The Suitability of Network Analysis of Social Media Data for the Prediction of Radical Innovations

Knut Jägersberg

School of Management and Governance

University of Twente

Enschede, The Netherlands

Master Thesis 29.01.2014

In Partial Fulfillment Of the Requirements for the Degree Master of Science In Business Administration,

Track Innovation & Entrepreneurship

First supervisor: Dr. E. Constantinides

Second supervisor: Dr. S.A. de Vries

UNIVERSITY OF TWENTE.

Abstract

Network patterns in social media meta-data are created, put in context, and found to be linked to innovations. Specifically, probabilistic patterns in historical network development are proposed to predict innovation. A series of studies on network developments in innovation related buzz was conducted and the patterns were tested in a quasi-experimental design, comparing natural groups of higher innovatively loaded buzz on radically innovative products with natural groups of less innovatively loaded buzz on 'traditional' products. Evidence for three predictive patterns could be found. In a second series of studies, two of the patterns were used to predict future innovations with a new methodology, for both known inventions as well as buzz sourced inventions, casted by a 'product-of-customer-interest' oracle methodology.

Management Summary

Research problem: Finding elements of an innovation prognosis software

Enterprises can gain game-changing competitive advantages from radical innovation. However, selecting R & D projects with radical potential still is a heuristic process, bounded by human rationality of management staff: knowing all relevant customer needs and especially predicting future needs is a matter of educated guessing by managers. This introduces uncertainty in the successful identification of promising future innovation. This research shows building blocks of a software, which helps to reduce uncertainty of prognostic market intelligence. The methodologies developed in this study advance prognostic market intelligence by externalizing prognostic knowledge and processing as big data on the customer voice found in social media and as software-based detection of patterns within that data for the forecasts.

Theoretical background: Stochastic patterns in network dynamics of social media activity

A social media monitoring service (Coosto) and a software toolchain were used to detect patterns to predict innovation. Sets of topics, which tend to be mentioned together in social media are found to construe network structures. Activity measures on these topics are observed, their co-evolution in the network structure is analyzed longitudinally, resulting in a dynamic network perspective on social media activity. These dynamics are analyzed for change tendencies by means of inferential statistics from the field of dynamic network analysis.

Methodologies: Detecting predictors of innovation and innovation forecasting

The first study sought to detect patterns, which predict innovation. Change tendencies, which are found by inferential statistics are contrasted between customer voices on radically innovative products and weakly innovative products. These differential patterns of stochastic change tendencies are the detected predictors of radical innovativeness.

The second study used these predictors to forecast innovations. New measurements of social media activity were taken and screened for activity networks with innovative change tendencies. This was done for activity networks on known, not-yet innovated inventions and on to be extrapolated, unknown inventions. Innovation forecasting for known inventions were straightforward: Customer voices on inventions, which had fitting patterns are found to have higher chance to be a radically innovative product then a weakly innovative product and that future innovation is simply the innovated invention. Innovation forecasting for unknown inventions, which are encrypted in the customer voice, was done in a more complicated process. First, all messages, which were constituent of the customer voice with innovative change tendencies, were subjected to the following decryption: Rapid content analysis by text-mining tools revealed the main theme complexes. The theme complexes were coded in ontological categories and the meaning-overlaps served as interest-weighting scheme. More prominent categories indicated parts of the theme complexes of higher customer interest. The relative interest was used to re-read the source themes from the raw customer voice in form of a product positioning statement, from themes of highest interest to lower interest to describe main product features and more augmented features.

Findings: Innovation prediction patterns and forecasted innovations

Contrasting 30 radically innovative and 32 weakly innovative networks revealed 3 predictive patterns in a quasi-experimental design. Two were used to forecast 9 innovations with increased likelihood to become radical, for example the commercialization of the bio-engineered production of oil from water (known invention) and the innovation of corporate-integrity-brokerage firms, which realize synergies in social responsibility marketing across organizational and sectorial boundaries (decrypted invention).

List of tables

List of figures

Abbreviations and definitions in order of mentioning

Terminology	Abbreviation	Definition
radical innovation	...	commercialization of a revolutionary new technology
disruptive innovation	...	commercialization creating new markets with new sets of value propositions, a radical innovation meeting customer needs
social meadia sphere	...	simplified network representation capturing the major outlines of a network structure of social media contents
buzz-network	...	abstract relationships amongst units of customer voice in social media, customer voice skeleton structures
network evolution	...	longitudinal change of network structure
network-behavior co-evolution	...	longitudinal change of an attribute associated to all nodes over the complete observed period in dependence of node-identity or network structual changes
trending topics	...	topics identified by Coosto as being mentioned together with a specified term
social network analysis	SNA	methods creating understandings of social network structures and attributes
dynamic network analysis	DNA	SNA methods creating understandings of network evolution, i.e. inferential statistics on network evolution
simulation investigation for empirical network analysis	SIENA	software for inferential statistical analysis of network evolution
	RSIENA	SIENA implementation in R, a general purpose but commandline based software for mathematical analyses of virtual all kinds
Visual Social Networks	Visone	multipurpose-software for SNA and DNA with (inferential) statistical analysis of network (evolution) with GUI 'remote-control' of RSIENA
network change tendency	...	distinctive biases in network evolution
information routing group	IRG	group of people with a common interest (i.e. 50) who engage in lateral communication (communication between individuals trespassing institutional or hierarchical boundaries) for mutual benefit via a computer network
lateral communication	...	communication between individuals trespassing institutional or hierarchical boundaries
buzz-spill over, buzz-force	...	degree of a topic arousing reply messages in relation to the likelihood of a message to penetrate or affect other messages in social media
Social Network Image Animator	SoNIA	primarily software to visualize network evolution in movies, also has a functionality to transscribe network data into adjecency matrices
Suggested Upper Merged Ontology	SUMO	a formalized, comprehensive ontology
normalized buzz-force, interest-intensity	NBF	buzz-force rescaled to an interval scale of 0-250 for each network. Rescaling serves to deal with the software limitation of only analyzing network behaviors of integer numbers from 0-250
Normalized Google Distance, Google Similarity Distance	NGD	symmetric conditional probability of co-occurrence of two words, a measure of semantic similarity of two words
	StOCNET	open software system for the statistical analysis of social networks using advanced statistical models

1. Introduction

1.1. Background of the research – The management problem of bounded rationality in innovation strategizing

Managers have to make choices in innovation strategizing for the sake of the success of a firm. The importance of innovation for firm success is backed up by the finding, that R&D expenditure is positively related with profitability and long-term growth (Geroski et al., 1993). How should management do innovation strategizing?

