Institute of Information Management

University of St.Gallen

Working Paper:

Data Mining and Social Network Analysis of Ideation Contests: A Repeated Measures Design

Authors: Thomas P. Walter
Andrea Back

Chair: Prof. Dr. Andrea Back

Date: 15.10.2013

University of St.Gallen –
School of Management, Economics, Law, Social Sciences and International Affairs (HSG)

Institute of Information Management
Mueller-Friedberg-Strasse 8
CH-9000 St.Gallen
Phone: +41 71 224 2545
Fax: +41 71 224 2716

1 Introduction

"Innovation is not the product of logical thought, although the result is tied to logical structure." Albert Einstein once said. Given the right conditions, ideation contests, which are web-based competitions of users who use their skills, experiences and creativity to provide a solution for a particular contest challenge defined by an organizer, can achieve outstanding results. For instance, recent solutions derived from ideation contests at InnoCentive, an ideation contest organizer that focusses on a broad range of contest domains such as engineering, computer science, math, chemistry, life sciences, physical sciences and business, include a low-cost rainwater storage system which now is in use across Africa, an off grid illumination device that now is used in areas of developing countries that are not connected to an electricity network or the concept of a school system which is based on cellular phone technology that distributes educational material by causing no-costs [1]. In a single company case study, market analyst Forrester Research illustrates the financial impact of using InnoCentive ideation contests in the research and development department of a large consumer products organization. The report arrives at the conclusion that ideation contests at InnoCentive achieve a return on investment (ROI) of 74%, with a payback period of less than three months [2].

Notwithstanding these success stories, it is not possible to plan innovation provoking ideation contests from scratch. Foremost, the initially mentioned logical structure of ideation contests itself, that is to say the incremental procedure which leads towards such impressive solutions (the so called ideation function), has not yet been fully unraveled. While we know lots of anecdotic examples where ideation contests have led to remarkable results, about the benefits of certain technical features supporting the ideation process or about the relevance of various influencing factors like monetary rewards or feedback mechanisms, yet little is known about the incremental process steps within ideation contests. Research still struggles with questions like "How many ideas does it require to achieve high ideation quality?", "Does time play a significant role when it comes to ideation quality?" or "How does idea aggregation work?". As a result, detailed findings on the processes within ideation contests are rare. Foremost, ideation contests are threatened as a black box, which transform some adjustable independent variables into a plurality of highly valuable ideas. Figure 1 illustrates this research gap.

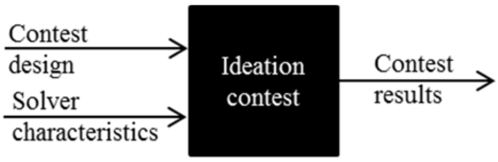

Figure 1 Analogy of an ideation contest as black box

Since [3] introduced his theory of brainstorming in 1957, researchers have been focusing on such input-output correlation. Likewise, basic tenets of ideation remained stable, albeit Information Systems (IS) like Electronic Brainstorming Systems (EBS) [4] or recent Crowdsourcing Platforms have rolled up the entire process of ideation. Today a variety of modern online ideation contests frequently provide a variety of corporate tasks to the anonymous crowd of internet users. The common understanding follows [5], who stated that as long as certain amounts of everyday people elaborate on these tasks, their aggregated results can excel the results a single expert achieves. MIT's Center for Collective Intelligence defines such situation as the appearance of collective intelligence (CI): *Groups of individuals doing things collectively (connected by computers) that seem intelligent* [6], [7]. Next to the laws of large numbers, frequently cited requirements of CI (or a wise crowd) are diversity, decentrality and independence in opinion and a subsequent process of aggregation [8–10]. But despite theory references the aggregation mechanism as highly relevant (for ideation quality or a CI), the aggregation within ideation (the ideation function itself) is a key question largely left behind by IS-research. Hence, this study's key research question is if and how the ideation function can technically be extracted from an ideation contest and if so, what we can learn on the aggregation process within the contests. The guidance of this question leads us to phrasing two proper research questions. In consideration of the appropriate procedural approach to extract and analyze the ideation function of ideation contests we ask:

> *RQ 1: How can the ideation function technically be measured and analyzed so that we can draw implications on the quality of ideation?*

As RQ 1 has a technical focus, our second research question is stated with regards to the research field of crowdsourcing, ideation contests and CI. Therefore we ask:

> *RQ 2: Which cognitive abilities of a solver crowd can be detected within ideation contests and how are they visible in intermediate and final results of an ideation contest?*

Thus, the second question aims on converting techniques from RQ1 into practice. Subsequent questions are how the final result of an ideation contests emerge from the crowd's submissions, whether ideas or concepts are built up on each other and to what extent winning ideas are a proof of an outstanding creativity or a result of CI? To elaborate these questions our survey will proceed as following. Chapter 2 will provide the theoretical background. We conduct a literature analysis on ideation theory, empirical studies on crowdsourcing and ideation contests and the methods of measurements applied during those studies. Chapter 3 focusses RQ 1 and develops a procedure model to measure and analyze the ideation function in online ideation contest. We suggest a repeated measures design, including data mining techniques and social network analysis to describe the ideation function. Chapter 4 puts the procedure model into practice. Applying repeated measures after every idea, we exploit an exemplary ideation contests called "*The motorbike of the future*", including 725 idea submissions. We discuss our findings and aim on drawing conclusions from this survey in chapter 5.

2 Theoretical Background

Ideation is defined as *the process of generating or conceiving of ideas and concepts that may be useful for attaining some desired state or outcome* [11]. Research has come up with a variety of techniques designed to increase the number or quality of ideas produced during ideation. Next to brainstorming [3], the Delphi method [12], or the five W's, ideation contests are nothing else than yet another technique [13]. Nevertheless, the ultimate purpose of every ideation techniques is to produce good, or outstanding ideas [14]. For the theoretical background of our survey, we intend to unravel the state of the art of the question how ideation contests produce good ideas. The following three questions serve us as guidance for the following chapters: *How does ideation work theoretically?*, *How do ideation contests work empirically?* and *Which research methods are applied to analyze ideation contests?* The first question is used to understand ideation as processes from a theoretical perspective. The second question hence is, for which of those theoretical assumptions empirical evidence has been found and the third question focusses the methods and techniques applied during those empirical studies. Herein, the aim is to detect the current usage of data mining and SNA techniques within the literature.

2.1 The underlying theory of ideation contests

[3] proposed an ideation protocol, coined brainstorming, for improving ideation. Following his theory, ideation should foremost aim for high quantities and defer judgment. The theory does not instruct clear process steps but rather provides guidelines such as to form groups of twelve people, to address one specific question at a time, to welcome unusual ideas or to combine and improve ideas. He suggests that ideas with better quality would be generated when people were hold back from criticizing one another's ideas, were open to wild or unusual ideas, focused on generating a large quantity of ideas, and sought to build and expand on the ideas of others [15]. Remarkably, [3] suggests that the first ideas that are mentioned are unlikely to be among best ideas. Subsequent, he suggests to focus on the second half of the ideation process, as more good ideas would be mentioned there. This theory indirectly induces that, all else being equal, the more ideas submitted, the more likely it is that good ideas are included. Until this very day, this argumentation is used as foundation of most studies on ideation contests. However, critics of Osborne's brainstorming approach argue that evaluation apprehension, production blocking, social matching and freeriding may obstruct high quality ideation [16]. Modern information systems (IS), mainly electronic brainstorming systems (EBS, [17]) provided opportunities to leapfrog those pitfalls by facilitating the ideation process electronically.

Ideation contests are a modern form of electronic brainstorming. Following the ideas of [3], [4], [17], [18], in theory ideation contests lead to a plurality of good ideas. In terms of ideation contests, good ideas (high ideation quality) are ideas that contain novel information, that are feasible to implement, that would attain the goal, and that would not create new unacceptable conditions [15], [19–21]. Furthermore, theory often suggests, that in best case, ideation contest can invoke a CI [6], [9], [22], [23], a so called *wise crowd* [8]. But similar to ideation, the theory does not imply a strictly required process instruction, but rather provides guidelines for successful implementation. [8] sums up diversity, decentrality and independence in opinions as well as the existence of an aggregating mechanism as requirements. However, he does not provide specific information how aggregation does work. [9] distinguishes between outreach, additive aggregation and self-organization to impose CI. Again, outreach and additive aggregation are seen as a required process steps. The value of outreach is seen in a larger number of opinions and additive aggregation is required to get the optimum out of such large quantities. The additive process suggests collecting a large number, but also variety of opinions and building the mean. However, [24] argues by using the law of large numbers and estimation games, but neglecting ideation.

Regarding the aggregation process within ideation, contrary conjectures on the ideation function are made by theories of groupthink [25], [26], the tipping point [27] and Bounded Ideation Theory [15]. Groupthink occurs, if the crowd's desire for harmony overrides a realistic appraisal of alternatives. Participants then try to minimize conflicts and reach a consensus decision without critical evaluation of alternative ideas or viewpoints [28], [29]. In the end groupthink can lead to the loss of creativity, uniqueness, and independent thinking, which in turn are required criteria for collective intelligence. Possible causes for groupthink are defined in high group cohesiveness, structural faults like the lack of norm-requiring methodological procedures and the situational context like recent failures or moral dilemmas. Whether groupthink occurs in a situation is largely a subjective perception. However, groupthink theory does not define how the aggregation of opinions can leads to groupthink in a step-by-step manner. Such assumption is taken by the theory of a tipping point [27]. The tipping point states that ideas, products, messages and behaviors spread just like viruses do. Hence, the tipping point reconditions the idea of network effects and critical masses [30] in ideation. The assumption is that a former linear or steady process is swapped by a marginal idea, an idea that has major impacts, a signal that stands out from the noise and hence, changes the direction of the entire ideation process. However, applied to ideation contest, the theory does not conjecture whether the effect should be of positive or negative nature. A positive effect could be explained by an eye-opening, game-changing idea that enters a new field and inspires others to be more creative. A negative effect could be explained by the situation of one idea representing the marginal idea. After the submission of this marginal idea, all follow up ideas do not add value to the ideation contest, neither in novelty nor feasibility. To compare those theoretical assumptions we make use of their assumptions on ideation functions, which is the relationship between the total ideation values (quality of ideation) produced during an ideation session and the total number of ideas (a time factor) contributed. Figure 1 sorts theories by their time of publication and illustrates their ideation function.

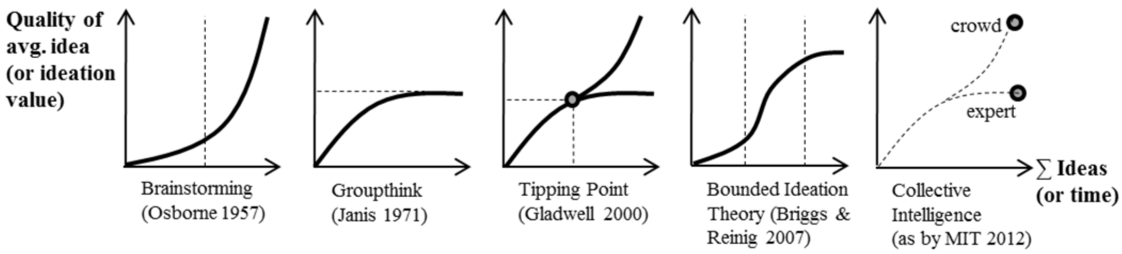

Figure 2 Ideation functions implied by different theories

[15] sums up [3], [4], [18] and [16], [25] and presents the arguments of a so called Bounded Ideation Theory, a new causal model of the ideation function. Their curve argues, that the ratio of good ideas to total ideas may be smaller early within the ideation process on due to limited understanding of the task, and then larger as understanding of the task increases, and then smaller again due to cognitive overload and physical exhaustion.

2.2 Empirical studies on ideation contests

Even theory remains inconclusive, the analysis of ideation is not merely a recent trend, but has long tradition, especially in IS-research. IS-researchers discuss the question of computer supported ideation processes since the early days of the web, from EBS [31], creativity software [32], over the development of wiki software [33], [34] until web-based ideation platforms [35], [36] or Group Wisdom Support Systems [37] . Moreover, with an increasing amount of web-platforms similar to InnoCentive (e.g. see NineSigma or IdeaConnection) and organizations which follow open innovation approaches [38] and start crowdsourcing idea generation processes on their own (repeatedly cited examples include MyStarbucksIdea, Dell IdeaStorm or IBM's jams), also the body of IS-literature dealing with the topic is growing. To get an overview see [39] or [40] for crowdsourcing taxonomies, [41–43] for literature reviews or [20] for a meta-analysis of 90 studies that deal with ideation quality.

