIs within the design, while low design scores represent a design team with little probability of success. Within this section, new notation and definitions must first be explained before the methodology is presented.

3.3.1. Notations and definitions

For this work, let c_i represent the competency vector of each individual. The competency vector, as introduced in Takai (2010), presents the ability of each individual based on a variety of distinguishing factors. The impact of the competency vector is presented through the use of a value added metric, V , where each individual, i , contributes to the value of the design through $V(i)$. The total value added metric is represented as $V_{D,I,M,Q}$, representing the specific KPI of

Table 3. Summary of network characteristics

Parameter	Equation	Design implication
Closeness	$C(G) = \frac{1}{\sum_j [d_{ij}]}, i \neq j$	<i>Closeness centrality</i> is calculated as the inverse of the sum of distances, d_{ij} , between node i and all other nodes within the network. This measure of centrality assumes that information flows through the shortest possible path. Closeness centrality has been linked to reduced R&D times within organizations (Borgatti 2005). In the context of information sharing in an open design network, individuals of high closeness have more immediate access to information developed by the other individuals within the network.
Betweenness	$B(G) = \sum_i \sum_j \frac{g_{ikj}}{g_{ij}}, i \neq j \neq k$	<i>Betweenness centrality</i> is calculated as the number of geodesic paths that pass through node k divided by the number of geodesic paths that exist between nodes i and j . Like closeness centrality, betweenness assumes that all information travels on the shortest path between two nodes. From the standpoint of an open design network, individuals with high levels of betweenness pass increased levels of information throughout the network. The loss of these members could significantly impact the network due to their high level of information sharing.
Eigenvector	$\lambda v = Av$	<i>Eigenvector centrality</i> is determined as the elements of vector v , where A is the adjacency matrix and λ is the eigenvalue. This measure of centrality indicates that a node with a high eigenvector score is adjacent to other nodes with high eigenvector scores. This would mean that members that are well connected within the group are also connected to other highly connected members. Eigenvector centrality has been used to identify highly influential members within a network.
Degree	$D(G) = \frac{\sum_i [D(n^*) - D(i)]}{[(N - 1)(N - 2)]}$	<i>Degree centrality</i> represents immediate influence within a design team, as it is defined as the number of connections incident upon a node. Higher levels of degree centrality within a network indicate that each member has a greater amount of incident information flow. Here, n^* is the node with the maximum degree, i is the current node and N is the total size of the network.
Diameter	$\text{dia}(G) = \max_{i,j} d_{ij}$	<i>Diameter</i> of a network is the maximum geodesic length between two nodes in the graph. Larger, less connected networks have a larger diameter. Networks containing isolates exhibit an infinite diameter as all nodes are not connected.
Density	$\rho(G) = \frac{2e}{N(N - 1)}$	<i>Density</i> represents the ratio of connections in a network against all possible connections. A network with a density of one contains all possible connections, indicating a completely connected graph. With respect to a product development team, higher density can increase knowledge sharing. However, it can also hinder the pace of the development process, as each member must share information with an increased number of individuals.

added value (D = demand, I = innovation, M = manufacturing, Q = quality). The specific design under review may have alternate or additional motivations, dependent upon the design or stakeholders involved, to determine the success of each design. However, the components of the design must be able to map to the abilities of the individuals within the crowd. Each component of the added value also has a corresponding weight, represented by $w_{D,I,M,Q}$.

During the calculation of the value added, each individual has a base score which is denoted by α . The base score of each individual represents their overall ability based on previous experience, while the competency vectors correspond to their specific domain knowledge. Various values for α were tested, including 10, 50, 100 and 1000 and while the overall score scaled up proportionately, the observed trends remained identical. Since the overall score is used for comparative purposes only, the precise assignment of α proved to be insignificant. Before beginning the development of the design scores, a few core assumptions must be introduced.

Assumptions. Given that the objective of this work is to study the relationship between design team formation and design success, some underlying assumptions must be understood. These assumptions are based on observations from the existing literature, but do not represent fully validated claims from an empirical perspective.

- (i) This simulation framework considers design improvements in a specific discipline to occur with the greatest effect when two members of the same discipline collaborate. Connections between dissimilar skill sets also improve the design, but with less overall impact. When there is no connection between members, the collaborative impact on the design score is not affected (Haque, Pawar & Barson 2000; Feng *et al.* 2010).
- (ii) Individuals with higher levels of experience will have a more profound impact on the overall design (McDaniel, Schmidt & Hunter 1988). Their experience is measured by their overall time spent at the company or working in a specific discipline. This consideration is represented through the use of each individual's base score.
- (iii) Additional individuals of the same discipline will follow a power law distribution of diminishing returns on design improvements. The order in which each individual of a specific discipline is added to the team determines the potential value that said individual can provide. Power law distributions have been shown to consistently prevail when observing social network structures where users are allowed to contribute openly (Barabási & Albert 1999; Barabási 2003). The use of the power law distribution causes members who initially work on a design component to generate the greatest impact on that component when compared against subsequent members.

Each team member has pre-assigned abilities and expertise based on their current job title and department, as shown in Table 4. This work evaluates six distinct and broad scope abilities. However, further decomposition of these abilities can lead to a model with higher resolution. Because of the current limitation of precise ability mapping, the competency matrix for this work is of size (6×6), representing six dimensions of skill decomposition.

Each column of this matrix shows the title of each member, while each row represents the corresponding ability. These columns form a competence

Table 4. Competence matrix

	Marketing	Sales	Research	Development	Quality	Production
Marketing	10	6	1	1	1	1
Sales	6	10	1	1	2	1
Research	3	1	10	7	3	2
Development	2	1	8	10	5	3
Quality	1	2	4	6	10	5
Production	1	2	4	6	6	10

vector for each member where each skill is assigned from 1 to 10. In this work, skills are assigned based on each individual’s working title, allowing for a slight overlap of abilities between disciplines. The development of a more complete skill decomposition is currently an active research topic (Kramer, Agogino & Roschuni 2016); however, it is outside the scope of this work.

The creation of the competency vectors allows for the evaluation of the collaborative ability of each individual. This process accounts for the similarity of abilities to develop the overall collaborative effect of individuals with different backgrounds. The degree of collaboration is calculated based on an adaptation of Takai’s degree of complementarity (Takai 2010), as follows:

$$\theta = \arccos \left(\frac{c_1 \cdot c_2}{|c_1| \times |c_2|} \right), \tag{1}$$

$$d = \frac{\theta}{90}, \tag{2}$$

$$l_{1,2} = e(1 - d). \tag{3}$$

The degree of collaboration, l , is calculated from the degree of orthogonality, d , between the competency vectors of individuals one, c_1 , and two, c_2 . While Takai’s original work looked at the degree to which each individual was able to complement the other individual’s skill set, this work takes $(1 - d)$ to indicate the ability of two individuals to collaborate on a specific design component. These values are in the range [0–1]. If two members have identical competency vectors, then they receive a degree of collaboration of $l = 1$, indicating that they have full ability to collaborate. If two members have zero common abilities, then they will have a degree of collaboration of $l = 0$, indicating that a collaboration effort on a specific component would not provide any additional design improvement. The final component, e , indicates whether a connection exists between members. This component can take on a value of $e = 1$ when a communication link exists and $e = 0$ when a communication link does not exist.