One way of innovation strategy is to pursuit radical innovation and it is a favorable strategy for a number of reasons. Creating radical inventions is a core entrepreneurial activity and important for wealth creation (Kirchhoff, 1991; Schumpeter, 1975), it has societal importance.

Innovation is the translation of an invention into the economy / the commercialization of an invention (US Department of Commerce, 1967). A radical innovation is thus commercialization of a revolutionary new technology (Christensen, 1997). Radical innovations are especially interesting for managers, because a radical innovation has the potential to become a disruptive innovation. Disruptive innovations create new markets with new sets of value propositions (Christensen, 1997). Radical innovation is disruptive, if the new commercialized technology meets customer needs. It makes sense to think of the radical-disruptive continuum as ranging from technologically revolutionary commercialization (i.e. Benz Patent-Motorwagen „Velo“ 1894) to game changing commercialization (i.e. Ford T model), changing the structure of a hole industry. A continuum concept between these two kinds of innovation is more realistic then thinking of them as two exclusive categories, because there are no purely radical innovations, which new technologies have no creative influence on existent markets. Commercialization itself always implies change of the market, even if only on a very small scale, i.e. by the offer of a single yet entirely new product. It makes more sense to consider disruptive innovation as building upon radical innovation on a continuum then to see them as opposed to each other. Managers, who strategize innovation along this continuum can realize great competitive advantage for their enterprise, they can either increase the technological competencies of their enterprise by radical innovation or even better, create their own monopoly structure by making use of first mover advantages in their new market by disruptive innovation. Both radical and disruptive innovation strategy are thus ways to enhance long-term growth.

With all these attractors towards radical-disruptive strategizing, what problem are managers facing when engaging in these kinds of innovation?

A major problem is uncertainty or the stochastic evolutionary nature of such innovations. Nelson & Winter (1977) suggested, that R & D projects as well as their selection are guided by a semi-stable set of search heuristics. These heuristics form innovation strategy with a probability distribution amongst different numbers and kinds of innovations. The selection process amongst these strategies is uncertain due to interaction effects with external conditions.

In other words, managers must select amongst various specifications of radical-disruptive innovation strategies and do so with bounded rationality (Simon, 1959), both environmental fit of the strategy as well as validity of the selection heuristics remain uncertain. This frames radical-disruptive innovation as risky undertaking in terms of returns of investments in R&D. Are there ways to drive innovation strategizing towards beneficial radical-disruptive patterns with less limitations from human bounded rationality?

1.2. Research Problem: design elements of an innovation forecasting software

The research problem of the study is the investigation of ways to reduce innovation strategy uncertainty by searching for patterns in customer voices on social media, which predict innovations. Social media customer voices are studied, as they are a readily available data base, which may indicate societies readiness to support an innovation.

The goal of the research can be described in analogy to the shift from weather prediction by weather proverbs to satellite data fed meteorology: By cartography of the drift of topics of societal interest within social media, cognitive readiness of consumers to ‘sprout’ a certain innovation could be modeled scientifically and replace cue based selection of R & D projects. Broadly stated, the vision behind this research is to inspire innovation strategizing by computer-supported prognoses of societal innovation-readiness, to inform strategic management decision making in analogue manner to a weather forecast adjusted choice of clothing. Ultimately, the study seeks to be the first step in a process that administers navigation amongst disjointed research directions for the development and improvement of the necessary algorithms and linkage of existent IT solutions for innovation prognoses. These prognoses start out as educated guesses but are aspired to become more and more reliable, like weather forecasts. The research is meant as a first step towards more advanced means of prognostic market intelligence, meaning a better understanding of customer needs and competitor behaviors at the same time.

1.3. Research Questions

1. What patterns discernable by computer-supported analysis are there amongst social media data, which predict radical innovations? (Study 1)
2. How could such patterns be used to predict future innovations from social media meta-data? (Study 2)

1.4. Justification for the research: Entrepreneurial benefits of prognostic market intelligence

To develop a method predicting an innovation with high chance of success has implications for entrepreneurs with an interest in future developments of customer needs and of the competitors. Here, this is called prognostic market intelligence. The implications count for both cases of predicting a forthcoming innovation as well as entirely customer voice steered R&D, meaning the prediction of an unknown innovation inspired by customer voices. Entrepreneurs with superior prognostic market intelligence for either of these two kinds of business ideas have competitive advantages over those without such knowledge. Entrepreneurs with this prognostic market intelligence also increase the chance of successful adoption by merely knowing an innovation in advance: As they believe to promote a future innovation, the prediction to innovate in the sense of setting up organizational resources to do so becomes a self-fulfilling prophecy. This effect, by which a person biases his or her own behavior by a self-prediction is known as "self-prophecy" (Sprangenberg et al. 2003). An entrepreneur believing in a statistically promising innovation can make that self-prophecy even stronger by mass-communicated marketing communications (Sprangenberg et al. 2003): The entrepreneur can actively create the target market by advertisements, which mention the scientific method used to select the innovation. These marketing communications can help reinforce the self-prophecy by persuading others by a scientific credibility of the innovations success. The mere social norm of taking science and its predictions as credible is likely to benefit the adoption of the innovation. Superior prognostic market intelligence can also have economic benefits for other entrepreneurs. A successfully predicted innovation can help to establish lots of complementary products. The predicted innovation can become a primary product which is likely to spark of other entrepreneurs to innovate complementary products. For example, Stremersch et al. (2007) found some support for the notion that such “indirect network effects” of primary products on complementary products do exist.

To sum up, managers, who think entrepreneurially and try to innovate, should be interested in the contribution of this research to the further advancement of prognostic market intelligence.

1.5. Theoretical domains

Addressing the research problem complements the literature of big data studies trying to predict economic events. An up-to-date literature review of 52 big data studies by Kalampokis, Tambouris & Tarabanis (2013) suggests that a number of studies dealt with the predictive power of social media data for economic events or topics like inflation (Guzman, 2011), feature films revenues (Goel et al., 2010), automobile sales and consumer confidence (Varian, 2011), future house prices and sales (Wu & Brynjolfsson, 2009) or Amazon sales rank spikes (Gruhl et al., 2005). None of these studies has dealt with the prediction of innovation based on social media data in particular, suggesting the need for further scientific inquiry in this area.