A number of recent empirical studies (for example [44–49]), which aim to detect success factors of ideation contests, have revealed that design patterns (rewards, feedback and rating mechanisms, task description and many more) have significant impact on the outcomes of ideation contests. Such factors can be seen as initial "settings" to an ideation contest. They all can be set and changed by an organizer prior to the start of a contest. A second set of independent variables frequently used in empirical studies (for example see [6], [50–53]) is given by variables that describe the characteristic of the solver crowd (age, gender, profession, origin, income, hobbies, educational level, attitudes and believes). Such factors are of demographic nature and describe what kind of crowd is involved in the ideation contest. The third set of independent variables deepens the description of the crowd by additional structural variables. For instance, researchers use the number of contacts a solver keeps within the solver network [53], the past experience of solvers within online communities [54], the attitudes of solvers to a seekers brand [48], attitudes towards sharing and collaborating within a group of solvers [21], [55], [56] or the structural position within the solver network [57], [58].

When it comes to measuring the impacts of these factors, it's merely their direct impact on the outcome, foremost the quantity and quality of ideas. In most studies the output, and in consequence the success, of ideation contests is measured indirectly. For instance, [45] use the amount of attracted solvers as indirect measurement of quality, [48] use the average rating of every idea on a five-star scale within a solver network, [59] measure quality by the boolean information whether a task was completed and [54] uses the information whether a submission was eventually implemented as primary dependent measure of quality. Also qualitative approaches can be found. For instance, [57] use data from external experts to measure quality and [60] take the evaluation from independent executives to compare crowd and expert submissions. As mentioned, only a few studies point to the characteristics of the ideation process when defining neither independent nor dependent variables. For instance, [61] show a temporal strategic analysis of solvers based on their decision when to enter a contest and when to submit an idea, [36] define a mixture of interactive methods to be applied during an ideation process. To the best of our knowledge, the closest ideation function based measurements is presented by [62], who develops a practical, web-based asynchronous ideation contest, which allows the implementation and test of various incentive schemes. In this survey the amount of ideas that refer to an initial idea as their root of inspiration are taken as quality measurement of this initial idea.

2.3 Research methods to analyze ideation contests

As the previous chapter shows, a tremendous volume of literature on ideation contests has been accumulated in IS-research. At the same time, a brief but specific literature review within this literature shows an unequal distribution of applied research methodologies. Ideation describing theory mostly are results of experiments in laboratory-like environments [3], [26], or are based on few remarkable incidents that are documented within exploratory single case studies [8], [27]. Applying those theories, explanatory research also runs experiments [18], [59], [63] or deepens insights by interviewing solvers [57], [64].

With an increasing amount of crowdsourcing platforms, a noteworthy amount of case study research approaches are realized. The goal is to describe the phenomena, as well as to develop ideation design artifacts such as taxonomies, models or patterns [6], [47], [49], [65–72]. Such research is often enhanced by descriptive statistics [36], [40], [51], [57], [73–75]. Evaluating these design artifacts and analyzing aforementioned independent variables, a large set of empirical studies are conducted. Using actual user data (taken from platform data) or indirect user data (from surveys) literature applies regression models such as OLS, PLS, ANOVA to analyze the effects of the aforementioned independent variables on the outcome of ideation contests [14], [33], [44–46], [48], [53], [55], [76], [77]. In a similar vein, research exploits logit models [50], [54], [78], cluster analysis [79] or various group comparison techniques like Cronbach's alpha, U-test or t-test [52], [58], [80]. Additionally we find game theoretical and other econometrical approaches to develop the structure of an ideation contest [35], [55], [62], [81], [82].

To this point, data mining as well as SNA techniques are remarkably neglected research methods. A brief, but precise literature review verifies this proposition. Therefore we scan The Association of Information Systems electronic Library, AISeL using “data mining”, “text mining”, “clustering”, and “social network analysis” as search terms to define our intended methodology and “ideation”, “idea contest”, “crowdsourcing” and “collective intelligence” to describe our primary field of interests. Thus, the simple question is how those methods are represented in the field? We pairwise combine all search terms by a logical AND and intend to find a combination of those terms in titles or abstracts of papers. The significant finding depicted in Table 1 is, that the 32 queries (title and abstracts) leads to only three hits. There are hardly any studies available that attempt to make a contribution to the field by using data mining or SNA as a research method, even though these methods are well-rehearsed and reputable within IS-research. Expanding the literature search by a forward and backward search on the findings reveals scattered coverage of the methods within the field.

Table 1 Results from a literature search for applied research methods within the field of ideation

Search Term Combination	Ideation	Idea contests	Crowdsourcing	Collective intelligence
Data mining	0	0	0	0
Text mining	0	0	[83]	[84]
Clustering	[85]	0	0	0
Social network analysis	0	0	0	0

For example, [86] apply text mining to depict crime networks, [87] apply four commonly used text classification algorithms and propose a text classification framework for finding helpful user- generated contents in online knowledge-sharing communities, [88] proposes a software tool that uses the concepts of swarm intelligence and text mining to analyze free/open source software development communities and [84] run text mining methodology on user opinions expressed via twitter to analyze the appearance of a collective intelligence. In an 2006 paper [85] suggest to use (but don’t apply) hierarchical clustering and multidimensional scaling (MDS) techniques in the design of group support systems and in an 2013 paper [83] apply text mining to select best ideas from crowdsourcing campaigns semi-automatically. Table 2 summarizes the provided background literature and therewith represents the research gap we address with our survey.

Table 2 Summary of background literature by representative findings and identified research gap

Research methodology	Focus on input/ output relations in ideation	Focus on ideation function (process)
Conceptual studies (exploring the phenomena by case studies, artifact design, taxonomies, and literature reviews)	[6], [20], [47], [49], [66], [69], [72], [89]11	[40]
Empirical studies (analyzing the impact of ideation contest design features and solver attributes by surveys among solvers and ideation contest data.)	[14], [44], [46], [48], [50], [52], [54], [55], [62], [76]	[33], [70]
Structural studies (analyzing ideation by applying data mining, text mining, clustering or social network analysis)	[53], [57], [58], [83–85]	**Research Gap**

This gap can be split into two parts, namely the research objective and the research methodology. First, we have shown that research is rather focused on analyzing the effects of input factors on the outcome of ideation contests. The ideation function itself, that is to say the process how ideas or concepts are aggregated and enhanced, represents a somewhat mistreated research objective. Second, we detect a unilateral commitment of research methodologies, foremost the use of descriptive and multivariate statistics. Structural research methodologies, such as text or data mining techniques or SNA are seldom used to analyze ideation contests. To the best of our knowledge, our study is the first that combines those two gaps.

3 Development of a procedure model to analyze ideation contests

This study exploits a procedure model to analyze the ideation function within ideation contests. Aiming on answering our first research question, we consider method engineering literature [90] to build a procedure model as systematic approach to analyze ideation functions. Therefore, we will explain all steps of the procedure model including relevant activities, techniques, as well as suggested tools and the expected outcome documents. In that way we intend the procedure model to be generic and repeatable within the domain of online ideation contests. During the development we adopt the blueprint of a standardized data mining and text mining procedure as described by [91–93]. To be exact, we will follow their common three step approach of pre-processing, processing or actual data mining and the final visualization. The following chapters explain those steps in detail.

3.1 Preprocessing

Once an ideation contests is chosen for analysis, all raw text data (the so called corpus) should be extracted from the website's database using a query language such as MySQL. The central unit of analysis is "ideas". To be able to analyze the ideation function, all extracted ideas must include attributes such as timestamp, rating, allotted reward or corresponding solver. The corpus then is imported to software for qualitative data analysis (QDA) together with its describing attributes. As central element of preprocessing, we suggest a manual coding of categories or concepts which are transported by the idea. Recoding existing data is widely seen as valuable within qualitative data analysis. An observed key mechanism in idea generation for product development is the association of one category of idea with another category [80]. For example, within our ideation contest example (c.f. the next chapter) the product "motorcycle" could be coded with a desired benefit category (eventually "fuel saving") and a product feature (eventually "a lightweight frame") to satisfy that benefit. Hence, we suggest that multiple codes per idea are possible. In that manner, also the combination [3], aggregation [8] or additive outreach [24] of separately mentioned concept categories is covered.

While coding ideas, existing taxonomies (for example, a taxonomy of motorbikes could list all parts of the bike) or a manual coding can be used to define a codebook. In any case, literature suggest to test the reliability of the coding by computing Krippendorff's alpha [94]. Therefore, at least two coders independently have to assign codes to ideas using a commonly provided codebook. Hereafter, inter-coders agreement is used to compare the consistency of coding between coders and can be useful to uncover differences in interpretation, clarify equivocal rules, identify ambiguity in the text, and ultimately quantify the final level of agreement obtained by coders. Acceptable Krippendorff alphas can be defined according to the importance of the conclusions to be drawn from the survey. [95] suggests to rely on data that achieves alpha-values of ≥ 0.8, consider data with $0.8 > \alpha \geq 0.667$ only to draw tentative conclusions, and discard data whose agreement measures $\alpha < 0.667$.

As final result of preprocessing, a code document matrix (CDM) can be computed. In a similar manner to the term document matrix (TDM) in text mining [93] the CDM is of *n by m* type, where *n* represents rows of ideas and *m* represents columns of allotted codes, that is to say concept categories. Hence, contrary to a TDM (which can be computed binary, by term frequency or inverse term frequency) the CDM is binary by definition, with "1" representing the fact that a code exists in an idea and "0" representing that it does not.

3.2 Ideation Process Mining: A repeated measures design

After pre-processing is conducted, a set of data mining techniques represent the core of our procedure model. The basic idea is to apply repeated measurements of the distance between concept categories (or ideas that are coded with these concepts) to be able to explore the changes of their relationships over time. The repeated measures design is a key constraint to our approach. Repeated measures design means to use the same research object (which are coded ideas) with every condition of the analysis [96]. Similar to a longitudinal study, we want to assess the change of concepts used throughout an ideation contest. Figure 3 illustrates our approach.

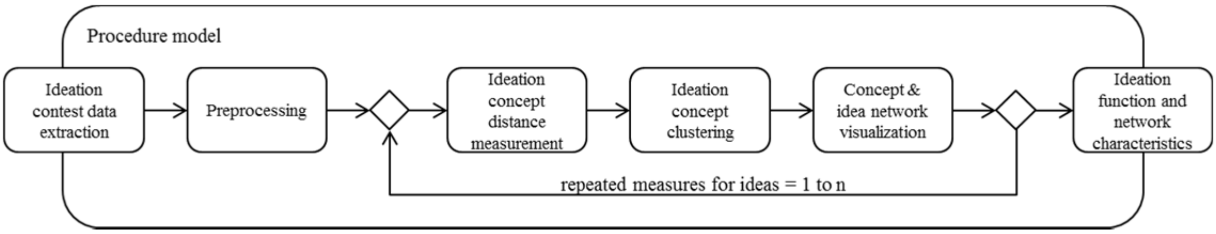

Figure 3 Procedure model

The relationships among concepts as well as idea similarity can be identified using distance measures, clustering, MDS and SNA. But those metrics will change as the process of ideation continue and it is those changes that we find particularly interesting and focus on. Hence, we suggest running all following steps of measurement (distances, clustering, MDS) in a cascading manner from the first to the last idea of the ideation contest.

Still, the procedure model might leave us with a cluttered outcome, especially as soon as the repeated measures design approaches larger N. This is a typical problem in reading the results from MDS or SNA [97], [98]. The final result (the overall situation) then might appear obscure and is of low explanatory character. A common approach in such situations is limiting the amount of total items and running analysis on a defined subset. In our context, subsets can easily be defined by idea's attributes. In that manner we suggest to use subsets of ideas for deeper analysis. In terms of a noised ideation function, that might result when too many concepts are mentioned, we suggest to limit ideas by a preselected amount concept. In terms of the ideation quality, we suggest to limit the analysis to rewarded ideas. Later the results can be compared with the overall contests. Both suggestions will be applied below.