In the context of a design project, when two connected individuals share project attributes, the level of commonality between each individual’s abilities is directly related to their capacity to support the work of one another. If two members have identical abilities, the information shared can be worked on collaboratively by both parties, while if two members share zero common abilities, the design component would be worked on independently by each member, corresponding to a degree of collaboration of $l = 0$.

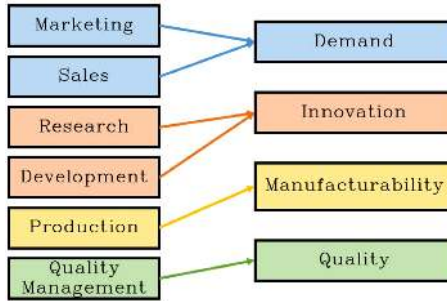


Figure 2. Direct mapping of ability to design component.

3.3.2. Value added calculation

To calculate the overall value added, each individual is assigned a competency vector based on their working title. The direct mapping of individual background to design component is shown in Figure 2. As collaborations are formed through network connections, we begin to observe overlap in these mappings accounting for the collaborative abilities of individuals.

Each individual member receives a value added score based on how many members of that discipline exist within that design team, following a power law distribution, and their overall experience. The initial individual of a specific discipline receives the most impactful score, while each subsequent individual increases the score, with a decrease in impact, following the value added function

$$V_{(i)D,I,M,Q}(x) = \frac{\alpha}{x} + \alpha y. \tag{4}$$

Here, x is the number of the individual in their corresponding discipline relative to when they were added to the team and y is the normalized days of employment. Days of employment are normalized between the most senior and the newest members in the pool of individuals to account for the experiential effects of individuals. The combination of value added across each individual and their collaborations is given as

$$V_{D,I,M,Q} = \sum_{i=1}^{i=T} \sum_{j=i}^{j=T} V_{(i)D,I,M,Q} \cdot l_{i,j}. \tag{5}$$

The inner loop calculates the sum of all collaborations for one individual, i , as they collaborate with the other individuals, represented by j . The outer loop repeats this process for each individual value, $V_{(i)D,I,M,Q}$, until all members of the team and all collaborations have been accounted for within team size T . It is worth noting that the index of the inner loop begins at $j = i$ as the networks developed are undirected, causing reciprocal collaborations to be identical.

The overall design improves as a result of the product of their individual value added function and their degree of collaboration for each discipline. This simulation framework supports strong collaborations between similar disciplines while still accounting for added benefit from complementary skill sets. However, connections between individuals with zero commonality are not supported as their resulting contribution reduces to zero. The specific design components that

increase in value are based on the title of the individual being analyzed and are proportional to their collaborative skill set.

These values are then combined in a weighted sum function to calculate the overall design score. For this work, all design attributes are weighted equally as the specific nature of the design and desired outcomes is not explicitly specified, as shown in Equation (6). For designs requiring higher aspects of innovation or quality, their weights, $w_{D,I,M,Q}$, could be adjusted accordingly.

$$\text{DesignScore} = w_D V_D + w_I V_I + w_M V_M + w_Q V_Q. \quad (6)$$

Following this calculation of each team's design score, the team structure is then reviewed as a network graph and social network metrics are calculated. At this point, each team can be evaluated for its composition and projected design ability. Any conclusions are specific to the group of individuals being considered and the assumptions being made, and the authors acknowledge the uncertainty in mapping individual abilities to design outcomes. However, the intent of this work is to develop a fundamental understanding of team dynamics in mass collaboration product development to be leveraged in practical applications.

3.4. Model validation

Verification of the results of the model proves difficult as a result of some of the simplifying assumptions applied in the calculation of the design score and the collaborative impacts of individuals. Because of this difficulty, this work looks toward Pedersen *et al.* (2000) to support the validation of this framework. Pedersen *et al.* decompose design method validation into four components: empirical performance validity, empirical structural validity, theoretical performance validity and theoretical structural validity.

The theoretical structural validity element is supported through the supporting literature for each assumption provided. This component requires the framework to rest on accepted constructs, each of which has been outlined where applicable. The empirical structural validity element is addressed using an applicable case study within the bounds of the framework. As this work aims to simulate design efforts in a mass collaboration environment, the case study outlined in the following section satisfies the empirical structural validity element of this framework. The application of the framework operating under the proposed assumptions allows for satisfaction of the empirical performance validity requirement.

The theoretical performance validity element arises from the ability to accept this framework beyond the presented case study. Due to the limitations resulting from the simplifying assumptions, this work is unable to satisfy the theoretical performance validity element. The assumptions made in this work force a limited applicability, and future work is required to observe whether this framework can provide similar results outside the bounds of this application.

4. Application

To apply the proposed simulation, a simulated heterogeneous crowd of approximately 180,000 members is utilized (Wang & Zaniolo 2015). This database is a simulated temporal data set used to model employees within an organization,

originally created to test database systems. Information about individuals includes their department, date of employment, age, salary and title. This information has been used in this work to provide additional attributes to the individuals of our generated crowd.

From the crowd, design teams are formed utilizing a subdivision of unique members who focus on specific design initiatives. The results are reviewed with distinct emphasis placed on network centrality, density and size. The following simulation study also includes a parametric analysis to understand the overall effects of the variation of team generation variables, such as team size and communication link threshold, utilized within the development. The inclusion of these variables allows for a more in-depth look at potential network structures as the network characteristics change. While the results cannot guarantee strict network development considerations, they highlight the usage of the simulation framework and its potential for team design.

4.1. Individual organization

The design capability of each group is evaluated based on the combination of skills in each developed team. While each individual may possess overlapping abilities across a range of disciplines, the decomposition of their abilities is represented to allow for a direct mapping between project goals and individual attributes.

To begin the analysis, we simulate 1,000 design teams, in which the organizational structure is an outcome of the random intersection model being applied. The first developed networks to be studied consist of 25 members, with a complete random intersection model and a communication link generation threshold of three. Application of a threshold of three indicates that two individuals must have a commonality of 25% within their pre-assigned traits, leading to networks of greater connectivity when compared to networks of higher commonality requirements. A level of three was also chosen to obtain a better understanding of how collaborations impact the design success. The following sections explore the resulting design scores and their distribution; overall network characteristics, including closeness, betweenness, eigenvector centrality, diameter, density and degree; and the top and worst performing design teams.

The potential number of combinations of individuals with a team size of 25 from a pool of 180,000 unique members is approximately 2.405×10^{131} , which does not even account for the possible connections that can be formed. The authors recognize that the generation of 1,000 teams only captures a very small fraction of the possible number of teams and connections available. Such a large search space lends itself to a more directed search using heuristic algorithms to intelligently search the solution space. This concept is currently being explored in other work (Ball & Lewis 2017); however, the primary motivation of this work is the exploration of network analysis metrics and how they relate to a variety of team formations and information flow characteristics.