In a related research also studying the use of social media data for future predictions, Ten Thij (2013) found that the evolution of network patterns amongst Twitter messages predicts future topics on Twitter. These findings suggest, that using the network perspective upon social media data is a fruitful direction of theory development for social media data based future prognosis. Network analysis software is thus chosen as computational backbone of the study.

1.6. Outline of the report

The literature review describes the relationships amongst the network theoretical frameworks employed in the search for patterns, which could predict innovations. Ways are shown to suit the network framework and the here handled RQ tailored-definition of networks is given, followed by an excurse for network analysis background knowledge, which is necessary for understanding the hypotheses following thereafter.

The methodology section describes how innovation predicting patterns were detected in study one, study two describes their application to forecast future innovations. How the data of these methodologies were analyzed is presented under data analysis. Discussion of the predictive patterns and the made predictions are the key topics of the last parts of the report.

1.7.Delimitations of scope: Envisioning spherical data from social media network data

Monitoring tools can be used to identify interdependent contents in social media. The wire frame for such interwoven topic nests can be called social media spheres, inspired by Levin's (1972) spheres of influence in social networks. Applying Levin's (1972) smallest space analysis on network structures in social media contents, it becomes possible to craft a simplified network representation. This representation captures the major outlines of a network structure of social media contents and is a great tool to provide overview on the network structures and to contextualize the findings of studies like the current one. However, creating these overviews is outside of the scope of this study, delimiting the focus of the study.

A sphere of influence in social media contents is a sector or cluster of concepts with a specified number of shared links, which is put in relation to all other linked concepts on a map. A complete spherical map is an approximation of Zeitgeist at a distinct timeframe. This perspective affords to interpret meta-data on social media contents as signature of online Zeitgeist or transforming all meta-data to a comprehensive answer on the question: What is going on in social media? An analogue for the relation between this question and a spherical map is similar to as the one between the following: What is the current weather? A weather map updated with big data from satellites. A spherical web of social media contents should just like a weather map be created by a computer program to be developed next to this research (another delimitation).

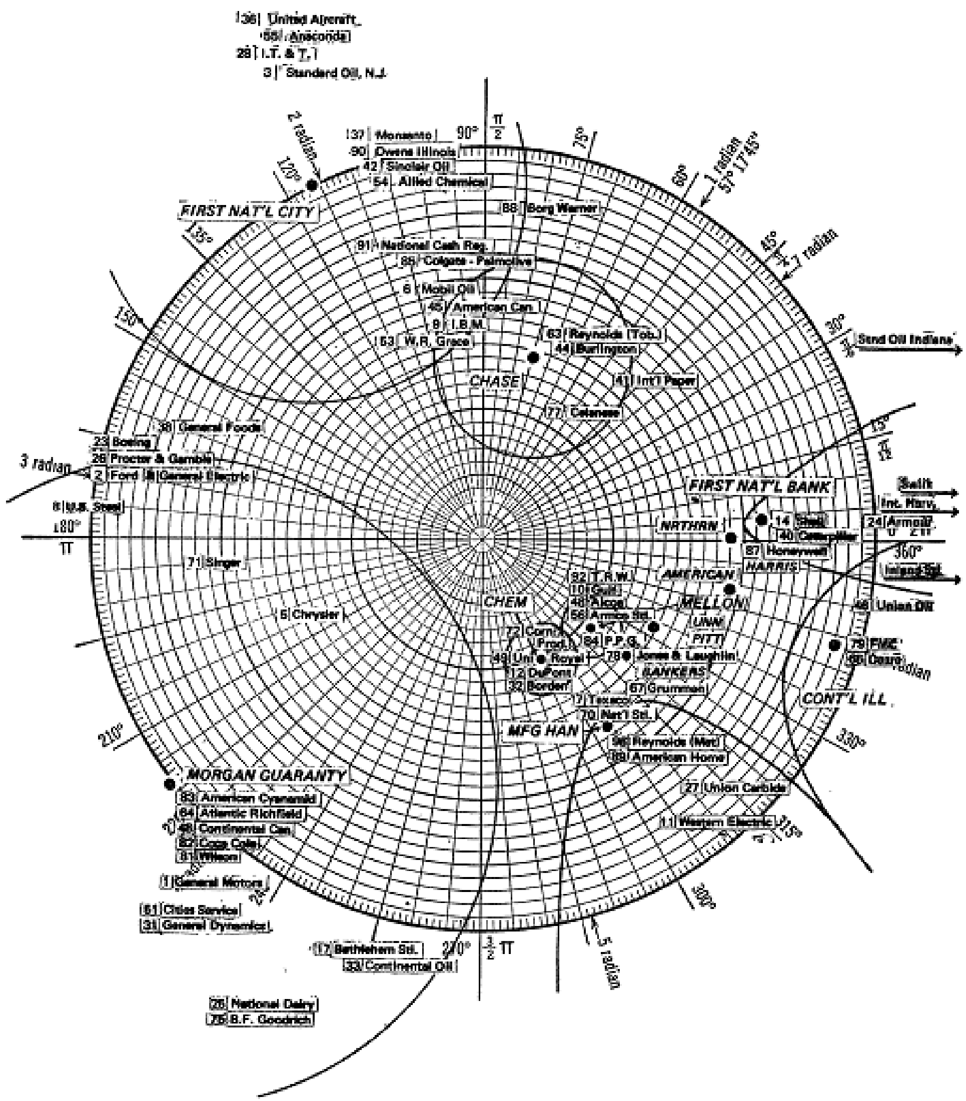


Figure 1 from Levin (1972) Map of a sphere of influence

This example shows spheres of influence of some industrials and banks in a social network. Cartography of spheres of influence in content networks on social media works analogue and describes the reach and relative position of individual buzz-clusters (curved lines). The result is a spherical map alike to the one above, which positions the spheres within a defined coordinate system. The spherical map is like a standardized semantic space, which can represent network patterns in social media data. These approaches to simplifying network representations could be used to contextualize network structures as suggested by the current studies in a way, which is better accessible for visual inspection. A big problem with network data generated from big data is the extreme density of such networks. The 'blob' of Facebook network structures shown at 5:15 of an intriguing lecture on challenges and opportunities for statistical inferences from network data by Neville (2011, see http://www.youtube.com/watch?v=1xLjYc7EUEU) makes clear that studies like this one need techniques for making the data visually interpretable, the concepts of Levine (1972) may help to contextualize network data.