3.2.1 Ideation Distance Measurement

Based on the CDM, distance matrices between concepts can be computed. The matrices are of *n by n* (one-mode) type, which means that concepts represented both, rows and columns and the values represent the distance between them. By default, distance measures allow statements on the co-occurrence of concepts within ideas. Co-occurrences are said to happen every time two concepts appear in the same idea. Hereafter distance matrices between concepts can be computed using Ochiai coefficients. Contrary to the usage of Cosine coefficients to measure distances based on a TDM, the coding of ideas by concepts results in the binary CDM. Hence, we suggest to use Ochiai's coefficient as the binary form of the cosine coefficient, to measure distances between codes: The distance d_{ij} between two codes i and j then is defined by $d_{ij}=\sqrt{\frac{a^2}{(a+b)*(a+c)}}$, where a represents ideas where both codes occur, and b and c represent cases where one code is present but not the other one. For example, two codes occurring together in various ideas could be represented by a variety of ideas taking up on the aforementioned idea of "saving fuel" (say code i) by building "a lightweight frame" (say code j). The similarity of two ideas will range from "0" to "1. This construes values of "1" meaning ideas are using exactly the same set of concepts to "0", usually indicating a total independence of ideas, and in-between values indicating intermediate similarity or dissimilarity of ideas. The more similar two documents are in term of the distribution of concepts, the higher the coefficient between them [91], [93].

3.2.2 Ideation Clustering

Clustering is the application of certain algorithms to automatically detect patterns within the CDM. Therefore, clustering applies the computed distance matrices to explore the similarities between various groups of concept categories. Often so called non-hierarchical (or centroid-based) clustering is applied, foremost the k-means algorithm [92], [99], [100]. We suggest to use k-means to partition the concept categories into *k* cluster in which each observation belongs to the cluster with the nearest mean. At the bottom, k-means is based on principal component analysis or minimalizing least squares [101]. In other words, clustering will minimize the average distance of a group of co-occurring codes to other groups of co-occurring codes. However, determining the exact *k* number of clusters in a data set is a frequent problem in data clustering, and is a distinct issue from the process of actually solving the clustering problem. In text mining, a frequently used method to determine the number of cluster is the elbow criterion [102], which suggests to choose the *k* number of cluster by the maximum R^2, so that neither dropping nor adding a cluster does rise the percentage of variance explained by the cluster [103].

3.2.3 Ideation Cluster Visualization

As a final step of the three-step data mining process [91] we are following, we suggest to read-out the results from clustering. We suggest cluster visualization based on MDS. MDS is often used in information visualization for exploring similarities or dissimilarities in data. Also known as principal coordinates analysis, MDS takes the distance matrices and outputs a coordinate matrix whose configuration minimizes average distances. We suggest using concept maps, as in our case a network of concept categories evolves through the repeated measures. For sufficiently small N, the resulting locations of the coordinate matrix can be displayed in a network-graph. This step can be supported by a plurality of SNA software tools, for example UCInet, Pajek, R or Network Workbench. Within the evolving concept map (network) of co-occurring concepts in ideas, nodes will represent a specific concept and weighted edges the fact how often two concepts co-occurred in ideas. The radius of a node will represent the relative frequency of the concept within the overall network. Cluster of concept are visualized short distances and unique color. For larger N, idea networks might not benefit as much from MDS as some distortion may result. As a consequence, some concepts which are computed as close to each other or which are parts of the same cluster may still be plotted far from each other due to the optimization of mean maximal distances within the entire network. In such cases, the suggested approach of limiting the amount of ideas to a subset and performing MDS on a fewer number of items usually produces a less cluttered map which can be interpreted again.

3.3 Ideation function and ideation network key characteristics

The repeated measures design generates a lot of new data. To be precise, for each idea that is added, we suggest computing the overall situation (network). This approach allows us to define new metrics that are closely related to the ideation function. To plot the ideation process we can look on the sums of concepts, ideas, words, participating solvers, rewards and ratings over time. These metrics already allow a first comparison of a specific ideation contest to the ideation functions implied by different theories (c.f. Figure 2). Furthermore, we suggest adding a growth-perspective, which is defined by the marginal input of every idea. Subsequent questions are whether an idea adds one or more new concepts (new nodes) or focusses on recombining already existing concepts (additional edges within the ideation network). Furthermore marginal ideas can also simply strengthen an edge (adding weight), which over time will lead towards two concepts being part of the same cluster. Additionally to these ideation functions describing characteristics we suggest to include concept network characteristics, especially betweenness centrality and network density, into the analysis.

3.3.1 Betweenness centrality

In [8] a central suggestion to achieve a collective intelligence is keeping loose connections. Referring to the idea that often it does not require very strong connections to spread an idea, this reintroduces [104] theory on the strength of weak ties. In terms of a concept network, a tie (edge) represents the fact that two concepts (nodes) co-occurred in one idea. Hence, following SNA literature, the betweenness centrality of a concept should be the most relevant measure to analyze such characteristic [105]. The betweenness centrality *g(k)* of a node *k* is defined as the share of times that a node *i* needs the node *k* in order to reach a node *j* via the shortest path [106]. Specifically, if g_{ij} is the number of shortest paths from *i* to *j*, and g_{ikj} is the number of these paths that pass through node *k*, then the betweenness centrality of node *k* is given by: $g(k) = \sum_i \sum_j \frac{g_{ikj}}{g_{ik}}$, where $i \neq j \neq k$ [105]. In our context the idea is that a concept which is in-between serves as transmission belt, like a bridge to a yet unseen concept combination. Such network position is defined as powerful, as it implies the ability to broker (or block the combination) of two separately mentioned concepts. Hence, ideas which introduce concepts of high betweenness centrality could be boundary spanning, or covering structural holes in the ideation contest. If later ideas use identic concept combinations, the betweenness centrality of such ideas will tend to be lower due to the existence of multiple shortest paths. Therefore we suggest comparing betweenness centrality of concepts for the subset of rewarded ideas (see above) at the final stage plus during intermediate stages (for example whenever a winning idea is submitted) of the ideation contests.

3.3.2. Concept Network Density

Finally, we suggest two different density measures for the concept categories network. Density describes the ratio of existing edges over possible edges (defined by the number of nodes) [107]. But as edges of the concept network are weighted, the density can also be defined as the average strength of edges across all possible (not all actual) edges [98]. The latter will take the network's cohesion, a growing strength between a distinct subset of nodes, into account, whereas the first measure uses the binary form of a connection between two nodes. We suggest to compute the network density D_t at time *t* using $D_t = \frac{2E}{N(N-1)}$, where *E* is the sum of weighted existing edges and *N* is the amount of Nodes of a network [107]. To sum up and close the development of our procedure model, Table 3 reassembles all suggested steps, including techniques, tools and pursued results.

Table 3 Procedure model steps to analyze ideation contest with data mining and SNA techniques

Step	Activity	Technique	Tools	Results
1	**Data extraction**	• Download the platform's database-dump. • Use a query language to extract the required information (ideas, timestamp, reward and/ or rating, user, etc.) in tabular format.	Query language like *MySQL*	• Plain ideation contest data (the corpus)
2	**Preprocessing**	• Code all ideas by a concept category. • Check inter-coder reliability by Krippendorff's alpha. • Compute a concept document matrix (CDM).	Software for qualitative data analysis (*Nvivo9, Provalis QDA miner, Netminer, etc.)*	• Coded Ideas • Reliability of coding • CDM
Repeat steps 3 to 5 for idea 1 to N. Repeat steps 3 to 5 for defined sub-sets of ideas.				
3	**Ideation distance measurement**	• Compute distances between concepts by Ochiai's coefficient based on co-occurrence in ideas.	Software for QDA or Statistics (*R, Stata, SPSS, Excel, etc.)*	• Code distance matrix
4	**Ideation clustering**	• Use k-means algorithm for clustering concepts. • Determine the amount of cluster using the elbow criterion (maximizing R^2-values).	Software for QDA or Statistics	• Category cluster
5	**Ideation clustering visualization**	• Use multidimensional scaling (MDS) to layout the concept maps (ideation networks).	Software for SNA (*UCInet, Netdraw, Pajek, NWB, etc.)*	• Concept cluster in network layout (concept maps) or as dendrogram
6	**Ideation function and ideation network characteristics**	• Compute the sums for ideas, words, solvers, concepts, rewards, ratings, edges over time. • Compute the growth rates of ideas, concepts and edges over time. • Compute network measures for ideas and concepts (betweenness centrality of ideas and overall network density measures (binary and weighted) over time. • Compare rewarded and unrewarded ideas.	Software for QDA, Statistics and SNA	• Ideation function • Process oriented ideation contest analysis • Ideation network analysis • Winning ideas in context

4 Results

To evaluate the suggested procedure model, we apply it to analyse a real-world ideation contests. As mentioned, this approach can be seen as single case study approach and therewith is also subject to the limits of case study research [108]. Our goal is not to generalize on ideation theory, but rather to explore how research could benefit from applying methods of data mining and SNA to the field. In the following we first will describe the example dataset and thereafter the result of applying the suggested prodedure model.

4.1 Dataset and Coding

We use data from a real world crowdsourcing website. The website was launched in 2007 and since then, 128 different ideation contests, including nearly 50'000 ideas, 20'000 ratings and 500'000US$ of rewards, have been completed. The website has 7.512 solver accounts. The average age of solvers is 41.8 years. 71.2% of solvers are male and 52.6% keep a university degree. The formal crowdsourcing procedure on the website runs as following: Various ideation contests are announced on the website simultaneously. Solvers can sign-up to the website (create a solver profile) for free, but in order to receive potential rewards they have to provide their bank account. During an ideation contests, solvers can submit various ideas, but also see, comment and rate the ideas of other solvers. The rating mechanism is similar to facebook's "like" button, or google's "+1" button, allowing solvers simply to express that they like another solvers idea, but not to differentiate the extent. However, the rating is solely a solver community feature with no consequence on winning a reward. Seekers pay an initial fee to the platform to get the contest online. Prior to the ideation contest seekers decide on the reward (total amount and structure), the duration of the contest, deliver a contest (task) description, announce their relevant criteria of rewarding (e.g. “most radical solution” or “feasible concepts with business impact”) and already suggest the type of expected answer (e.g. “We would like to receive plain text.”). After an ideation contest is closed, all submitted ideas are rated by a seeker’s expert committee and best ideas are rewarded. Table 3 summarizes the overall website’s and our chosen ideation contest’s descriptive statistics.

Table 4 Descriptive statistics of the dataset

Characteristic	Crowdsourcing website	Ideation contest example
∑ ideation contests	128	1
∑ ideas (avg. per contest, std. dev.)	49’284 (385.0, 89.1)	725
∑ words • Avg. per contest (std. dev.) • Avg. per idea (std. dev.)	2’089’472 • 16’324.3 (6222.1) • 42.4 (17.2)	45’759 • - • 62.94 (13.2)
∑ solvers (avg. per contest, std. dev.)	8’512 (305.3, 181.9)	402
Avg. solver age (std. dev.) in years	38.12 (10.3)	41.8 (7.7)
Solver’s gender (male, female) in %	66.2/ 33.8	71.2/ 28.8
∑ rewards (avg. per contest, std. dev.) in US$	524’800 (4100.0, 522.3)	5000
∑ ratings total • Avg. per contest (std. dev.) • Avg. per idea (std. dev.)	211’921 • 1’655.6 (310.5) • 4.30 (0.82)	2175 • - • 3.0 (0.96)
∑ Contest duration (avg. per contest, std. dev.) in days	11’712 (91.5, 14.2)	104

As mentioned, we chose one particular ideation contests as an example to test our procedure model. This contest was called “*The motorbike of the future*” and its seeker was a global manufacturer of motorbikes. The contests lasted for 104 days, during which 725 ideas were submitted by 402 solvers (1.82 ideas per solver; 6.98 ideas per day). 27 solvers submitted at least five ideas and 84 at least three ideas. The seeker rewarded twelve different ideas, aggregating a total reward of 5’000 US$. The task description of the seeker was rather broad: *“Motorcycle, scooter, moped riders and enthusiasts: We want your ideas! What will be the perfect motor bike of the future – what features should it have, how much would it cost, who would buy it? What will be the next exciting trend in motor bikes that will fascinate the youth of tomorrow? We are looking for unique, breakthrough products and services for the future of motorcycles, scooters, mopeds or the like vehicles. The solutions should embody innovative technology, passion and modern lifestyle.”*

We followed the procedure model (c.f. Figure 3, Table 2) to analyze all ideas that were submitted to this task. While extracting the raw ideation contest information (the ideas in plain text) from the crowdsourcing website’s database using MySQL, we also included various types of arrays of every idea. This included the idea’s timestamp, solver ID, reward and rating. All this information was imported to the *Provalis Research QDA-Miner* software. Within the software, all ideas were coded by responding categories. A first coder coded all ideas by a category following basic rules of coding [109]. The goal of categorizing ideas was to describe how (which part or by which act or process) motorbikes should be innovated. The depth of coding was ceded to the coder. For example, various ideas addressed environmental issues. However, as they suggested various aspects, resulting codes included E-*motor, Hybrid, Fuel-saving, CO_2-emission, Solar power*, h*eat-reusage*, etc. Likewise, ideas that addressed a business issue included codes such as *Vendor Service, Brand- refreshment, Versioning/ Customization, Salesprice, etc.* The first coding resulted in a codebook containing 70 concept categories. This codebook was provided to two other coders, who also coded all ideas, being allowed to only using codes from the codebook. Inter-rater reliability was assessed by calculating Krippendorff’s alpha for all 70 codes. The reliability for most of the codes is above .67, the commonly canonical value. Perfect agreement was achieved for concept categories that described parts or types of a motorbike (for example *Wheels*, *Seats*, *Storage*, *Frame*, *Handlebars*, *Sidecar*, *Quad*, *Trike*, *Jetski* or *Snowmobile* all resulted in values of 1.0). Lowest agreement was achieved in rather technical categories or categories that described the benefit of an action. For example, the agreement values of *fuel cell*, *heat re-usage*, *on-board electronics*, *novel vendor service*, *brand-refreshment* or *visibility* are .62, .66, .66, .65, .66, .67, respectively. Remaining codes received values above 0.67. Given the difficulty of the specific task (and the fact that none of the coders is a motorbike experts), those results seem to be satisfactory [95], [109], [110].