4.1.1. Design score

To demonstrate that the team generation captured a wide variety of team compositions, Figure 3 shows the distribution of the design scores.

The distribution of the simulated design scores follows a normal distribution with an average of 936.5 and a standard deviation of 159.6. The top score from

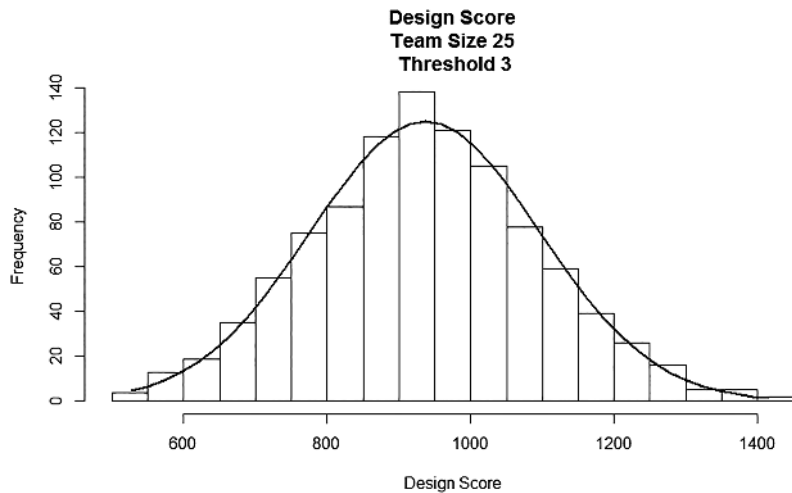


Figure 3. Histogram of design scores.

this set of teams is 1415.5 and the worst performing team scores a 526.6. It is important to obtain a complete coverage of design team potentials as this allows for teams of widely varying abilities to be studied. While the sample of teams only represent a limited portion of the potential combinations, given the approximate normal distribution, the authors believe that the results found can be considered characteristic of the pool and model being utilized. In the following sections, overall trends in network composition are discussed.

4.1.2. Network characteristics

Closeness centrality, shown in Figure 4(a), represents the number of information channels necessary for one individual to reach another individual and not the physical closeness of each individual. High physical closeness in a distributed mass collaboration context is rarely possible as members of the crowd are widely dispersed and they must have the ability to work with other individuals regardless of their location. Average closeness centrality follows a positive linear correlation with respect to design score.

This result yielded an R -squared value of 62.3%, indicating a reasonably strong correlation. Another thing to note is that this result excludes teams with a closeness value of zero, indicating the presence of isolates or incomplete network graphs. From this result, we conclude that increased levels of closeness centrality, indicating a shorter geodesic path between members, generate design teams with higher design potential. This idea points to the development of new information channels for members who currently experience individual levels of low closeness centrality, following the notion that direct lines of communication between members help to improve the collaborative efforts of the design team.

It is also worth noting that closeness centrality assumes that the item being passed between the edges follows the shortest path between nodes. Because of this assumption, the correlations found for this metric are most applicable to items that are spread in series, such as CAD models or shared design variables, as opposed to concepts or ideas, which adhere to more of a parallel duplication process.

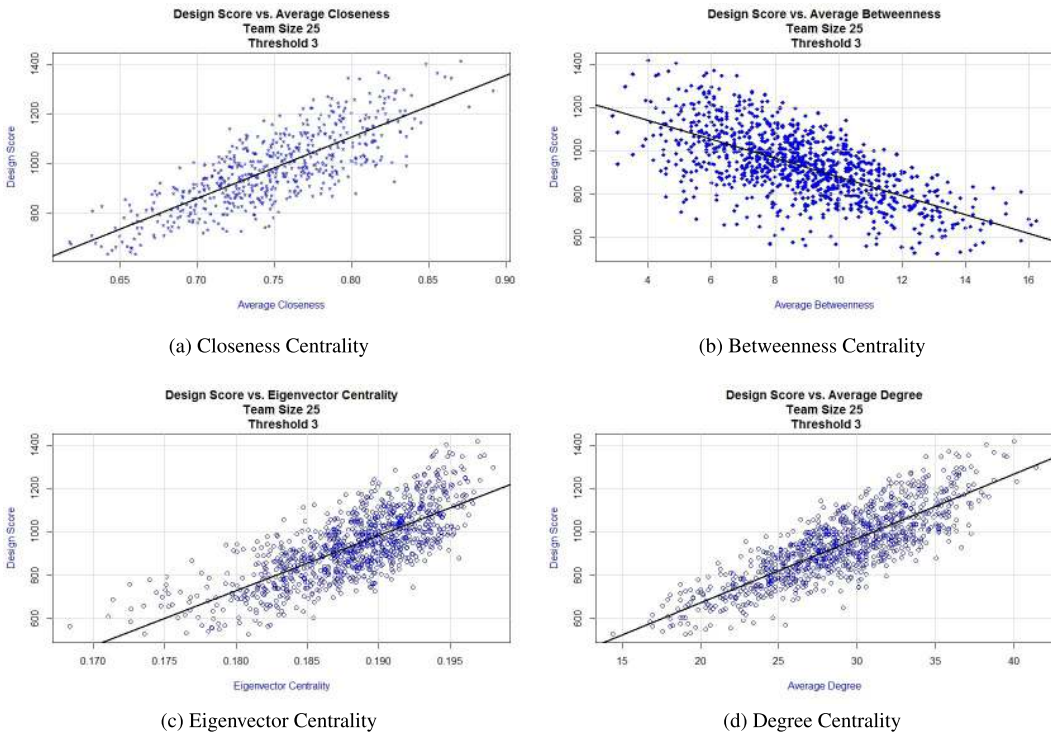


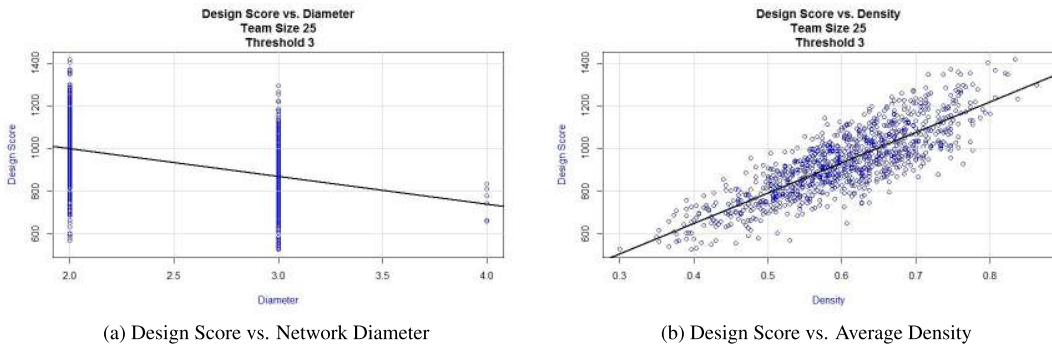
Figure 4. Network graph centrality indicators.

Betweenness centrality, shown in Figure 4(b), exhibits a negative linear correlation. The correlation between design score and average betweenness exhibits an R -squared value of 40.7%; however, it still has a noticeable effect on design performance.