Each sphere is a part the overall buzz-network pattern, which can be brought together in maps like the one above.

This study tries to identify patterns in social media sphere evolutions, which predict innovations. As weather charts are observed over time and stochastic models are built to make future weather predictions and climate change predictions from lots of weather predictions, the current study is meant to envision ways to construe a Zeitgeist prediction from social media spheres followed by a prediction on future innovations from lots of Zeitgeist predictions. Google Zeitgeist is an example for the Zeitgeist-online buzz metrics link: Google Zeitgeist is a yearly review published by Google, which shows topics of large societal interest and influence by means of i.e. the development of search queries on words related to those topics (i.e. queries on the controversial Pussy Riot band in 2012).

The research questions are studied here for a more direct link then the link between Zeitgeist, their representations on spherical maps and innovative constellations of these two. The research questions deal with small network structures as predicting innovations. Coming up with Zeitgeist predictions from spherical map constitutions or even setting up a single spherical map from network data from social media require much more developmental work then feasible in the thesis. Instead, the predictive power of small clusters of online contents is studied here, further delimiting the focus of this study.

1.8.Conclusion

A network analysis approach is taken in this study for computer supported analysis of social media data, which is studied for its innovation prediction possibilities. It is shown, that this is a theoretically interesting approach towards analyzing social media data, which may afford the offspring of new generations of social media data analysis methodologies, even beyond the scope of this research.

All in all, searching for design elements of an innovation prognosis software in network analysis software domains was chosen for theoretical reasons as well as out of personal interest in social network theory.

2. Research Issues

2.1. Introduction

This study focuses on singular network structures or in other words on individual buzz networks (one field of influence between buzz topics, later introduced as trending topics). Networks are regular patterns of information exchange (Haythornthwaite, 1996), the information is represented as nodes and exchange relationships are represented as connectors or links between the nodes. The activity of individual buzz-topics (as nodes) and their changing associations (as connectors) are measured over time. This results in a dynamic network structure. Network nodes can be assigned to have arbitrary attributes (i.e. a color or an numeric quantity of something).

These dynamic networks are chosen as theoretical focal point. The buzz networks are observed over time. The study seeks to address the research question, if network structure and network attribute co-evolution predict innovation in social media contents.

The network structure between buzz-topics of this study is defined like this: The length of the links depends on proportional relationships between the activity on topics, which tend to co-occur within some timeframe. These topics are called trending topics, they indicate the membership in a common network. Both are measurable with data from the social media monitoring service Coosto.nl.

Defining these networks can be suited to the researchers needs by answering the question:

What kind of relationships in the data am I interested in?

The relationships in the data investigated in the current study were less on the network structural change but more on a behavioral-coevolution (explained in the next paragraphs) within the network structure. The reason for this choice was that although the structural change is per se an interesting field of research, it is hard to define what structural change may be responsible for emergence or dissolution of trending-topic connections without knowing the trending-topic algorithm. As the research seeks to study patterns, which participate in the arise of innovations, observing the interest of people in the networked topics (indicated by the amount of response only messages) seemed as a straightforward way to show the use of network analysis to predict innovations: It is straightforward to expect that topics, which are connected by being each-others trending topics, may have co-evolving patterns of interest. For example, the same people who read about topics being mentioned together on the same kind of websites (i.e. car magazines) may show similar interest in both topics. A test of a new Volvo (having the image of being safe) and the results of the NCAP crash tests may arouse similar interest curves, depending upon their current need for safety (i.e. increasing after some large accident has been reported in the media). Such patterns of co-evolving interest amongst trending topic-topics are the here handled definition of buzz-networks, more specifically, research question 1 could be specified like this: Do co-evolution patterns amongst the interest in a trending topic network predict innovations and what kinds of co-evolution patterns do so? Trending topics are topics, which are found to be mentioned together in social media by a social media monitoring service called Coosto.

It should be noted here, that the network structural change could be studied as well, however, a theoretically interesting mathematical relationship between two contents must first be defined. In other words, a huge number of more specific research questions can be answered by buzz-network analysis. The definition of the network then follows the particular research question. For example, an more mathematical, alternative specification of a research question to look for patterns predicting innovations could go like this. Assume that a certain threshold of a Google Similarity distance between an topic and the topic 'innovation' could be found to correlate with the process of innovation. Do patterns of the evolution of a trending topic structure linked by the clothing-in towards the threshold value predict the innovation?

In the current, initial studies, things have been kept simple by focusing on a behavioral co-evolution (response-only messages of a topic) within an evidenced network structure (trending topics, which are found to be related to each other with Coosto.nl). But the use of network analysis to study innovation is only limited by the imagination of the researcher, testing relationships between innovation predicting concepts. Network analysis does not have to be on relationships between humans, but may deal with much more artificial relationships, too (i.e. relationships in particle physics, Wikipedia, 2013)).

The theoretical background of this study is network analysis theory, which is explained in detail in the next paragraphs.

2.2. Necessary excurse on parent theories to indicate assumptions made in the research problem theory and hypotheses: Statistical analysis of network evolution

It is considered a necessity to inform the reader about the employed parent theories in more detail for a number of good reasons. Only by this excurse, the reader can be given a chance to critically appraise some of the subtle assumptions made in this research, without it, the research would not be transparent. This excurse is also mandatory to establish the terminology, upon which the research problem theories and hypotheses are based. The information presented in this chapter is furthermore necessary to make the research replicable on some of the technically more advanced hypotheses. The reader is advised that *inferential statistics in network analysis* are not an easy subject matter. A fellow researcher comments: "If doing statistical network analysis was easy, then everyone would do it" (Keegan, 2013). The complex nature of inferential statistics in network analysis makes a brief discussion of the parent theories not viable, doing so would sacrifice scientific rigor.