4.2 Repeated measures to analyze concept cluster development

Applying steps 3 to 5 from the procedure model enables us to analyze and visualize how the solver's ideas are built up and relate to each other while creating an ideation network. We apply the repeated measures design by using the ideas' timestamps and calculating a new CDM after every idea that is added during the ideation contest. Code distances are used to cluster ideas, which are visualized by applying MDS to draw concept maps. In other words, following the procedure model we are able to describe the ideation process by using 725 concept maps and the included network characteristics. Figure 4 illustrates four instances, that is to say the concept maps after ten, 20, 30 and 40 of the 725 ideas.

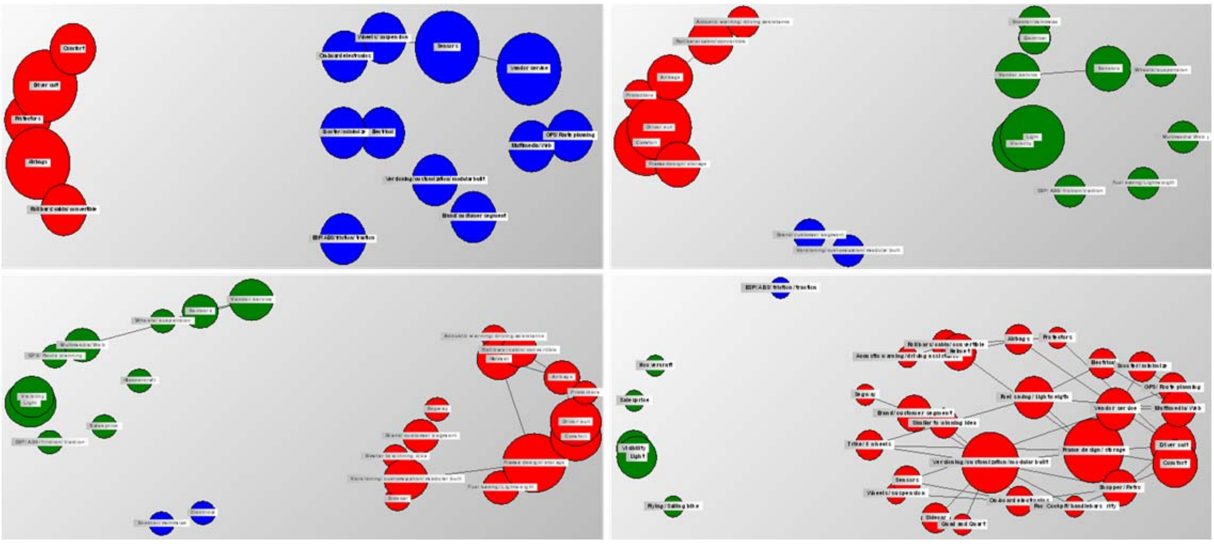

Figure 4 Concept maps based on CDM, clustering and MDS after ten, 20, 30 and 40 of 725 ideas (top left to bottom right).

Stringently, early ideas contain high levels of novelty. To be precise, in terms of words or categories, the very first idea that is submitted to an ideation contest will have endless novelty by definition. In our case the first idea was the following: *"My idea is a networked motorcycle. The bike of the future should include lots of networked features. I think, that based on GPS and your online profiles it should be possible that your bike suggests locations for you to stop, find the way to appointments that you have in Outlook or is recognized by your home whenever you approach your garage by 100m or so. A networked bike would also include a feature that allows you to stream your music from your home computer or a feature that allows your mechanic to get remote access to your bike's system."*

This rather comprehensive idea already mashes up a technical idea (new sensors) with new business offers (remote access by a mechanic, streaming). Appurtenant concepts are *Sensors, On-board-electronics and Vendor service,* and they are represented within the blue cluster, visible in the top left image of Figure 4. As ideas are visible to other solvers, subsequent ideas try to circumvent initial ideas, which leads to other concepts also visible in Figure 4. Consequently, after 10 ideas the following situation emerges: Twelve different concepts are mentioned within these ten ideas. The elbow criterion suggests two cluster (R^2 = 0.8655). Apparent by larger node-radius, four concepts (*Driver Suit, Airbags, Sensors, Vendor Service*) already are mentioned at least twice and as the edges signalize concepts also briefly co-occur in ideas (as within the first one). With further ideas being submitted, nodes are added (due to new categories), node sizes change (due to concept frequencies), edges are added (due to co-occurrence of categories), the weight of edges changes (due to multiple co-occurrence of categories), the amount of cluster changes (due to the elbow criterion) and a concepts membership to a cluster changes (due to the distance measures). Figure 5 illustrates this by opposing the concept networks at halftime (N=362) and finish (N=725) of the ideation contest.

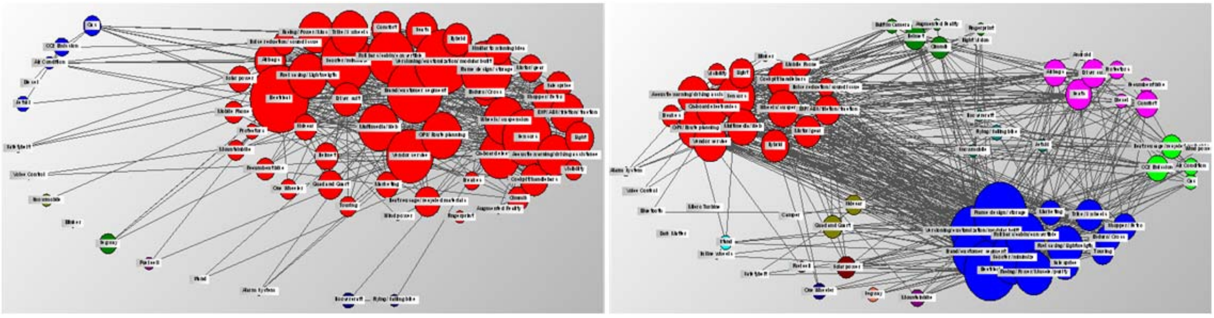

Figure 5 Concept maps of an ideation contests after N=362 (left) and 725 (right) ideas

Figure 5 illustrates that by the nature of ideation, the likelihood of novel concepts decreases over time. Table 4 supports this hypothesis, by presenting equivalent network characteristics. By the end of the first half, the majority of concepts (60 of 70 or 85%) are already submitted. Conversely, the second half of ideas that are submitted to the contests significantly increase the amount and weighting of edges. This is given by the concept network's density of 11.8% (208 edges of 1770 possible edges between 60 concepts) at N=362 against 20.1% (486 of 2415 for 70 concepts) at N=700 ideas. In other words, whereas the first half of ideas is rather focused on brainstorming and rapid mentioning of new concepts, the second half of ideas is rather focused on recombining and aggregating concepts that already have been mentioned. Considering the weighting of this effect becomes even stronger. At N=362, 298 edges are created (making 90 or 30.2% duplicates) while at N=700, 820 edges are created (making 334 of them or 40.7%) duplicates. Thus, not only the creating of new edges accelerates during the second half, but also the weighting. Using a negative expression, this means that during the second half 99.0% of the mentioned concepts are copies, and also nearly every second combination (46.7%) of concepts has already been mentioned.

Table 5 Ideation concept network characteristics, split by two halves of an ideation contest.

Network characteristic	N=1 to N=362	N=363 to N=725	N= 725
∑ Nodes (binary)	60	10	70
∑ Nodes (frequency)	621	954	1575
∑ Node duplicates frequency (%)	561 (90.3)	944 (99.0)	1505 (95.6)
∑ Edges (binary)	208	278	486
∑ Edges (frequency)	298	522	820
∑ Edge duplicates frequency (%)	90 (30.2)	244 (46.7)	334 (40.7)
Network density (%)	11.8	29.5	20.1

In that manner, clustering leads to the emergence of visible, dominant concepts as well as so called single item cluster (a single concept defining a cluster of its own). For example, as by the time of N=362, ideas which included safety concepts are caught within the major cluster (colored red in the top-left network of Figure 5), but increasing co-occurrence of a subset of safety concepts leads to the pink colored cluster the left image. This represents the fact that between N=362 and N=725 ideas, safety concepts like *Airbags, Protectors* and *Safetybelts* remarkably co-occur, but also together with concepts like *Comfort,* new kinds of *Seats* and *Driver Suits.* In other words, as the contests proceeds, solvers adopt the safety issue and increasingly recombine it with an issue of higher comfort. The repeated measures design shows, that ideation is by no means a linear process. It becomes visible, that there is a brainstorming phase in the beginning of the contest and an elaborating- or framing-phase in the second half of the ideation contest. While novelty decreases in terms of new concepts, further ideas aggregate and hence, help to carve out more specific or detailed concepts. Figure 6 shows this by presenting the ideation contests metrics and characteristics as suggested in step 6 of our procedure model.

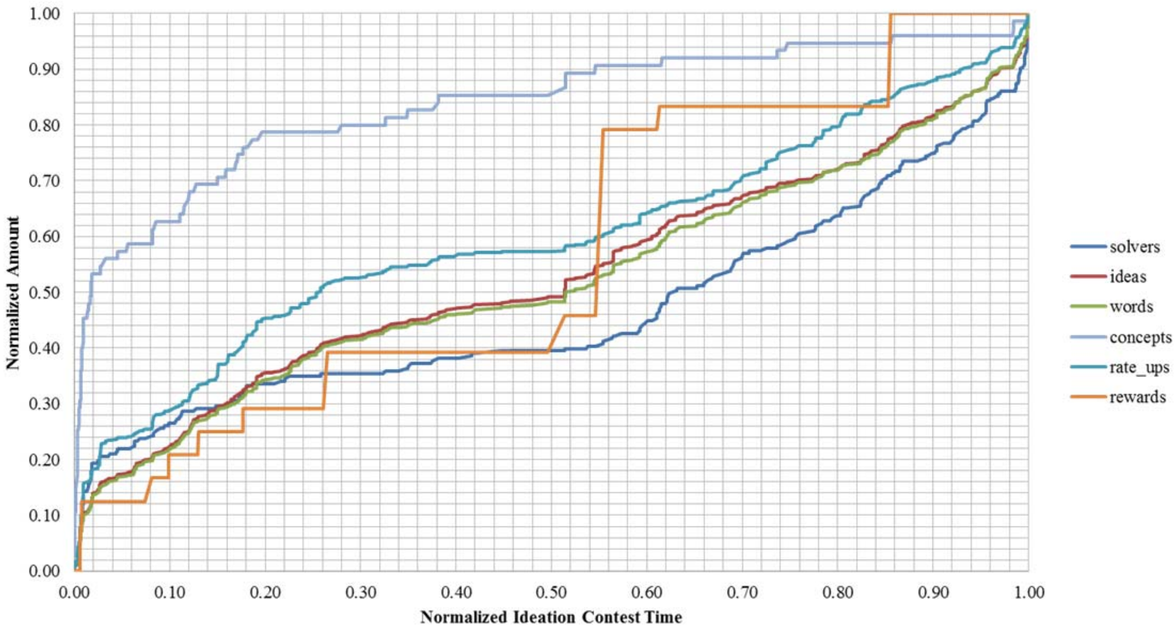

Figure 6 Ideation contest characteristics (ideation function)

During the first 20% of the contest's allotted time already 80% of the concepts are mentioned, but the vast amount of submitted ideas during the remaining 80% of the time do not add new concepts, but just recombine or worse) repeat existing ideas. The ideation function shows some solvers as productive early mover. The first 20% of all participating solvers already mention over 50% of all concepts in the first 18% of ideas and within less than 5% of the contest time. By half time close to 90% of the concepts are mentioned, but yet only 35% of the contest participating solvers have joined the contest. Hence, approximately 60% of the users seem to be laggards (maybe freeriders), as they join the contest at a time already 90% of the concepts are mentioned.