Higher levels of betweenness centrality are known to indicate critical nodes, as they fall within the information paths of multiple adjacent nodes (Borgatti & Everett 2006). From these results, higher potential design scores have teams consisting of a lower number of critical nodes. This result is promising as teams with better design potential do not rely on any individual node to transfer large flows of information. The removal of any individual node would not have a significant impact on the entirety of the team.

As with closeness centrality, betweenness centrality assumes that information follows the shortest path between nodes. In the context of an open design initiative, this may not always hold true for concepts or ideas, indicating that nodes of high betweenness generally control the flow of specific items as opposed to ideas.

Eigenvector centrality, as shown in Figure 4(c), exhibits a positive correlation to design score with an R -squared value of 56.4%. Higher levels of eigenvector centrality lead to members of the design team having strong influences on other members of the design team. This consideration can also point to ‘group think’, reducing the variety of ideas and limiting innovation, as members of the design team can be persuaded to agree with influential members so as to follow the sentiment of the group.



(a) Design Score vs. Network Diameter

(b) Design Score vs. Average Density

Figure 5. Network graph characteristics.

Unlike closeness and betweenness, eigenvector centrality does not rest on the assumption of shortest path flows. Because of this, the results shown for eigenvector centrality apply to the transfer of ideas and concepts regarding designs. This implies that the observation of a positive correlation supports the diffusion of design concepts within the team.

Degree centrality, shown in Figure 4(d), shows a positive linear correlation between the average degree of the design team and the potential design score, with an R -squared value of 68.3%, indicating a relatively strong relationship.

Higher degree for each individual member, however, increases the amount of information flow each individual must be responsible for. As the degree increases, the level of individual involvement increases as they now receive information from additional members. This metric must be carefully leveraged so as to increase the degree of members where it is most advantageous for the entire network.

Degree centrality is also an indication of the immediate influence of a given node. In the context of a design effort, immediate influence is shown through the direct sharing of design variables or CAD models between two collaborating individuals. Since this metric only considers nodes directly incident upon one another, increased levels of degree centrality lead to more connected designs.

Three discrete values for *diameter* are observed, as shown in Figure 5(a). Networks of lower diameter provide stronger design potential. It is noted that the top performing design team had the lowest potential diameter while the highest values of network diameter all contained design scores that fell below the average. While diameter does not appear to be a strong indicator of design score, it can be concluded that networks where information channels between members must include multiple intermediary nodes perform worse.

The final network characteristic reviewed is the *density* of each network. From Figure 5(b), another strong positive linear correlation is observed as teams with greater density tend to generate higher design score potentials.

The density of a network is in direct relation to the average degree centrality of that network. The overall impact of density on design scores is a result of the same network dynamics as when average degree centrality was considered. However, density is an overall network characteristic while degree is individualized.

A multiple linear regression was also performed with respect to the design score and the four centrality metrics. However, due to the high multicollinearity between predictors, measured by their variance inflation factors (Craney & Surles 2002), the results were excluded from this work.

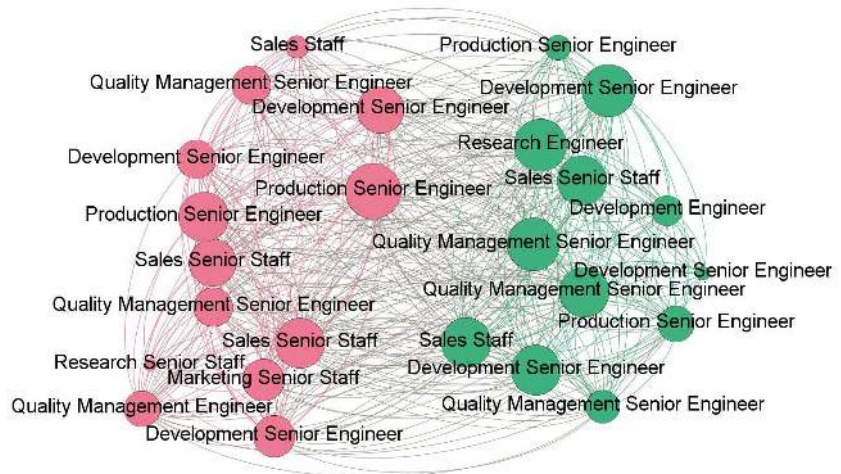


Figure 6. Top performing team network structure.

Table 5. Network properties of the top performing team

Diameter	Density	Eigenvector centrality	Average degree	Average betweenness	Average closeness
2	0.8333	0.1969	40	4	0.8708

4.1.3. Top performing network

The top performing design team received a design score of 1415.45. This network was very well connected, with strong connections between members of similar disciplines allowing for increased collaboration efforts, as shown in Figure 6. This network had a density of 0.833, indicating that approximately 83% of the potential connections between nodes were utilized. It also had a diameter of two, allowing for a close flow of information between all members.

As shown in Figure 6, node size is proportional to its degree, as it was previously determined that degree held the highest statistical significance when considering levels of network centrality. The top performing team has multiple nodes of high individual degree, supporting the spread of information within the network.

The colors for each node represent the cluster in which they belong. This network developed two distinct clusters of individuals represented by the pink and green nodes using a community detection algorithm based on a heuristic optimization approach to find high-modularity partitions, outlined in Blondel *et al.* (2008).

Table 5 highlights all network metrics attributed to the network graph of the top performing design team. Each of the network parameters outlined falls well above the average values determined from the entire simulation of the 1,000 teams.

Another consideration regarding the top performing team is related to the number of members from each discipline. This team is composed of a relatively even distribution of members, with the exception of only one marketing specialty and two research members, as shown in Table 6. The wide variety of disciplines

Table 6. Individual members on the top performing team

Background knowledge	Number of members
Sales	5
Marketing	1
Quality assurance	6
Research	2
Development	7
Production	4

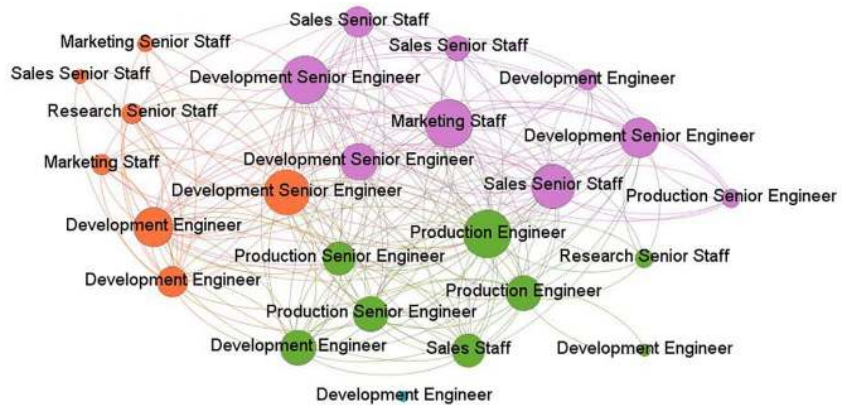


Figure 7. Worst performing team network structure.

allows for a well distributed design effort. The exact combination of individuals leading to the most successful designs would depend on the design task being considered.