2.2.1. Basic model of network evolution

In this research and in Visone as dynamic network analysis (DNA) tool, network evolution patterns are understood within the actor-oriented model of network evolution. The actor-oriented model of network-behavior co-evolution (Burk, Steglich & Snijders, 2007) suggests that each actor or node ‘decides’ upon their outgoing network ties in order to optimize his / hers / its position in the network to meet short term preferences and constraints and due to some residual unknown element (random deviation). Each decision changes both the network ties as well as behavioral variables in response to the current network structure and the behavior of other actors. The actor-oriented model as used in Visone software serves to reconstruct *likely* trajectories of network evolution between the made observations (Burk, Steglich & Snijders, 2007).

2.2.2. Software implementation of the network evolution model

This introduction is on the essentials of Burk, Steglich & Snijders, 2007; Ripley, Snijders & Preciado, 2011; Snijders, van de Bunt & Steglich, 2010:

A factually observed network evolution, as depicted in adjacency matrices like the following one, is observed for changes between each observation.

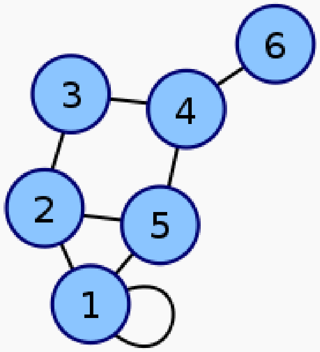
Labeled graph	Adjacency matrix
	$\begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$ Coordinates are 1-6.

Figure 2 Adjacency matrices (Wikipedia commons pictures, 30.12.2013)

The model of the network evolution can look solely at the changes of network ties from ‘snapshot’ to ‘snapshot’ and at changes of a network attribute (which may be numerical like the later introduced normalized buzz-force) from snapshot-period to snapshot-period.

Such a network attribute, if it is observed for all nodes at all times, is called a network behavior (Burk, Steglich & Snijders, 2007). In social network analysis, observing the relationships between people and simultaneously measuring the smoking frequency of each participant is an example for a network-behavior co-evolution. One can observe, in how far social relationships and smoking co-evolve in such a network. Here, the co-evolution of topics related to each other as being each others trending topics and the later introduced normalized buzz-force is observed. Again, the networks evolution is modeled as change rates of either the network ties (as vanishing or emerging or staying stable) or of some network-behavior (change between measurements, i.e. increase in smoking frequency).

2.2.3. Software estimation of network evolution

Next, the here employed software packages need to estimate a continuous version of the network evolution (from the snapshots), this allows to test the network evolution for tendencies as well as the network-behavior co-evolution for tendencies. To estimate network evolution, a huge number of micro-steps in between the observations or snapshots is created (see for more background upon all of this Burk, Steglich & Snijders, 2007). The micro-steps are necessary to construe a network evolution with only one little change at a time (as a continuous-time Markov chain), for which statistical testing of network evolution is possible.

In figure 2, ties between node 2 and 5 and 5 and 4 may emerge or vanish between two snapshots, so several micro-steps are construed with only one change at a time (i.e. first 2 and 5 emerges, then 5 and 4). The estimation algorithm in the software here of course runs a lot of estimations to construe a 'likely' trajectory of network evolution, by statistically checking the construed evolution for convergence (or significant differences) with the factually observed snapshots at their respective periods.

2.2.4. How the network evolution simulations are tested for effects

This continuous-time Markov chain construal of network evolution can be tested statistically (by the software). It can be tested for effects of network structural tendencies (i.e. out-degree effect as the tendency of network nodes to have outgoing connections to other participants, aka network density), as well as structural change tendencies (i.e. effects tested with the software network creation and endowment functions) or network-behavior co-evolution (i.e. average similarity effects, introduced later on).

In the next step, the construal of network and network-behavior construal can be tested for these effects. A model of effects, which are expected upon theoretical grounds then has to be specified and tested. Ripley, Snijders & Preciado (2011) advice to start with a simple model, testing it and adding effects and retesting it. In case of lacking knowledge / expectations of the network structure effects, effects which help to capture network evolution tendencies may also be chosen on a more inductive basis (Snijders, van de Bunt & Steglich, 2010), i.e. based on a good match of the model with the data as signified by significant effects of those effects as well as their tendency to increase the significance of the other effects of the model.

The significance of an effect indicates, if the simulated network evolution provides evidence for the presence of the effect. For this the following is done: The networks evolution has been simulated with 'virtual' micro-steps between the 'real' or observed values. Only these virtual 'redrawings' of the network evolution can be tested for statistical tendencies or change tendencies.

By selecting a set of theoretically expected tendencies, a model is specified, which may or may not capture the change tendencies of the simulated evolution (representing the observed evolution). For each change tendency within this model, the null hypothesis that the parameter of the change tendency is 0 (it is not present) is tested, rejection of the null hypothesis means evidence for the presence of the effect. Significant effects within the model are evidence, that the observed network evolution follows a trajectory, which is ‘biased’ towards the tendency over the whole course of evolution.

For example, a significant density effect means that over the whole of the simulated micro-step changes of vanishing and emerging ties shows a tendency of participants to have outgoing connections to other participants. Significant density effects are basic evidence for the presence of a network structure over the period at all.

2.2.5. Advanced period-by-period analysis of network evolution

Next to the models on probabilistic network evolution (the tendency for effect X) and the evidence from them for some network tendency in common language, the networks rate change data between the snapshots can also be reviewed for analysis. The software allows to ‘catch’ the algorithms estimates of rate change between snapshots (of the network-evolution and the network-behavior evolution) and their standard errors (amongst other data). The estimated number of opportunities for behavior change (also called behavioral rate change) between two snapshots can be used for comparative analysis of period bound behavior evolution between independent networks.

2.3.1. Research problem theory for model specification: Non-randomness in Buzz Network Evolution

It is important to define a theoretically motivated point of departure in the definition of a network evolution model, as the selection of included effects should be guided by theoretical knowledge (Snijders, van de Bunt & Steglich, 2010).

What theoretical frameworks can motivate assumptions on buzz network evolution patterns and help in making a choice amongst the large number of available testing options? This and the next few paragraphs serve to answer this question.

Information routing groups (IRGs) are a group of people with a common interest (i.e. 50) who engage in lateral communication (communication between individuals trespassing institutional or hierarchical boundaries) for mutual benefit via a computer network (Andrews, 1984).