Like [3] conjectures, the second half of the ideation contest is different. But in our case it is not the high quality that causes separation, but rather it's merely changing network characteristics. 60% of the solvers join the contest within this phase creating 50% of the final raw material (in terms of ideas or words). On first sight, these ideas are of lower novelty, as they do not add new concepts. On the other hand, those ideas appear to be equally valuable as they receive 50% of all ratings and also more than half of the total rewards. Consequently, we will deepen the analysis of idea and concept aggregation within the next step.

4.3 How a specific concept develops within an ideation contest

As aforementioned, performing a MDS on a large number of items usually produces a cluttered map that is hard to interpret. This is the case if we consider the concept maps for higher N (c.f. the network at N=725 in Figure 5). To avoid the noise, created in those maps and provide more detailed view on the ideation process we will limit the amount of ideas within this step. As a result, we will again analyze the co-occurrence of concepts, but only in regards to a centralized concept. We use the concept of *Versioning/ Customization,* which is the concept with the highest betweenness centrality within the finished ideation network (at N= 725). The concept is used by 75 ideas which suggest that the motorbike of the future should be built modular, so that a vendor or the customer himself can reassemble various parts and hence, customize the motorbike towards differentiating customer needs. Below, we will present the results of applying the procedure model on this limited subset of N=75 ideas, neglecting all the other (650) ideas.

The first co-occurrence of the concept happens with the submission of the 23rd overall idea. As the 2nd idea overall already suggested variety in frame designs and storage capacity to create customizable motorbikes (coded with *Frame/ Storage* and *Versioning/ Customization*), the 23rd idea reintroduced the concept, of customization, but from a business model perspective, suggesting focusing on new kinds of customer groups (coded with *Brand/ Customer Segment* and *Versioning/ Customization*). With the ideation contests continuing, more ideas adopt the concept of *Versioning/ Customization*. Thus, due to repeated measures and MDS

Versioning/ Customization represents a centered concept whose frequency will be plus one for every idea added. This is visible by increasing node size of the centered concept in Figure 7.

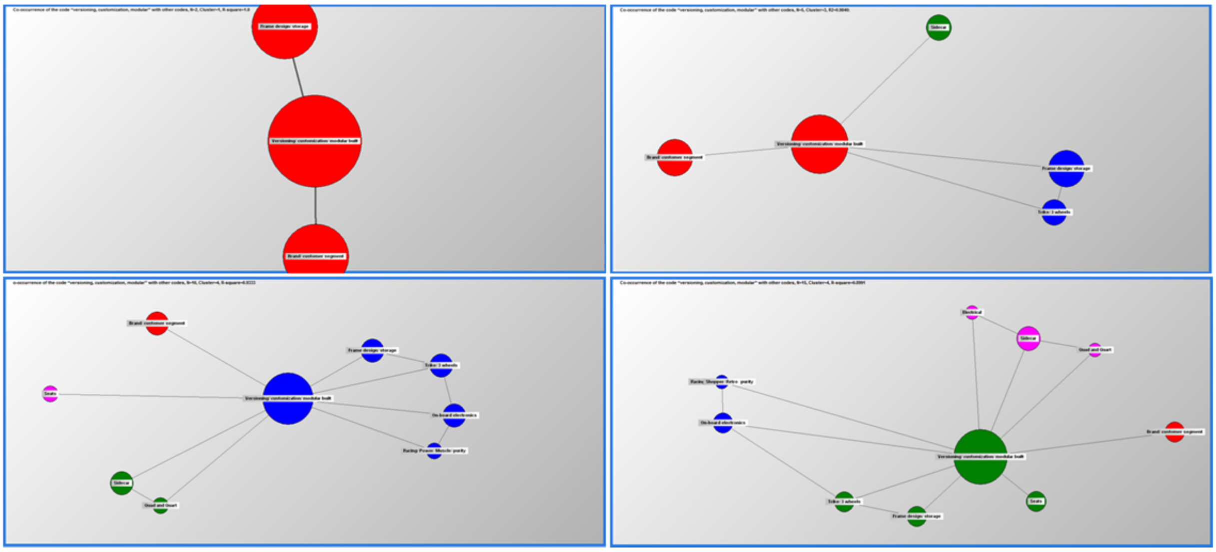

Figure 7 Cluster development of concepts co-occurring with the centered concept of *Versioning/ Customization* at N=2, 5, 10 and 15 ideas using the centered concept

Figure 7 shows, that concepts are rarely copied during this stage of the contest, but rather enhanced. For instance, at N=5 (by the 34th overall idea) the first concept (from N=1) is enhanced. The solver takes the initial idea (to *vary frames and storage*) and combines it with the idea to build it as a *trike* (c.f. Figure 7, top right, blue cluster). At N=15, four cluster are built. For example, the pink cluster collects ideas that suggest reinventing the *sidecar*. However, this idea is enhanced by the ideas of an *electrical engine* and an idea which suggests *building bike and sidecar as modules of a quad* (which is a 4 wheel motorbike). But the idea of an electrical quad itself has not yet been mentioned at this time of the contest as we can recognize by the missing edge between those concepts.

During the ideation contest, discrete "sub-sub-concepts" such as *customizable motorbikes with sidecars,* are gleaned, reintroduced and enhanced with other concepts, but also copied, neglected or forgotten. Staying with the mentioned example of a *sidecar* illustrates further characteristics of the ideation function. Ideas suggest special protection solutions for the passenger sitting in the sidecar (at N=18), multimedia to entertain the passenger (at N=42), or highly comfortable seats for the sidecar (at N=44). However, at the same time the relative frequency of concepts within this sub-network is decreasing. The reason for this is that other ideas (for instance the multimedia solution) get adopted by different, faster developing concepts. This leads to gaining density of the red colored cluster, which is created around the centered concept. Hence, at N=36, the concept of an *electrical engine* switches from the cluster around the idea of a *sidecar* towards this centered (red colored) cluster. Repeated co-occurrence of codes within the centered cluster (for example a *scooter-type* of bike that has *roll bars* and hence, can be seen as a *convertible*) leads to higher density and cohesion within this cluster. Figure 8 illustrates these changes of the code network structure, foremost the situation of the central cluster becoming more powerful.

Despite this development, it remains hard to determine a certain tipping point [27] nor opinion building or a groupthink situation [25] in this ideation function. Likewise the situation does not reflect ideation as it is suggested by [3] and even provides an inverse ideation function to the one that is suggested by bounded ideation theory [15]. Like the ideation network of the overall ideation contest, also the sub-network of the developing concept is built in phases. Again, the second half (from N=38 until N=75) is not less than productive than the first half. Although it does not stand for many novel concepts (nodes), 70% of the new combinations or aggregation (edges) is set here. To sum up the repeated measures steps from our procedure model, this single case study shows a three-phased ideation function, depicted for the single cluster development in Figure 9.

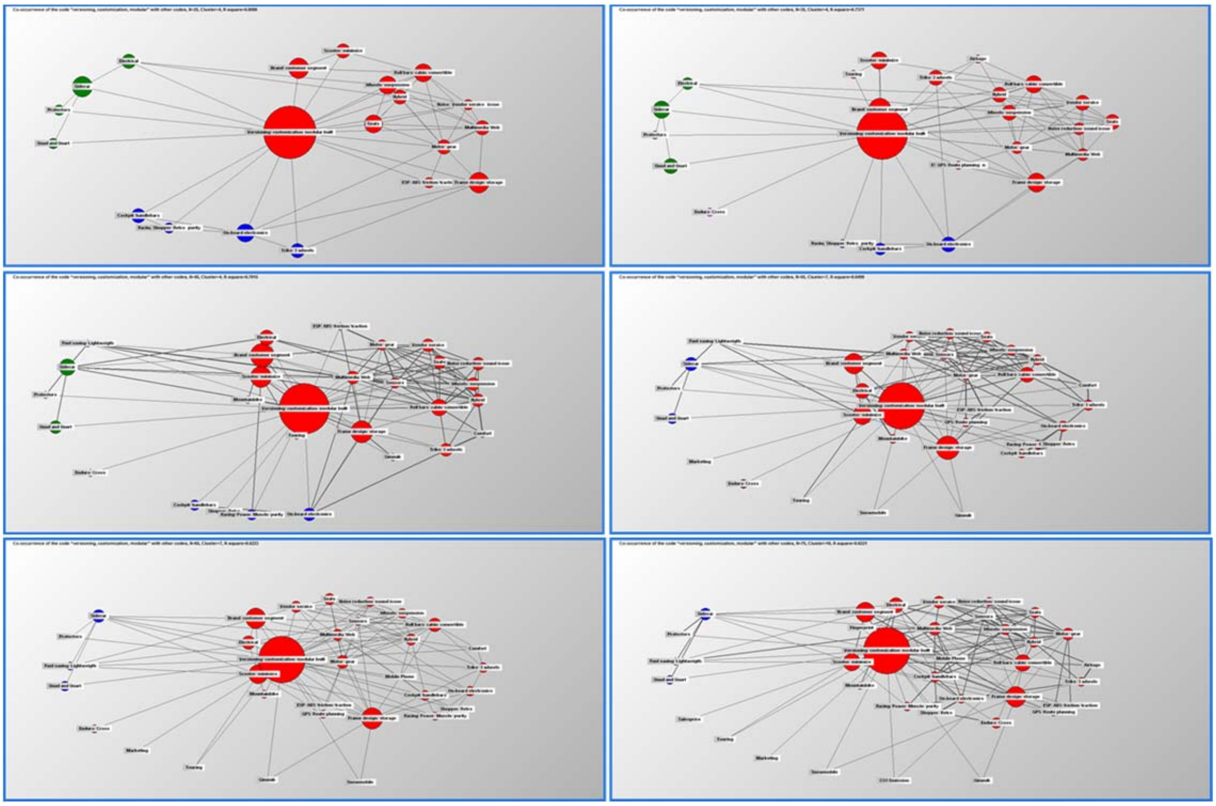

Figure 8 Cluster development of concepts co-occurring with the centered concept of *Versioning/ Customization* at N=25, 35, 45, 55, 65 and 75 (from top left to bottom right).

From the ideation contest's start, early movers and the early majority starts brainstorming. During this phase a large variety of novel concepts is introduced. Some of those ideas will be solid ground for other solvers to adopt on and some of those early ideas will be neglected by the crowd. During the middle part of the ideation function solvers already focus on aggregation. This results in strong cluster further being strengthened while some "outstanding" ideas (representing single-item cluster) still are not further pursued by the crowd. The final phase of ideation focusses on strengthening already existing co-occurrences of various concepts (weighting of existing edges). The negative description of this phase is "copying". A clear sign of saturation is the network's density (c.f. Figure 9) or the ratio of new edges per new idea.