4.1.4. Worst performing network

The worst performing design team received a design score of 526.64. From the network graph shown in Figure 7, it is observed that there were multiple members that had a degree of one or two, indicating that they were partially removed from the design effort. This led to poor information sharing from these members, thus decreasing their design score. This network also had a density of only 0.397, indicating that only approximately 40% of all possible connections were utilized.

When reviewing the network graph for the worst performing team, it is observed that there are fewer nodes of individually high degree, as the average degree for this team was significantly lower. It is also observed that there are four clusters that have formed, one of which is the isolated development engineer represented by the blue node. This lack of connectivity negatively impacts the team’s performance.

Table 7 highlights all network metrics attributed to the network graph of the worst performing design team. Each of these network properties presented falls

Table 7. Network properties of the worst performing team

Diameter	Density	Eigenvector centrality	Average degree	Average betweenness	Average closeness
3	0.3967	0.1809	19.04	13.12	0

Table 8. Individual members on the worst performing team

Background knowledge	Number of members
Sales	5
Marketing	3
Quality Assurance	0
Research	2
Development	10
Production	5

below the average values determined from the complete simulation of the 1,000 teams.

When reviewing the worst performing team for the distribution of individual members, it is observed that this team heavily consists of development engineers, with zero quality assurance engineers, as shown in Table 8. Because of this breakdown, the team does not sufficiently capture the entire design process, creating a poor overall design.

Comparison of the top performing team with the worst performing team furthers the idea that teams with greater connectivity, increased skill distribution and increased levels of information flow tend to create higher potential design scores. Another characteristic that is significantly different between the top and the worst teams is the experience level and variety of individuals on the team. The top performing team has 20 senior members, indicating greater ability in their respective disciplines, while the worst performing team only has 14 senior members.

4.1.5. Network generation comparisons

Next, we consider the impact of the communication link generation method, considering randomly formed networks, probabilistically guided networks and directed networks.

As shown in Figure 8(a), the network formed from random trait assignment consistently provided the most effective design team, with the probabilistically guided network receiving the second highest average marks and the directed network performing the worst of all three. These results, however, must be carefully interpreted as the average density, Figure 8(b), indicates that directed networks were also the least connected networks. It was previously identified that there exists a strong positive correlation between design score and network density, potentially leading to the variation in design scores observed.

The decreased overall design scores can be explained by the limited potential for collaboration efforts, expressed through a decreased network density, as the probability of communication links was decreased. The probabilistically guided and directed networks limit the overall amount of potential connections, as these

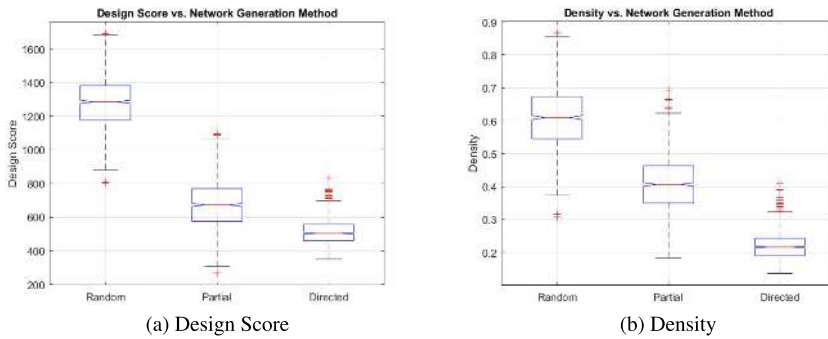


Figure 8. Network generation method statistics.

are now dependent on the variety of members within the design team. The limited potential for communication links can also be quantified by the average degree of each network, as it was only 19.3 for the partially random network compared with 28.9 for the fully random network generation. The directed network is the most restrictive as it only allows for individuals with the same background knowledge to communicate. The decreased design score is primarily attributed to the decrease in the probability of collaboration.

Figure 9 illustrates a summary of the network metrics across the three types of network formations. Random trait assignment lead to teams with the highest ability and networks of the greatest density. Partially random networks have the highest levels of betweenness centrality and largest average diameter.

The directed network received average closeness and betweenness values of zero as there were isolated groups within the networks. These isolated groups meant that no member within the network could communicate with any other members.

4.2. Parametric analysis

To further understand the effects of network construction, a parametric analysis is performed to observe the impact of varying levels of team size and communication link generation on design score and network structure characteristics.

As expected, it is observed that the design score increases with increased team size and decreased threshold value, as shown in Figure 10. As the number of individuals on each team increases, along with their probability of communication, shown through decreasing threshold values, the design scores also increase.

4.2.1. Network characteristics

Closeness centrality, shown in Figure 11(a), is studied relative to varying team size at constant lines of threshold values. Threshold values of one allow for networks of much greater density as individuals require only one trait similarity before links are formed, while threshold values of six require individuals to share half of their traits before they collaborate. As the team size increases, the average closeness centrality for each group decreases for threshold values of one and two. When looking at higher threshold values, the closeness centrality remains constant at

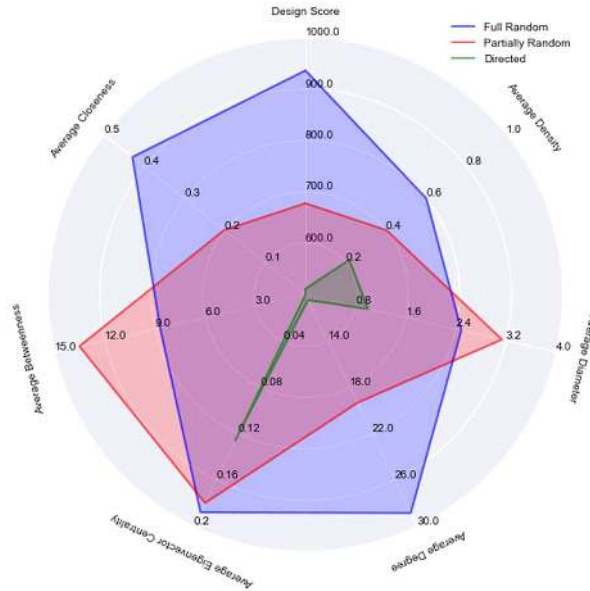


Figure 9. Summary of network statistics.

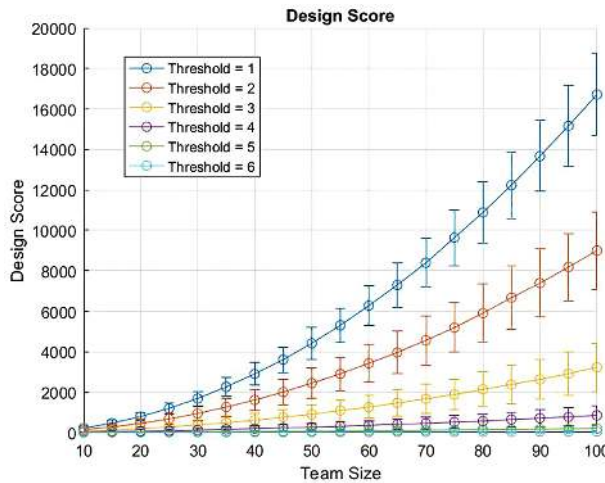


Figure 10. Design scores with 95% confidence intervals.

zero due to incomplete network graphs, with the minor exception of insignificant closeness levels for a threshold of three with team sizes of under 30 members.