Practically all social media (or Web2.0 applications), as a collection of open source, interactive and user-controlled applications supporting the creation of informal user's networks to facilitate the flow of knowledge amongst its members (Constantinides & Fountain, 2008), are lateral media. Lateral media are structures supporting lateral communication (Masternewmedia, 2006). An informal example are gossip groups and a formal example are IRGs. IRGs play the role of the coordinating mechanism, distributing relevant information each social media user:

An updated, more technical definition describes an IRG as one of "semi-infinite set of interlocking and overlapping groups containing individuals who use software and email to automatically mediate and exchange information via lateral communication. Due to the principle of six degrees of separation, a specific message is highly likely to meet *any relevant but unknown target* by the process of lateral diffusion" (Masternewmedia, 2006).

This means, that information is passed on to individuals, who are likely to have use of the information, although no one ever has to know the emergence or members of the chain of people passing on the information. Next to this, groups, for example IRGs, are an emergent property of social networks (Haythornthwaite, 1996). Hence, some social networks in social media have the special characteristic of information intelligence, some networks or IRGs within 'know' how and where to channel relevant information. Relevant for this information intelligence, Andrews (1984) suggested that IRGs (emergent from those networks) gradually become an entity with a large body of tacit knowledge, a rich and well-integrated information exchange group, like a co-operative brain working by cross-fertilizing conversations.

To sum up, the information intelligence imposed on social media users should result in selective and well-governed control of activity in buzz-networks. Patterns of co-evolution in the here studied trending topics networks are likely to occur, because topics, which are frequently mentioned together are likely to be read by people included in equivalent IRGs (i.e. news postings on the same topic on different sites of the same genre, i.e. car reviews).

2.3.2. Theory relevant to specify the network evolution model: New product category learning patterns in buzz network evolution

Information intelligence in social media may result in information processing processes, which mimic information processing of humans. Others also describe learning mechanisms in the world wide web, which remind of human learning processes, for example emerging semantics in in web-based e-learning systems (Zhuge, 2009).

The following processes are inspired by theories from human information processing like semantic network theory of memory. The Collins and Quallian (1975, see also Ashcraft, 2005) model of semantic memory suggests a network structure of semantic memory and some processes for retrieval of information from that structure. New category learning in buzz networks may work by a cascade of events similar to category learning in human semantic memory networks. First of all, the networks evolution might depend on the event of buzz-spill over (psychological terminology: “spreading activation between concepts”). Topics within a network are more or less likely to arouse responses in discussion followers. These responses are likely to induce responses on those topics, which are associated with the original buzz topic. A buzz can perpetuate itself throughout the network, it can spill over to its associated concepts. This spill over does explain two things: For a part, network evolution (i.e. how close two concepts are associated) depends on this spill over, because two associated topics, which are repeatedly under discussion at the same time are more closely associated in social media (technology and user-minds). More importantly here, the spill over can have different sources at the same time and encounter each other at a particular discussion topic due to the networks structure (“encountering of in-phase spreading activations in a semantic network”). When the same patterns of buzz-spill over encounterings repeat themselves over time (“intersection recognition in semantic networks”), then the network becomes restructured in ways, which makes future buzz-spill over more efficient. If this series of events is repeated itself in systematic, patterned manner as repeated encounterings in several concepts at once, then a new category is likely to emerge (“categorization”), which offers a more efficient pathway between all these prominent buzz transmission concepts. Such a new category can be exploited as a new product category.

2.3.3. Some Probabilistic Network Evolution Patterns are Proxies of New Category Learning

There is a theoretical shortcut for the category learning process: All necessary is in-phase buzz-spill over encountering at the same semantic nodes over a longer time. A new category should emerge if this condition is given, although it is not easily pre-specified because the emergence cascade is much more complex than the ones indicated above. This appetence for a new category is rather fuzzy and stochastic.

However, this fuzzy category learning makes it suitable for probabilistic accounts on network evolution. Some probabilistic network behavior co-evolution effects map out patterns, which make encounterings very likely if a certain network structure is assumed (more details in the following section). Observing these effects over a longer period means that a network is prepared to learn a new (product) category. This readiness can be exploited to identify a promising a new innovation with high chance of adoption, especially under the participants of the networks IRGs.

2.3.4. Stochastic Proxies of New Category Learning in Buzz Networks

New category learning, as the theoretical shortcut explained above, can be operationalized by observing stochastic network evolution effects longitudinally, if an egocentric network structure is used to model the network. An egocentric network structure is "a picture of a typical actor in any particular environment and show how many ties individual actors have to others, what types of ties they maintain, and what kind of information they give to and receive from others in their network." (Haythornthwaite, 1996)

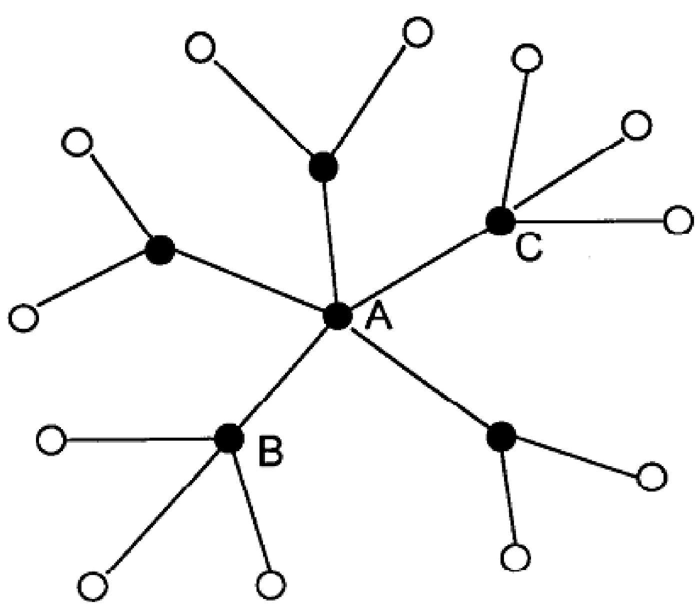


Figure 3 Centralized egocentric network (from Haythornthwaite, 1996), the buzz topic is positioned at A

A group of network-behavior coevolution effects of such an egocentric network centralized around the buzz topic (i.e. innovation) imply repeated encountering of buzz spill over (or interest-intensity changes) in the buzz topic:

Interest-intensity, which co-evolves for all nodes in an egocentric layout, means that waves of interest-intensity changes move throughout the network (collective rise and fall). These waves encounter in the ego node due to the network's egocentric layout. Network behavioral effects, which can indicate interest-intensity co-evolution and by extension, new category learning, are (based on Snijders et al., 2007):

Average similarity effect. This effect expresses the preference of the focal node (alpha-node) to have similar scores compared to the average similarity score of its attached nodes (i.e. like B and C). In aboves egocentric setup with one focal node only, average similarity effects mean that the buzz-spill over evolution of the centralized buzz topic A is probabilistically similar to the average buzz-spill over evolution of all individually captured associated topics.