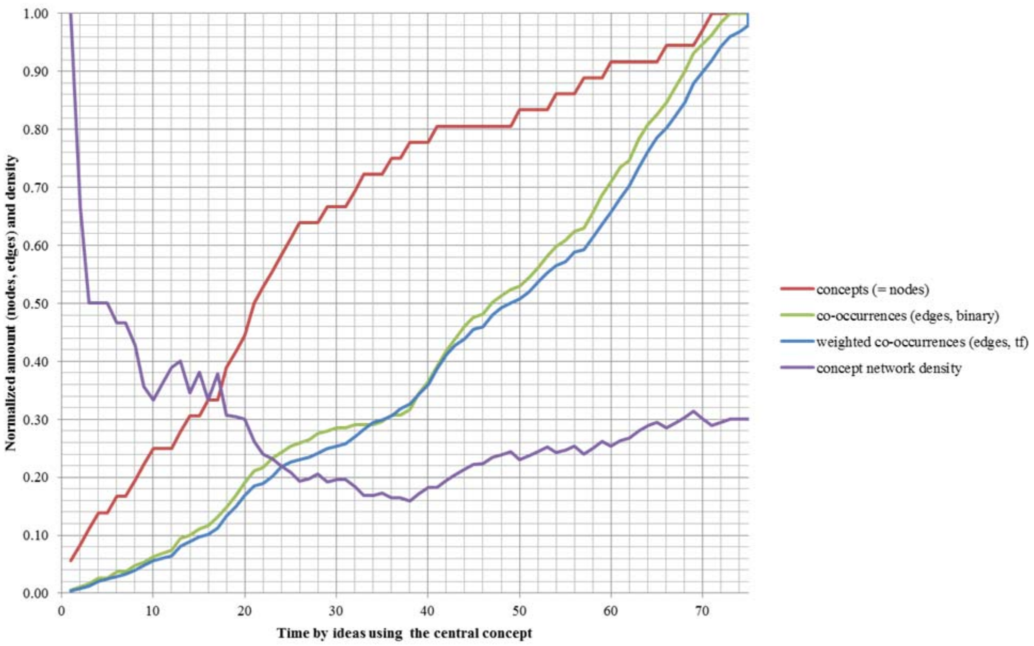

Figure 9 Ideation function of a centered concept within an ideation contest

To deepen the aspect, how ideation characteristics are reflected by the final results of an ideation contests we will provide a final comparative analysis of rewarded and unrewarded ideas in the next chapter.

4.4 Where good ideas come from

As literature review and theory have shown, the issue of ideation quality opens a multi-sided discussion between researchers. To this point, research only achieved uncertainty what factors have influence on ideation quality and how ideation quality should be measured. In addition to factors which measure the impact of contests design features, extrinsic or intrinsic motivation on quality, we are able to determine if ideation network characteristics (introduced in 3.3) may are aligned with the quality. Using our existing data provides two opportunities of indirect quality measurements. First, we can make use of the solvers ratings of their own ideas and second, we can use rewards given by the seeker. As ratings might be biased by individual connections among solvers as well as their different perception and hence, practice of the rating mechanism [44], we will make use of the rewards that are given by a seeker's expert committee. Therefore, we act on the assumptions, that the expert committees follow a fair rewarding process, are able to cope with the quantity of ideas (which is a problem according to [83] and also stick to their announced measurement criteria.

In the selected case the seeker's task description calls novelty (*We are looking for unique, breakthrough products and services.)* and feasibility *(The solutions should embody innovative technology, passion and modern lifestyle.)* as relevant criteria of quality and reward twelve out of the 725 ideas. Therefore, our next step is to turn to the "best versus the rest" of the ideas in terms of their network characteristics. As noted above, in terms of an evolving ideation network, an idea is novel if it creates a new node (introduces a new concept) or a new edge (is the first that combines two nodes which have been mentioned before, but separately). A third nuance of novelty can be defined in the fact, whether the combination of nodes in an idea already has been used at the time of submission. For instance, an idea might use four different concepts (nodes) and hence, six edges, but none of the edges represents novelty as they might have been used pairwise by various seekers already. However, the situation of combining exactly those four concepts might still be novel. Table 6 applies those metrics to compare twelve rewarded and the 713 unrewarded ideas from our example.

Table 6 Comparison of concept network metrics for rewarded and unrewarded ideas

Concept network metrics	**Node frequenc y**	**New nodes created**	**Nodes weighted**	**Edge frequency**	**New edges created**	**Edges weighted**	**Novel node combination at time of submission**	**Between-ness centrality**
12 rewarded ideas								
• **mean**	3.0	0	3.0	3.50	1.33	2.17	75.0%	17.88
• **(SD)**	(1.33)	(0.0)	(1.33)	(2.23)	(1.09)	(0.88)		(12.20)
• **median**	3	0	3	4	1	2		22.35
713 unrewarded ideas								
• **mean**	2.21	.10	2.11	1.09	.68	.41	37.2%	9.15
• **(SD)**	(0.69)	(0.08)	(0.86)	(0.92)	(0.51)	(0.21)		(15.62)
• **median**	2	0	2	1	1	1		8.35
Mann-Whitney U-test							-	
• **Z-value**	-1.355	.887	-1.821	-2.015	-1.956	-1.744		-2.26
• **(p-value)**	(.045)	(.023)	(.067)	(.052)	(.50)	(.38)		(.016)

As Table 4 shows, 70 different concepts are used within 725 ideas. Enhancing existing concepts leads to a total concept frequency of 1575, meaning on average, every idea includes 2.17 concepts. Remarkably, none of the rewarded ideas introduces a new concept, although on average they combine more concepts than unrewarded ideas (3.0 to 2.21). In other words, all rewarded ideas use concepts, which, by the time the rewarded ideas are submitted, already exist within the contest. Hence, the assumption is that rewarded ideas do well in enhancing these concepts. Looking at the edges stretches evidence to this hypothesis. Within the 725 ideas the 70 concepts are tied by 468 edges. Additionally, 334 times an edge is strengthened by repetition. As a result, on average an idea coins 1.09 edges, of which 0.68 represent a new connection and 0.41 a repetition. In contrast, the twelve rewarded ideas create 3.5 edges on average of which 1.33 are new combinations and 2.17 repetitions. Additionally, by the time they are submitted, 9 of the 12 rewarded ideas establish a new, and hence so far unique, combination of concepts.

We can see in Table 6, that the rewarded ideas have higher betweenness centrality than unrewarded ideas (17.88 to 9.15). Also there is a lot of variation in betweenness centrality for the overall network (for example SD = 15.62 relative to a mean betweenness of 9.15 for all unrewarded ideas). This makes sense, because we know that as edges can be created by the use of one out of of 70 concepts, most connections between ideas can be made in this network without the aid of any intermediary. Hence, combining the network numbers and the fact that there cannot be a lot of betweenness, the difference between rewarded and unrewarded ideas is a significant hint that I this case, rewarded ideas are somehow bridging structural holes within the ideation network. Figure 10 illustrates the network of all ideas. Rewarded ideas are colored yellow and the edges towards their ten closest neighbors (with shortest distances in the distance matrix) colored red. We can see that some of the rewarded ideas (for example the three depicted in the middle-top position of the network) are among closest neighbors to each other, stating they are very similar according to coded concept categories. Even though the network of 725 ideas is a bit cluttered and per definition produces some distortion, we are able to recognize rather central positions of rewarded ideas.

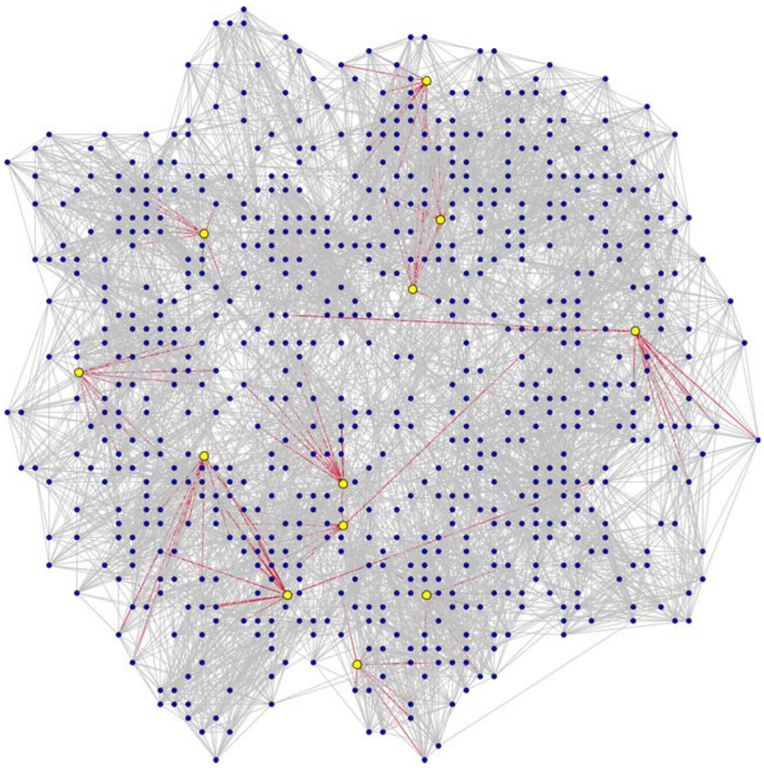

Figure 10 Idea network, based on coding. Rewarded ideas in yellow, edges to 10 nearest neighbors in red

Summing up this analysis, we can say that in this particular example, rewarded ideas are not the ones which introduce new concepts, but rather those which combine them in a clever fashion. The network characteristics of rewarded ideas suggest that they are rich in information (given the fact that they include 3 concepts on average), even though particular concepts have been applied before. Rewarded ideas are able to introduce a novel way of combining those concepts (given the fact that on average they produce 1.33 new combinations). This means that even though their content richness includes some repetitive nature, the ideas over all represent novel combinations. As a result they are positioned central, in-between all other ideas in the ideation network. A hypothesis, taken from these results could be, that high quality ideas might be boundary spanning. Their main strength might be situated in reaching across different concept borders in order to build relationships, interconnections and interdependencies that finally solve a complex problem.

5 Discussions

We recombine long existing methods, which are to say coding, distance measures, clustering mechanisms, MDS and SNA to develop a procedure model which aims to analyze the process of ideation within online ideation contests. Our intention is to develop an approach that might help to understand how innovative ideas are formed and embedded within ideation contests. We focus on the process of ideation (often called the ideation function), not on input factors or the results. We put the developed procedure model to practice by applying it to analyze an ideation contests, taken from a crowdsourcing platform. The empirical results are based on 725 ideas, which were submitted to an ideation contest searching for innovative concepts to create the motorbike of the future. The analysis of repeated measures for the overall contest and the development of a single concept as well as the comparative analysis of rewarded and unrewarded ideas allows us to discuss some theoretical as well as managerial contributions as well as analogous limitations.

Literature on ideation is often vague, and does not proof the assumed ideation function by empirical studies. To the best of our knowledge, our study is the first one that uses a repeated measures design to analyze an ideation contests in a step-by-step manner. Furthermore we make use of methodologies which, in our opinion, are underrepresented in the rapidly growing research field of online communities. The results from applying our procedure model to a real world dataset suggest that ideation theory might not be sufficient in explaining the process (ideation functions) of modern online ideation contests.

Contrary to [3] conjecture that the value of ideas in ideation will increase over time, we find that increasing time itself changes the context of ideation. Ideas which are submitted during the second half of an ideation contests find utterly different starting conditions and hence, are hardly comparable with ideas that are submitted in the early stage. In terms of novelty we find a degressive collinear ideation curve within our example. The reason for this is that an early brainstorming phase is highly productive and forecloses the chance of later ideas to find a concept that has not yet been mentioned. Albeit this seems to be frustrating in terms of ideation novelty, late ideas still give the impression of being highly useful in terms of feasibility. As the nature of ideation changes within the second half, ideas base on existing concepts and enhance them, mostly by recombining and drawing analogies. Hence, in our case we are also not able to provide evidence to the theories of groupthink [25] or the tipping point [27]. Additionally, our ideation function is converse to the one conjectured by bounded ideation theory [15]. The difference we suggest is constituted by the fact that in modern online ideation contests, solvers mostly are competitors, not collaborators. In that manner, taking a solver's perspective, theories like tipping points, bounded ideation theory, collective intelligence or groupthink do not provide helpful strategies to get an idea rewarded.

As a result from our survey, we argue that the often-quoted quality vs. quantity discussion [6], [14], [15], [20], [21], [24], [48], [57], [60], which essentially implies that it requires high amounts of diverse and autonomous ideas to include highly valuable good ideas or even breed a collective intelligence, might not be the best way of understanding the problem. Of course, we can find evidence for such assumption, as rewarded ideas are also a product of previous ideas. But at the same time such quality vs. quantity conjecture misses the point of ideation. Our study suggests that albeit a seeker's task description, solvers address different quality criteria throughout a single ideation contest. Within hundreds of solvers participating, chances to score high in quality criteria such as novelty or feasibility are not uniformly distributed over time. As a result solvers that participate early in an ideation contest might be more likely to briefly search what kind of concepts are already mentioned and simply add a different concept to be the first that has mentioned it. Dealing with saturation in terms of concepts, solvers that enter during later stages have to scrutinize previous ideas to be able to differentiate themselves. Therewith their task is different. Eventually their only chance to stand out is a particularly clever combination of concepts. Hence, as a plurality of follow this strategy, the amount of ideas, words as well as the network density rises. The result of this process is that researchers as well as seekers perceive ideation contests as a whole like noise from which it is hard to detect the signals [83].