Focusing on threshold levels of one and two, it can be concluded that network closeness decreases as team size increases. This is a result of increased team size leading to greater geodesic paths between individuals. Due to the addition of team members, the average distance between each node increases as communication between members now spans across a greater number of individuals. This result indicates that as teams are formed with significantly different individuals, their networks become less centralized.

When observing the betweenness centrality, shown in Figure 11(b), it is evident that as team size increases, the betweenness centrality of each network also increases. As additional members are introduced to the design team, connections between members have a greater chance of passing through other members, causing the average betweenness of the entire network to increase. Based on this, caution must be taken when developing networks of large team size as these contain additional critical nodes that control large information flows.

The impact of threshold level illustrates a curious result, as a threshold of three creates the most impactful change across varying team sizes, while a threshold of one has a less profound effect. Lower threshold levels lead to a greater overall probability of developing connections between individuals. Because of this increased probability, we observe that with a threshold of one, the betweenness does not increase as quickly as when looking at thresholds of two, three and four since the increase in connections also supports the development of direct lines of communication. When considering higher thresholds, the probability of an increased number of additional communication links decreases, also decreasing the number of direct paths between individuals, forcing information flows to pass through intermediary nodes.

We observe that at 25 individuals, the betweenness centrality of networks with a threshold of three begins to increase above that found for a threshold of two. This phenomenon can most likely be attributed to the decreasing probability of additional direct lines of communication in higher thresholds, causing networks with a threshold of three to increase betweenness at a greater rate as more individuals are added to the network. It is also observed that thresholds of five and six only exhibit a very minor impact as the probability of new communication links remains low, not significantly impacting each individual's betweenness level.

An increase in the team size creates a decrease in the eigenvector centrality, as shown in Figure 11(c), for each design team following a decreasing power regression line. Because this correlation follows a power regression, changes to team size for smaller teams have a much greater impact on eigenvector centrality when compared with larger team sizes. When team sizes remain small, eigenvector centrality is high due to the limited pool of potential member connections. At smaller team sizes, the probability of influential members connecting other influential members is higher due to the decreased pool of potential member connections. As team sizes increase, this probability decreases, causing the decrease in eigenvector centrality to a point where adding additional team members creates a negligible effect.

The correlation between average network degree and team size is shown in Figure 11(d). Degree centrality increases linearly with respect to team size as the number of potential connections increases due to the increased probability of other team members sharing the required number of traits for a connection.

Comparison of team size with network diameter, as shown in Figure 12(a), reveals results that do not show much discernible pattern across both team size and communication thresholds. It is also observed that these values had much greater variability, notably limiting any statistically supportable conclusions. One observation that can be noted, although cautiously, is that the average values of diameter appear to increase asymptotically toward a constant. This constant also appears to increase with increasing threshold value. Thresholds of one and two reach their constant values, diameters of 2 and 3 respectively, immediately.

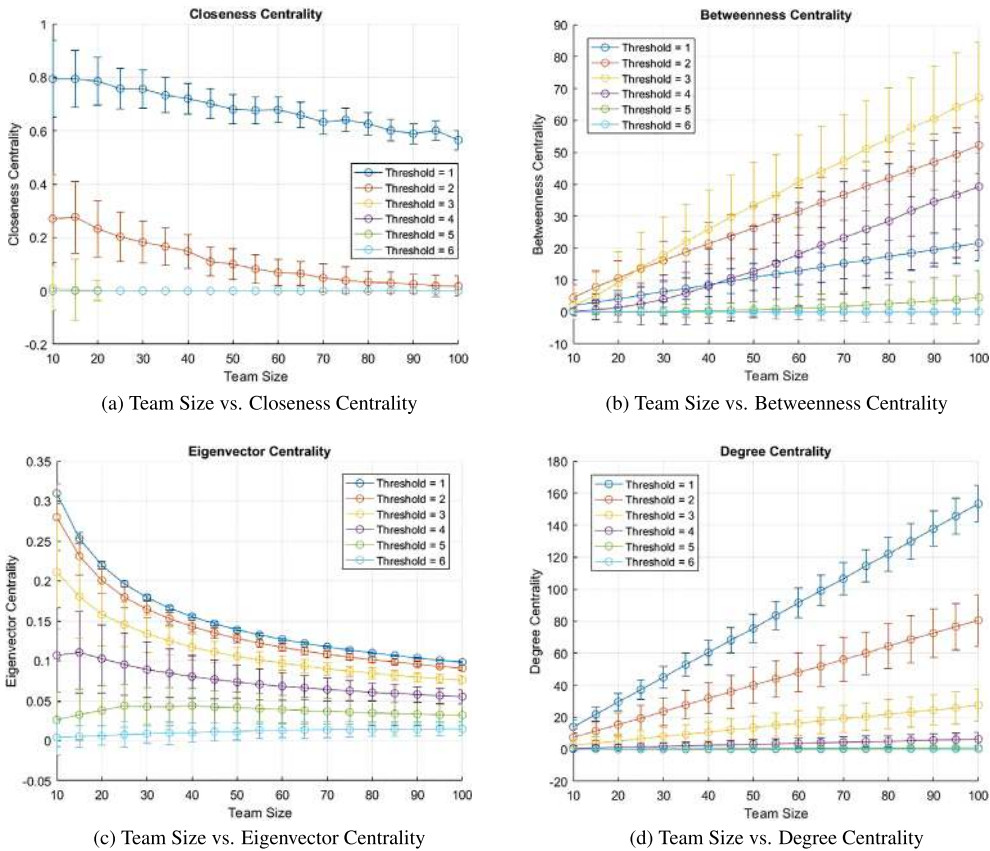


Figure 11. Social network metrics with 95% confidence intervals.

A threshold of three requires a team size of approximately 25 before a constant diameter of 4 is reached and a threshold of four reaches a constant value of approximately 5 at a team size of 55. It also appears that the deviation in the results is a function of team size as a threshold of one has a near-zero level of deviation for a team size of 60 or above. Based on this trend, the authors believe that if additional larger teams were considered, all threshold values would begin to settle at a constant diameter with decreasing levels of variability.

Networks generated with thresholds of three and four both follow increasing patterns along a positive power regression. Networks of threshold level three smooth out at a network diameter of four and networks of threshold level four smooth out at a network diameter of five. The cause of the increasing diameter level for smaller team sizes for these two threshold values is due to incomplete network graphs for low team sizes. With a low number of team members and a higher threshold value, isolates and disconnected clusters form, causing network diameters of zero. When taking the average across these networks, the incomplete graphs begin to disappear, as networks become more connected, causing the average diameter to level out around the constant value for each threshold. For example, this phenomenon occurs at 20 individuals for a threshold of three.

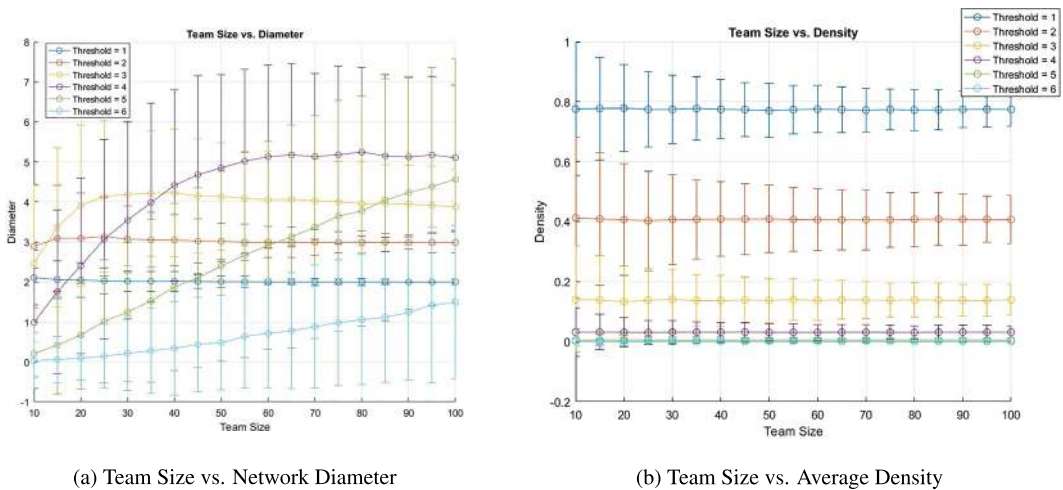


Figure 12. Network graph characteristics with 95% confidence intervals.

For networks with thresholds of five and six, the impact of these incomplete graphs is much more prominent. These values appear to increase linearly with respect to team size; however, they are expected to smooth out, in a similar fashion to thresholds of three and four, as the number of incomplete graphs decrease.

When reviewing team size against network density, Figure 12(b), it is observed that there is no discernible effect of team size on network density. As the team size increases, the network density across constant threshold lines remains relatively constant. It is observed that network density decreases significantly with increased threshold level. This is due to the increased potential for communication links when the threshold value is low.

Based on these results, it is preferable to support each design team with additional lines of communication to allow for greater sharing of design activities, especially for teams of greater size. When considering crowdsourced design, this could come in the form of using members with similar traits and complementary abilities. With the traits of individuals being used to generate connections between them, crowdsourced networks would benefit from the combination of individuals who share common interests to support collaboration efforts. Thus, it is advantageous to develop additional modes of communication between members, potentially through increased content sharing or trait matching.

While this work allows for the initial analysis of simulated design teams, there exist further possibilities to extend this simulation framework to generate a more robust and adaptive model, allowing for the theoretical performance validation of the framework. Currently, individual abilities are restricted to broad estimations of their overall competencies. Further understanding of the specific attributes of each individual and how these map to design improvements is required before self-organizing mass collaboration efforts can expand.

Additionally, it is important to note that increased connections between individuals are not penalized, supporting the trend of increased communication between members. In practice, additional communication links can increase time of development and costs. This work reviews a design initiative under constant

iteration, similar to open source projects, such that development time is assumed to not significantly impact the design score.

5. Conclusions

This work presents a conceptual simulation that quantifies the design ability of ad hoc design teams generated from a crowd while reviewing the network structure of each team. Predicted design improvements are estimated after the initial characteristics of each individual have been quantified, based on the composition and network structure of each team.

Overall, it is determined that increased connectivity through information flows and member positioning allows for greater design ability. While this result is intuitive, quantifiable metrics allow for stronger network development guidance when considering large-scale mass collaboration projects. It is also found that as the team size is increased, greater emphasis must be placed on open information flow within the network.

An increased grasp of network development also requires the inclusion of additional team formation methods such as *recommender systems*, to allow for individuals to join efforts to which they are well suited, *learning algorithms*, to properly map competencies to design improvements, and *agent based modeling*, to observe time-dependent effects on network formations. With these additional insights, self-governing design networks could potentially support complex development projects.

References

- Aitamurto, T. & Landemore, H. E. 2015 Five design principles for crowdsourced policymaking: assessing the case of crowdsourced off-road traffic law in Finland. *Journal of Social Media for Organizations* 2 (1), 1–19.
- Ball, Z. & Lewis, K. 2017 The design of the crowd: organizing mass collaboration efforts. *ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, DETC2017–68127*, American Society of Mechanical Engineer.
- Barabási, A. L. 2002 *Linked: The New Science of Networks*, Perseus Books Group.
- Barabási, A. L. & Albert, R. 1999 Emergence of scaling in random networks. *Science* 286 (5439), 509–512.
- Blondel, V. D., Guillaume, J. L., Lambiotte, R. & Lefebvre, E. 2008 Fast unfolding of communities in large networks. *Journal of Statistical Mechanics* 2008 (10), P10008.
- Borgatti, S. P. 2005 Centrality and network flow. *Social Networks* 27 (1), 55–71.
- Borgatti, S. P. & Everett, M. G. 2006 A graph-theoretic perspective on centrality. *Social Networks* 28 (4), 466–484.
- Borgatti, S. P. & Foster, P. C. 2003 The network paradigm in organizational research: a review and typology. *Journal of Management* 29 (6), 991–1013.
- Brabham, D. C. 2008 Crowdsourcing as a model for problem solving: an introduction and cases. *Convergence* 14 (1), 75–90.
- Brabham, D. C. 2009 Crowdsourcing the public participation process for planning projects. *Planning Theory* 8 (3), 242–262.
- Brabham, D. C. 2010 Moving the crowd at threadless: motivations for participation in a crowdsourcing application. *Information, Communication & Society* 13 (8), 1122–1145.