Consequently, our study draws three final conclusions. The first is, that in order to understand ideation, research has to analyze ideation. This addresses the mentioned research gap of threating ideation contests as a black box. The consequence is that surveys among solvers, or their perception of contests characteristics, are less significant than the real behavior and action happening inside an ideation contest. As we've shown, there is a plurality of yet only occasionally used methods, which can be applied to measure ideation processes. Still, like theories, such methods must undergo rigorous empirical testing to see if propositions made are useful. Our second conclusion is, that the discussion about where good ideas comes from might has to take a step back and concentrate on the quality defining criteria separately. Researcher's common quality criteria, which often are a measure of aggregated characteristics like novelty, feasibility, relevance or elaboration, might be to manifold and of no avail for ideation contests. Instead of searching for magic bullets, crowdsourcing platforms or research could split ideation contests into different phases and aligning them with different quality criteria.

Therefore our third conclusion is that crowdsourcing platforms might try to offer ideation contests in a bit by bit fashion. As structuring process steps is a common approach (see [12], [94]), the ideation function from our study reasons an ideation contest to run as following: An ideation contests could start by a broad task description by the seeker and a few brainstorming days. The assumption is, that as there will be an early mover crowd, this phase will already collect obvious ideas, analogies, and curios concepts. Maybe limiting the amount instead of the timespan might be a helpful adjustment. Then the seeker ends this phase, preselects concepts and rewards the bests based on novelty or the ideas ability to surprise, etc. Subsequently, the seeker introduces a new ideation contest using the feedstock from brainstorming as input (maybe in the task description or as compulsory reading to enter the contest). The second ideation contests may intend to achieve higher feasibility, to elaborate on the existing ideas in detail or strengthening their relevance. Some contest parameters such as a required minimum length of an idea or required tagging of ideas by using the concepts from brainstorming might help as guidance and to avoid a cluttered overall result. Our hypothesis is, that such a split might attract different types of solvers. Creative thinker (maybe without a high educational background) will be more efficient within the first part, while gifted tinkerers (say puzzlers) and inventors will enhance the second phase.

Still it can and should not be concluded from a study of a single data sample of the applied procedure model, which combination or combinations of these techniques should be implemented in any particular situation. However, the identification of helpful techniques in this study may guide future efforts to determine ideal combination of data mining and SNA during the analysis of ideation.

6. References

[1] InnoCentive, "Winning Solutions and Solvers at InnoCentive," *innocentive.com*, 2012. [Online]. Available: http://www.innocentive.com/for-solvers/winning-solutions/low-cost-rainwater-storage-system. [Accessed: 14-Nov-2012].

[2] M. Bishop, "The Total Economic Impact TM Of InnoCentive Challenges," Camebridge, 2009.

[3] A. F. Osborn, *Applied Imagination*. Scribner, 1957, p. 317.

[4] J. F. Nunamaker, A. R. Dennis, J. S. Valacich, D. R. Vogel, and J. F. George, "Electronic Meeting Systems to Support Group Work," *Communications of the ACM*, vol. 34, no. 7, pp. 40–61, 1991.

[5] J. Howe, "The rise of Crowdsourcing," *Wired Magazine, Vol. 14 (6)*, pp. 1–4, 2006.

[6] T. W. Malone, R. Laubacher, and C. Dellarocas, *The collective intelligence genome*, vol. 38, no. 3. 2010, pp. 21–31.

[7] MIT Center of Collective Intelligence, "Handbook of Collective Intelligence," 2012. [Online]. Available: http://scripts.mit.edu/~cci/HCI/index.php?title=Main_Page#Diversity. [Accessed: 26-Jan-2012].

[8] J. Surowiecki, *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. New York: Little Brown, 2004.

[9] E. Bonabeau, "Decisions 2.0: The Power of Collective Intelligence," *MIT Sloan Management Review*, vol. 50, no. 2, 2009.

[10] T. W. Malone, R. Laubacher, and C. Dellarocas, "Harnessing Crowds : Mapping the Genome of Collective Intelligence," *Elements*. pp. 1–20, 2009.

[11] T. Connolly, L. M. Jessup, and J. S. Valacich, "Effects of Anonymity and Evaluative Tone on Idea Generation in Computer-Mediated Groups," *Management Science*, vol. 36, no. 6, pp. 689–703, 1990.

[12] N. Dalkey and O. Helmer, "An experimental application of the Delphi method to the use of experts," *Management Science*, vol. 9, no. 3, pp. 458–467, 1963.

[13] S. W. Knoll and G. Horton, "Changing the Perspective: Improving Generate thinkLets for Ideation," *HICSS 2010 Proceedings*, pp. 1–10, 2010.

[14] K. Girotra, C. Terwiesch, and K. T. Ulrich, "Idea Generation and the Quality of the Best Idea," *Management Science*, vol. 56, no. 4, pp. 591–605, Feb. 2010.

[15] R. O. Briggs and B. A. Reinig, "Bounded Ideation Theory : A New Model of the Relationship Between Idea- quantity and Idea-quality during Ideation," in *HICSS 2007 Proceedings*, 2007, pp. 1–10.

[16] M. Diehl and W. Stroebe, "Productivity loss in brainstorming groups: Toward the solution of a riddle," *Journal of Personality and Social Psychology*, vol. 53, no. 3, pp. 497–509, 1987.

[17] A. R. Dennis and J. S. Valacich, "Computer Brainstorms : More Heads Are Better Than One," *Journal of Applied Psychology*, vol. 78, no. 4, pp. 531–537, 1993.

[18] T. M. Amabile, *Creativity in context: Update to the social psychology of creativity*. Westview Press, 1996, p. XVIII, 317 S.

[19] C. Terwiesch and K. T. Ulrich, *Innovation Tournaments: Creating and Selecting Exceptional Opportunities*. Boston: Harvard Business School Press, 2009, p. 425.
[20] D. L. Dean, J. M. Hender, T. L. Rodgers, and E. L. Santanen, “Identifying quality, novel, and creative Ideas: Constructs and scales for idea evaluation,” *Journal of the Association for Information Systems*, vol. 7, no. 10, pp. 646–699, 2006.
[21] I. Blohm, U. Bretschneider, J. M. Leimeister, and H. Krcmar, “Does Collaboration among Participants Lead to Better Ideas in IT-Based Idea Competitions? An Empirical Investigation,” in *2010 43rd Hawaii International Conference on System Sciences*, 2010, pp. 1–10.
[22] P. Levy, *Collective Intelligence: Mankind’s Emerging World in Cyberspace*. Plenum Trade, 1997, p. 277.
[23] J. M. Leimeister, “Collective Intelligence,” *Business & Information Systems Engineering (BISE)*, vol. 02, no. 04, pp. 245–248, Jun. 2010.
[24] E. Bonabeau, “Decisions 2.0 : The Power of Collective Intelligence,” *MIT Sloan Management Review*, vol. 50, no. 02, pp. 45–52, 2009.
[25] I. L. Janis, “Groupthink and group dynamics: A social psychological analysis of defective policy decisions,” *Policy Studies Journal*, vol. 2, no. 1, pp. 19–25, 1973.
[26] M. E. Turner and A. R. Pratkanis, “Twenty-Five Years of Groupthink Theory and Research: Lessons from the Evaluation of a Theory,” *Organizational behavior and human decision processes*, vol. 73, no. 2/3, pp. 105–15, Feb. 1998.
[27] M. Gladwell, *The Tipping Point: How Little Things Can Make a Big Difference*, vol. 1. Back Bay Books, 2000, p. 304.
[28] A. Pinsonneault, H. Barki, R. B. Gallupe, and N. Hoppen, “Electronic Brainstorming: The Illusion of Productivity,” *Information Systems Research*, vol. 10, no. 2, pp. 110–133, Jun. 1999.
[29] G. DeSanctis and R. B. Gallupe, “A foundation for the study of group decision support systems,” *Management Science*, vol. 33, no. 5, pp. 589–609, 1987.
[30] C. Shapiro and H. R. Varian, *Information Rules: A Strategic Guide to the Network Economy*. Boston: Harvard Business School Press, 1997, pp. 1–352.
[31] A. R. Dennis, J. S. Valacich, T. Connolly, and B. E. Wynne, “Process Structuring in Electronic Brainstorming,” *Information Systems Research*, vol. 7, no. 2, pp. 268–277, 1996.
[32] K. R. MacCrimmon and C. Wagner, “Stimulating Ideas through Creativity Software,” *Management Science*, vol. 40, no. 11, pp. 1514–1532, 1994.
[33] A. Kittur and R. E. Kraut, “Harnessing the wisdom of crowds in wikipedia,” in *Proceedings of the ACM 2008 conference on Computer supported cooperative work CSCW 08*, 2008, p. 37.
[34] C. Wagner, “Wikis And Beyond: Harnessing Collective Intelligence Through Web 2.0 Technologies,” in *AMCIS 2009 Proceedings*, 2009, p. Paper 602.
[35] C. Terwiesch and Y. Xu, “Innovation Contests, Open Innovation, and Multiagent Problem Solving,” *Management Science*, vol. 54, no. 9, pp. 1529–1543, Sep. 2008.
[36] J. Bergman, A. Jantunen, and J.-M. Saksa, “Enabling Open Innovation Process Through Interactive Methods: Scenarios and Group Decision Support Systems,” *International Journal of Innovation Management*, vol. 13, no. 01, pp. 139–156, 2009.
[37] C. Wagner and A. Back, “Group Wisdom Support Systems: Aggregating the Insights of Many through Information,” *Information Systems*, vol. IX, no. 2, pp. 343–350, 2008.
[38] H. W. Chesbrough, *Open innovation: The new imperative for creating and profiting from technology*. Boston: Harvard Business School Press, 2003.
[39] A. C. Rouse, “A Preliminary Taxonomy of Crowdsourcing,” in *ACIS 2010 Proceedings*, 2010.
[40] D. Geiger, S. Seedorf, T. Schulze, R. Nickerson, and M. Schader, “Managing the Crowd : Towards a Taxonomy of Crowdsourcing Processes,” in *AMCIS 2011 Proceedings*, 2011, pp. 1–11.
[41] Y. Zhao and Q. Zhu, “Evaluation on crowdsourcing research: Current status and future direction,” *Information Systems Frontiers*, Apr. 2012.
[42] A. C. Bullinger and K. Moeslein, “Innovation Contests – Where are we ? Innovation Contests – Where are we ?,” in *AMCIS 2010 Proceedings.*, 2010, p. Paper 28.
[43] J. L. Jones, R. Webb Collins, and D. J. Berndt, “Information Markets : A Research Landscape,” *Communications of the Association for Information Systems*, vol. 25, no. Article 27, pp. 289–304, 2009.
[44] C. Riedl, I. Blohm, J. M. Leimeister, and H. Krcmar, “Rating Scales For Collective Intelligence in Innovation Communities: Why Quick and Easy Decision making Does Not Get It Right,” in *ICIS 2010 Proceedings*, 2010, p. Paper 52.
[45] Y. Yang, P.-Y. Chen, and P. Pavlou, “Open Innovation : An Empirical Study of Online Contests,” in *ICIS 2009 Proceedings*, 2009, pp. 1–16.
[46] O. A. Acar and J. Van Den Ende, “Motivation, Reward Size and Contribution in Idea Crowdsourcing,” in *Dime-Druid Winter Conferece*, 2011, p. 29.
[47] J. Haller, A. C. Bullinger, and A. Neyer, “Beyond the Black Box of Idea Contests,” in *EURAM Annual Conference*, 2009, pp. 1–37.

[48] T. P. Walter and A. Back, “Towards Measuring Crowdsourcing Success: An Empirical Study on Effects of External Factors in Online Idea Contests,” in *The 6th Mediterranean Conference on Information Systems*, 2011, pp. 1–12.

[49] W. Ebner, J. M. Leimeister, and H. Krcmar, “Community engineering for innovations: the ideas competition as a method to nurture a virtual community for innovations,” *R&D Management*, vol. 39, no. 4, pp. 342–356, Sep. 2010.

[50] K. R. Lakhani, P. A. Lohse, J. A. Panetta, and L. B. Jeppesen, “The Value of Openness in Scientific Problem Solving,” *Biotech Business*. pp. 1–57, 2007.

[51] K. J. Boudreau and K. R. Lakhani, “How to Manage Outside Innovation How to Manage Outside Innovation,” *MIT Sloan Management Review*, vol. 50, no. 4, pp. 69–76, 2009.

[52] M. K. Poetz and M. Schreier, “The Value of Crowdsourcing: Can Users Really Compete with Professionals in Generating New Product Ideas?,” *Journal of Product Innovation Management*, vol. 29, no. 2, pp. 245–256, Mar. 2012.

[53] O. Nov, M. Naaman, and C. Ye, “Motivational , Structural and Tenure Factors that Impact Online Community Photo Sharing,” in *Proceedings of the Third International ICWSM Conference*, 2009, pp. 138–145.

[54] B. L. Bayus, “Crowdsourcing and Individual Creativity over Time: The Detrimental Effects of Past Success,” *Wired*. 2010.

[55] N. Franke and S. Schah, “How communities support innovative activities: an exploration of assistance and sharing among end-users,” *Research Policy*, vol. 32, no. 1, pp. 157–178, Jan. 2003.

[56] Z. Emden, R. J. Calantone, and C. Droge, “Collaborating for New Product Development: Selecting the Partner with Maximum Potential to Create Value,” *Journal of Product Innovation Management*, vol. 23, no. 4, pp. 330–341, Jul. 2006.

[57] J. Björk and M. Magnusson, “Where Do Good Innovation Ideas Come From? Exploring the Influence of Network Connectivity on Innovation Idea Quality,” *Journal of Product Innovation Management*, vol. 26, no. 6, pp. 662–670, 2009.

[58] C.-L. Hung, J. C.-L. Chou, and T.-P. Dong, “Innovations and communication through innovative users: An exploratory mechanism of social networking website,” *International Journal of Information Management*, no. 2010, pp. 1–10, Jan. 2011.

[59] D. Chandler and A. Kapelner, “Breaking monotony with meaning: Motivation in crowdsourcing markets,” *University of Chicago mimeo*. 2010.

[60] M. K. Poetz and M. Schreier, “The value of crowdsourcing: can users really compete with professionals in generating new product ideas?,” *Innovation*, vol. forthcomin, no. Fourthcoming, pp. 1–37, 2009.

[61] Y. Yang, P. Chen, and R. Banker, “Impact of Past Performance and Strategic Bidding on Winner Determination of Open Innovation Contest,” in *Pre ICIS Workshop on Information Systems and Economics*, 2010, p. 29.

[62] O. Toubia, “Idea Generation, Creativity, and Incentives,” *Marketing Science*, vol. 25, no. 5, pp. 411–425, Sep. 2006.

[63] Eilbeck J.C., Enol’skii V.Z., Aiken M., Vanjani M., and Paolillo J., “A comparison of two electronic idea generation techniques,” *Information & Management Management*, vol. 30, no. 2, p. 9, 1996.

[64] T. Kohler, J. Fueller, K. Matzler, and D. Stieger, “Co-Creation in Virtual Worlds: The Design of the User Experience,” *MIS Quarterly*, vol. 35, no. 3, pp. 773–788, 2011.

[65] W. Ebner, J. M. Leimeister, U. Bretschneider, and H. Krcmar, “Leveraging the Wisdom of Crowds: Designing an IT-Supported Ideas Competition for an ERP Software Company,” in *HICSS*, 2008, pp. 417–417.

[66] J. M. Leimeister, M. Huber, U. Bretschneider, and H. Krcmar, “Leveraging Crowdsourcing: Activation-Supporting Components for IT-Based Ideas Competition,” *Journal of Management Information Systems*, vol. 26, no. 1, pp. 197–224, Jul. 2009.

[67] A. C. Bullinger, A.-K. Neyer, M. Rass, and K. M. Moeslein, “Community-Based Innovation Contests: Where Competition Meets Cooperation,” *Creativity and Innovation Management*, vol. 19, no. 3, pp. 290–303, Aug. 2010.

[68] D. C. Brabham, “Crowdsourcing as a Model for Problem Solving: An Introduction and Cases,” *The International Journal of Research into New Media Technologies*, vol. 14, no. 1, pp. 75–90, Feb. 2008.

[69] L. Muhdi, M. Daiber, S. Friesike, and R. Boutellier, “The crowdsourcing process: an intermediary mediated idea generation approach in the early phase of innovation,” *International Journal of Entrepreneurship and Innovation Management*, vol. 14, no. 4, pp. 315–332, 2011.

[70] E. Dubach, L. Muhdi, D. Stöcklin, and F. Michahelles, “Crowdsourcing for ‘ Kiosk of the Future ’ – A Retail Store Case Study,” in *AMCIS 2011 Proceedings*, 2011, p. Paper 324.

[71] J. Helquist, J. Kruse, and M. Adkins, “Group support systems for very large groups : A peer review process to filter brainstorming input review process to filter brainstorming input,” in *AMCIS 2006 Proceedings*, 2006, p. Paper 16.

[72] T. P. Walter and A. Back, "Crowdsourcing as a Business Model : An Exploration of Emergent Textbooks Harnessing the Wisdom of Crowds," in *BLED 2010 Proceedings*, 2010, pp. 555–568.
[73] T. Schulze, M. Indulska, D. Geiger, and A. Korthaus, "IDEA ASSESSMENT IN OPEN INNOVATION : A STATE OF PRACTICE," in *ECIS 2012 Proceedings*, 2012, p. Paper 149.
[74] D. C. Brabham, "Crowdsourcing the Public Participation Process for Planning Projects," *Planning Theory*, vol. 8, no. 3, pp. 242–262, Jul. 2009.
[75] F. T. Piller and D. Walcher, "Toolkits for idea competitions: a novel method to integrate users in new product development," *R&D Management*, vol. 36, no. 3, pp. 307–318, Jun. 2006.
[76] L. Chen and D. Liu, "Comparing Strategies for Winning Expert-rated and Crowd-rated Crowdsourcing Contests : First Findings," in *AMCIS 2012 Proceedings*, 2012, p. Paper 16.
[77] K. Frey, S. Haag, and V. Schneider, "The Role of Interests , Abilities , and Motivation in Online Idea Contests," in *Wirtschaftsinformatik 2011 Proceedings*, 2011, p. Paper 71.
[78] Y. Yang, P.-Y. Chen, and R. Banker, "Winner Determination of Open Innovation Contests in Online Markets," in *ICIS 2011 Proceedings*, 2011, no. Allio 2004, pp. 1–16.
[79] S. Adamczyk, A. C. Bullinger, and K. M. Moeslein, "Knowing is Silver , Listening is Gold : On the importance and impact of feedback in IT-based innovation contests Knowing is Silver , Listening is Gold : in IT-based innovation contests," in *Wirtschaftsinformatik 2011 Proceedings*, 2011, p. Paper 97.
[80] C. Wagner and A. Hayashi, "A New Way To Create Winning Product Ideas," *Journal of Product Innovation Management*, vol. 1994, no. 11, pp. 146–155, 1994.
[81] N. Archak and A. Sundararajan, "Optimal Design of Crowdsourcing Contests," in *ICIS 2009 Proceedings*, 2009, pp. 1–16.
[82] Y. Huang, P. V. Singh, and K. Srinivasan, "CROWDSOURCING ' BLOCKBUSTER ' IDEAS : A DYNAMIC STRUCTURAL MODEL OF IDEATION," in *ICIS 2011 Proceedings*, 2011, pp. 1–17.
[83] T. P. Walter and A. Back, "A Text Mining Approach to Evaluate Submissions to Crowdsourcing Contests," in *HICSS 2013 Proceedings*, 2013, pp. 1–10.
[84] M. Böhringer and P. Helmholz, "' What are they Thinking ?' - Accessing Collective Intelligence in Twitter," in *24th international Bled Conference*, 2011, pp. 310–320.
[85] A. Hendaoui and M. Limayem, "Idea consolidation process in face-to-face meetings : a new approach to organize and integrate individuals perspectives," in *ECIS 2006 Proceedings*, 2006, p. Paper 130.
[86] Y.-H. Tseng, Z.-P. Ho, K.-S. Yang, and C.-C. Chen, "Mining term networks from text collections for crime investigation," *Expert Systems with Applications*, vol. 39, no. 11, pp. 10082–10090, Feb. 2012.
[87] G. A. Wang, V. Tech, and W. Fan, "A knowledge adoption model based framework for identifying helpful user-generated content in online communities," in *32nd International Conference on Information Systems (ICIS)*, 2011, pp. 1–11.
[88] L. P. Linden, "A Tool for Identifying Swarm Intelligence on a Free / Open Source Software Mailing List," in *SAIS 2008 Proceedings*, 2008, p. Paper 29.
[89] A. C. Bullinger, A.-K. Neyer, M. Rass, and K. Moeslein, "Community-Based Innovation Contests: Where Competition Meets Cooperation," *Creativity and Innovation Management*, vol. 19, no. 03, pp. 290–303, Aug. 2010.
[90] S. Brinkkemper, "Method engineering : engineering of information methods and tools," *Information and Software Technology*, vol. 38, pp. 275–280, 1996.
[91] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Pearson Addison Wesley, 2006, p. 769.
[92] T. Miller, *Data and Text Mining: A Business Applications Approach*. Upper Saddle River: Prentice-Hall, 2005, p. 455.
[93] R. Feldmann and J. Sanger, *The text mining handbook : advanced approaches in analyzing unstructured data.* Boston: Cambridge University Press., 2007.
[94] J. S. Armstrong, *Principles of forecasting: A handbook for researchers and practioners*. New York: Springer, 2001, p. 849.
[95] K. Krippendorff, "Measuring the Reliability of Qualitative Text Analysis Data," *Quality Quantity*, vol. 38, no. 6, pp. 787–800, 2004.
[96] S. W. Menard, *Handbook of Longitudinal Research: Design, Measurement, and Analysis*, vol. 1. Academic Press, 2008, p. 668.
[97] W. de Nooy, A. Mrvar, and V. Batageli, *Exploratory Social Network Analysis with Pajek*. New York: Camebridge University Press, 2005, p. 334.
[98] J. Scott, *Social Network Analysis: A handbook*. London: Sage Publications, 2000, p. 208.
[99] M. Steinbach, G. Karypis, and V. Kumar, "A Comparison of Document Clustering Techniques," in *Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2000, no. X, pp. 1–2.
[100] A. Hotho, "A Brief Survey of Text Mining," *Forum American Bar Association*, pp. 19–62, 2005.
[101] C. Ding and X. He, "K -means clustering via principal component analysis," *Twenty-first international conference on Machine learning - ICML '04*, pp. 1–9, 2004.

[102] F. Can and E. A. Ozkarahan, “Concepts and Effectiveness of the Clustering Methodology for Text Databases,” *ACM Transaction on Database Systems*, vol. 15, no. 4, pp. 483–517, 1990.
[103] Z. Niu, D. Ji, and P.-N. Tan, “Using cluster validation criterion to identify optimal feature subset and cluster number for document clustering,” *Information Processing & Management*, vol. 43, no. 3, pp. 730–739, 2007.
[104] M. S. Granovetter, “The Strength of Weak Ties The Strength of Weak Ties1,” *The American Journal of Sociology*, vol. 78, no. 6, pp. 1360–1380, 1973.
[105] S. P. Borgatti, “Centrality and network flow,” *Social Networks*, vol. 27, no. 1, pp. 55–71, Jan. 2005.
[106] L. C. Freeman, S. P. Borgatti, and D. R. White, “Centrality in valued graphs: A measure of betweenness based on network flow,” *Social Networks*, vol. 13, no. 2, pp. 141–154, 1991.
[107] S. Wassermann and K. Faust, *Social Network Analysis: Methods and Applications*. New York: Camebridge University Press, 1994, p. 825.
[108] R. K. Yin, *Case Study Research: Design and Methods*, 3rd ed. Thousand Oaks, USA: Sage Publications, 2003, p. 180.
[109] P. Mayring, “Qualitative Content Analysis,” in *Forum Qualitative Social Research 2000 Proceedings*, 2000, vol. 1, no. 2, pp. 1–10.
[110] T. M. Amabile and S. Kramer, *The Progress Principle: Using Small Wins to Ignite Joy, Engagement, and Creativity at Work*. Harvard Business Review Press, 2011, p. 272.