- Brabham, D. C.** 2013 *Using Crowdsourcing in Government*, pp. 1–42. IBM Center for the Business of Government.
- Burnap, A., Ren, Y., Gerth, R., Papazoglou, G., Gonzalez, R. & Papalambros, P. Y.** 2015 When crowdsourcing fails: a study of expertise on crowdsourced design evaluation. *Journal of Mechanical Design* **137** (3), P031101.
- Certoma, C., Corsini, F. & Rizzi, F.** 2015 Crowdsourcing urban sustainability data, people and technologies in participatory governance. *Futures* **74**, 93–106.
- Change.org** [Online]. Available: <http://www.change.org>. [Accessed: 15-Feb-2017].
- Chiu, C. M., Liang, T. P. & Turban, E.** 2014 What can crowdsourcing do for decision support? *Decision Support Systems* **65**, 40–49.
- Contractor, N. S., Wasserman, S. & Faust, K.** 2006 Testing multitheoretical, multilevel hypotheses about organizational networks: an analytic framework and empirical example. *Academy of Management Review* **31** (3), 681–703.
- Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., Leaver-Fay, A., Baker, D., Popovi, Z. & FoldIt Players** 2010 Predicting protein structures with a multiplayer online game. *Nature* **466** (7307), 756–760.
- Cormier, P., Litterman, B. & Lewis, K.** 2011 Empirically derived heuristics to assist designers with satisfying consumer variation in product design. In *ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, DETC2011–48448*, American Society of Mechanical Engineers.
- Craney, T. A. & Surles, J. G.** 2002 Model-dependent variance inflation factor cutoff values. *Quality Engineering* **14** (3), 391–403.
- Deijfen, M. & Kets, W.** 2009 Random intersection graphs with tunable degree distribution and clustering. *Probability in the Engineering and Informational Sciences* **23** (4), 661–674.
- Dorn, C. & Dustdar, S.** 2010 *Composing near-Optimal Expert Teams: A Trade-off between Skills and Connectivity, On the Move to Meaningful Internet Systems: OTM 2010*. pp. 472–489. Springer.
- Erdős, P. & Rényi, A.** 1960 On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* **5**, 17–60.
- Esser, J. K.** 1998 Alive and well after 25 years: a review of groupthink research. *Organizational Behavior and Human Decision Processes* **73** (2), 116–141.
- Farasat, A. & Nikolaev, A. G.** 2016 Social structure optimization in team formation. *Computers and Operation Research* **74**, 127–142.
- Feng, B., Jiang, Z.-Z., Fan, Z.-P. & Fu, N.** 2010 A method for member selection of cross-functional teams using the individual and collaborative performances. *European Journal of Operational Research* **203** (3), 652–661.
- Fitzpatrick, E. L. & Askin, R. G.** 2005 Forming effective worker teams with multi-functional skill requirements. *Computers & Industrial Engineering* **48** (3), 593–608.
- Franzoni, C. & Sauermann, H.** 2014 Crowd science: the organization of scientific research in open collaborative projects. *Research Policy* **43** (1), 1–20.
- Fuge, M. & Agogino, A.** 2014 How online design communities evolve over time: the birth and growth of OpenIDEO. In *ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, DETC2014–35243*, American Society of Mechanical Engineers.

- Fuge, M., Tee, K., Agogino, A. & Maton, N.** 2014 Analysis of collaborative design networks: a case study of OpenIDEO. *Journal of Computing and Information Science in Engineering* **14** (2), P021009.
- Hahn, J., Moon, J. Y. & Zhang, C.** 2008 Emergence of new project teams from open source software developer networks: impact of prior collaboration ties. *Information Systems Research* **19** (3), 369–391.
- Hamel, G.** 2011 First let's fire all the managers. *Harvard Business Review* **89** (12), 48–60.
- Haque, B., Pawar, K. S. & Barson, R. J.** 2000 Analysing organisational issues in concurrent new product development. *International Journal of Production Economics* **67** (2), 169–182.
- Howe, J.** 2006 The rise of crowdsourcing. *Wired Magazine* **14** (6), 1–4.
- InnoCentive** [Online]. Available: <http://www.innocentive.com>. [Accessed: 15-Feb-2017].
- Karonski, M., Scheinerman, E. & Singer-Choen, K. B.** 1999 On random intersection graphs: the subgraph problem, combinatorics. *Probability and Computing* **8** (1–2), 131–159.
- Kramer, J., Agogino, A. M. & Roschuni, C.** 2016 Characterizing competencies for human-centered design. In *ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, DETC2016-60085*, American Society of Mechanical Engineers.
- Lappas, T., Liu, K. & Terzi, E.** 2009 Finding a team of experts in social networks. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 467–476. Association for Computing Machinery.
- Mcdaniel, M. A., Schmidt, F. L. & Hunter, J. E.** 1988 Job experience correlates of job performance. *Journal of Applied Psychology* **73** (2), 327–330.
- McPherson, M., Smith-Lovin, L. & Cook, J. M.** 2001 Birds of a feather: homophily in social networks. *Annual Review of Sociology* **27** (1), 415–444.
- NASA Open Innovation Service (NOIS) Contracts** [Online]. Available: <https://www.nasa.gov/open/plan/open-innovation.html>. [Accessed: 15-Feb-2017].
- OpenIdeo** [Online]. Available: <http://www.openideo.org>. [Accessed: 15-Feb-2017].
- Panchal, J. H.** 2009 Agent-based modeling of mass-collaborative product development processes. *Journal of Computing and Information Science in Engineering* **9** (3), P031007.
- Panchal, J. H.** 2015 Using crowds in engineering design towards a holistic framework. In *2015 International Conference on Engineering Design*, pp. 27–30. Design Society.
- Panchal, J. H. & Fathianathan, M.** 2008 Product realization in the age of mass collaboration. In *ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, DETC2008-49865*, American Society of Mechanical Engineers.
- Pedersen, K., Emblemsvag, J., Bailey, R., Allen, J. K. & Mistree, F.** 2000 Validating design methods & research. In *ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, DETC2000/DTM-14579*, American Society of Mechanical Engineers.
- Ren, Y., Bayrak, A. E. & Papalambros, P. Y.** 2016 Ecoracer: game-based optimal electric vehicle design and driver control using human players. *Journal of Mechanical Design* **138** (6), P061407.
- Schenk, E. & Guittard, C.** 2009. *Crowdsourcing: What Can be Outsourced to the Crowd, and Why*, Workshop on Open Source Innovation, Strasbourg, France. p. 72.
- Shirky, C.** 2008 *Here Comes Everybody: The Power of Organizing without Organizations*. Penguin.

- Surowiecki, K.** 2005. *The Wisdom of Crowds*. Anchor.
- Takai, S.** 2010 A game-theoretic model of collaboration in engineering design. *Journal of Mechanical Design* **132** (5), P051005.
- Tuarob, S. & Tucker, C. S.** 2015 Automated discovery of lead users and latent product features by mining large scale social media networks. *Journal of Mechanical Design* **137** (7), P071402.
- Tapscott, D. & Williams, A. D.** 2008 *Wikinomics: How Mass Collaboration Changes Everything*. Penguin.
- Wang, F. & Zaniolo, C.** 2015 Employee Temporal Data Set, <http://timecenter.cs.aau.dk/Data/employeeTemporalDataSet.zip> [Accessed: 1 Dec 2017].
- Wang, M., Chen, W., Huang, Y., Contractor, N. S. & Fu, Y.** 2015 A multidimensional network approach for modeling customer–product relations in engineering design. In *ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, DETC2015–46764*, American Society of Mechanical Engineers.
- Wasserman, S. & Faust, K.** 1994 *Social Network Analysis: Methods and Applications*. Cambridge University Press.
- Wi, H., Oh, S., Mun, J. & Jung, M.** 2009 A team formation model based on knowledge and collaboration. *Expert Systems with Applications* **36** (5), 9121–9134.
- Wilson, R. J.** 1996 *Introduction to Graph Theory*. Longman Group Ltd.
- Wu, D., Rosen, D. W., Panchal, J. H. & Schaefer, D.** 2016 Understanding communication and collaboration in social product development through social network analysis. *Journal of Computing and Information Science in Engineering* **16** (1), P011001.
- Zhu, M., Huang, Y. & Contractor, N. S.** 2013 Motivations for self-assembling into project teams. *Social Networks* **35** (2), 251–264.