Total similarity effect. Total similarity of buzz-spill over indicates, that the focal node showed the tendency to have similar scores compared to the sum of spill-over similarity scores of the associated nodes. Total similarity of buzz-spill over means, that the buzz-spillover evolution of the centralized buzz topic A is probabilistically similar to the collective buzz-spillover evolution of the complete topic nest around A.

Average alter effect. If the focal nodes average values of the spill-over behavior increases, then associated nodes average values also show the tendency to have higher values as well. Average alter effects of buzz-spillover behavior means that the centralized buzz topics spill over was likely to co-evolve with the spillover of the associated topics.

2.3.5. Model Specification to capture innovative new category learning

Buzz-spill over encountering, which is critical for new category emergence, is implied by a model of network evolution with these three effects: Buzz-spill over of A, which tends to evolve similar with all individual topics associated to the central topic A and in the whole nest of A's associated topics and which is likely to co-evolve upwards and downwards with its associated topics is a spill over evolution, which has likely to indicate real spill over co-evolution in the network. A model, which combines these three effects is likely to indicate a probable co-evolution of spill-over, which comes from the networks real, historical co-evolution of spill-over. Evidence for historical co-evolution of spill-over over a longer time (i.e. a year) means, that the spill-over of the cluster was likely to encounter each other in the centralized buzz-topic (i.e. innovation, in A) and that the encountering event was likely to be repeated frequently.

A model with these three effects with significance of these effects is evidence for repeated spill-over encountering in the centralized buzz-topic and suggests, that the centralized buzz-topic emerged from the spill-over encounterings of its associated topics. This is especially true if cases of historically known new category emergences (i.e. new innovations) are compared to cases without new category emergences (i.e. traditional products).

A longitudinal observation of these network behavior effects operationalizes the new category learning process in buzz-networks. In conclusion, the model specified to capture innovative network evolution is theoretically motivated to concentrate on average similarity, total similarity and average alter effects of interest-intensity as network behavior.

2.4.1. Research problem theory on pattern expectancies: Hypotheses on archetypes of innovative customer voice dynamics

The current research studies how buzz networks and how the tendency of each topic (or node) to arouse responses in social media (as network behavior) co-evolve. There are a number of pattern expectations. These pattern expectations are on the interest-intensity behavior of network-structured customer-voice in social media, i.e. interest-intensity is taken as the main indicator of customer voice and it hypothesized to predict radical innovations by the following relative patterns. The predictive pattern are always relative in the sense, that they mean that an innovation becomes more radical than the ones of the to be compared with, traditional product sample. Some radicalness may still be residual in the traditional product sample, innovations with lower radicalness then the traditional product sample are thus not predictable by this method.

All patterns are of rather abstract nature, an attempt to make them more graspable is made in the diagrams depicted under each hypothesis. All example evolutions depicted below mean classes or archetypes of buzz-word interest evolution. These examples serve as representatives of a whole bunch of possible evolutions, which fit in the probabilistic-archetypical evolution categories. With other words, the depicted example evolutions are only potential realizations of the patterns and not the patterns themselves. A further complication is, that the predictive patterns mean the presence of relatively more archetypical evolutions, they mean that the categories of the example evolutions below are more likely to occur in the more innovative sample then in a less innovative or traditional product sample of customer voice.

Next to these example evolutions, each hypothesis is accompanied by a figure showing the proposed causal relationships. An overview of the hypothesized innovation predictors:

1. Higher probability of interest-intensity-oscillation in more innovative buzz
2. Higher prevalence of co-evolution in interest-intensity
3. Higher volatility of interest-intensity rate changes
4. Higher prevalence of cross-network interest-intensity rate change co-evolution
5. Period specific differences between cross-network interest rate change averages

5.4.2. Hypothesis 1: Innovation related buzz networks should exhibit a lower tendency towards interest-intensity behavior subjecting itself to a negative feedback then non-innovative buzz networks because a natural difference between their source of interest-intensity.

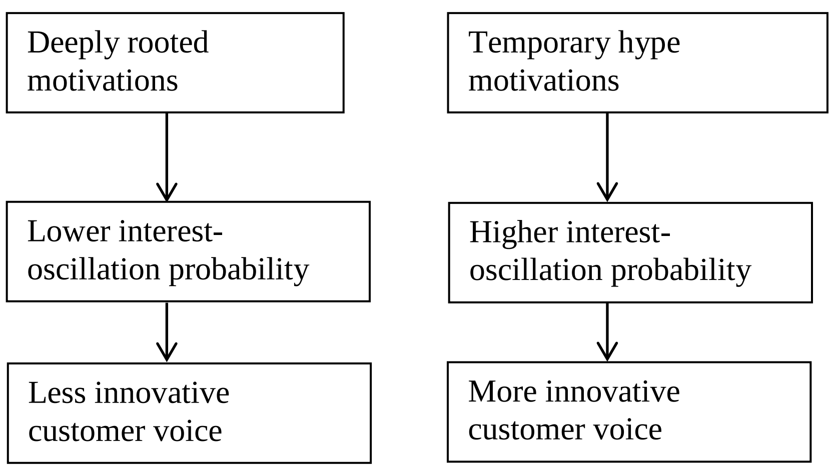


Figure 4: Innovation Predictor-Customer Voice relationship in hypothesis one

A negative quadratic shape effect (for more details see Snijders, Van de Bunt & Steglich, 2010) of interest-intensity behavior means, that interest-intensity subjects itself to a negative feedback: The more interest-intensity tends to increase, the less becomes the ‘push’ to get even higher or the more the interest-intensity decreases, the less becomes the ‘push’ to get even lower. In the long run, the negative quadratic shape effect of interest-intensity means that interest tends to oscillate with the same acceleration and deceleration pattern around some value like a pendulum around the point closest to earth. The figure below shows an example of a negative quadratic shape effect function. Here, interest-intensity would decelerate fiercely to get higher at a value of 3 and decelerate to get lower below 1:

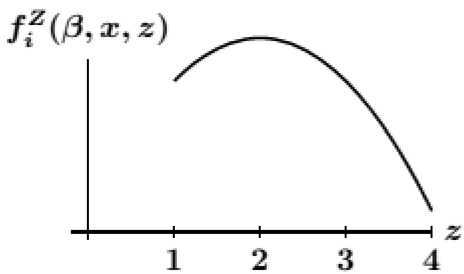


Figure 5: Negative quadratic shape effect (from Snijders, van de Bunt & Steglich, 2010)

This is still a rather abstract representation. It means, that each of the network nodes or buzz-word interest-intensity has a higher probability to decrease above the value of 2 and a higher probability to increase below the value of 2. The following figure shows a short, fitting course of interest-intensity co-evolution, it is a dramatization:

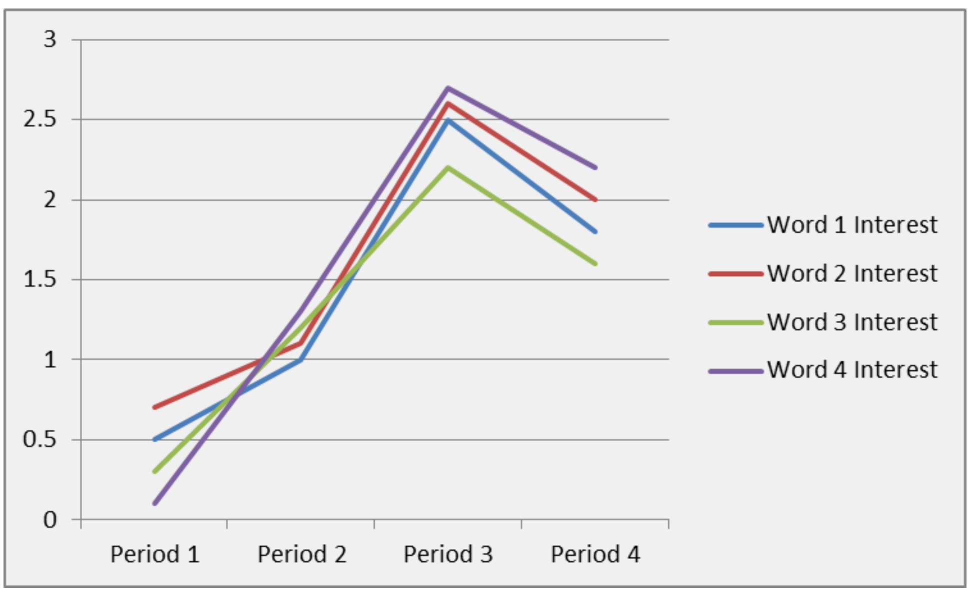


Figure 6: Interest-Intensity evolution example in compliance with negative quadratic shape effect

In this example, the probability of interest in any buzz-word to rise below the value of 2 is higher: The interest intensity curves for each buzz-word increases from period 1 to period 2. However, the probability of interest in any buzz-word to fall above the value of two is also given: Interest-intensity for each buzz-word decreases from period 3 to period 4, as soon as the interest value of 2 was exceeded by each word. This is a dramatization, as it would be sufficient if for example only three of the four interest lines decrease together above the value of two or increase below 2. Also, the strength of the ‘pendulum movement’ of all words taken together around the value of 2 may be much weaker.

Innovative products do not arouse a range of interest-intensity as strictly limited as the one of less-innovative products, because interest in more traditional products comes from deeply rooted cultural values and habits, which may be more or less activated during a period but do not vanish completely or come out of nothing. Traditional products are unlikely to capture peoples’ attention by a storm. The source of interest-intensity change of innovation related buzz is likely to be be sourced by much more temporary trends or fashions, for example think of the hype prior to the launch of the Ipad.

Both the arguable propensity of innovation buzz to be sourced by hype and the propensity of traditional product buzz to be sourced from well-established, ‘settled’ motivations for interest (values) give reason to expect less tendency for negative of interest-intensity on itself in innovation buzz then in buzz on traditional products.

A final general point on the abstraction of all patterns described here needs to be made. The example evolution above is one of many possible examples, which fit with the pattern. To show that the example may look very different, the following purposefully awkward drawing is made. It is still a relatively simple, turnaround and mirrored variant of the pattern, more difficult alternatives are also possible. It shows that factual evolutions can be very different from one another and still fit the same pattern:

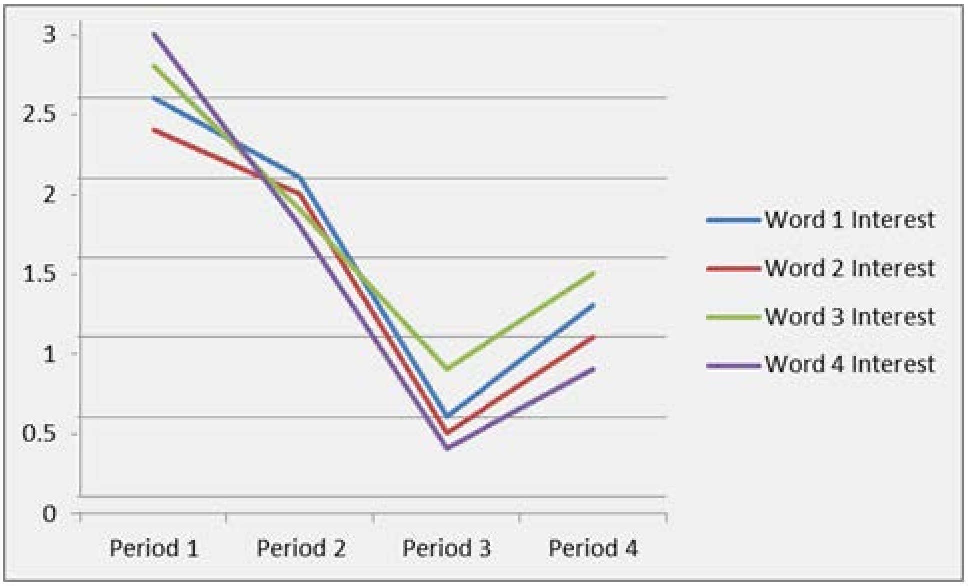


Figure 7: Simple alternative example evolution in compliance with quadratic shape effect

5.4.3. *Hypothesis 2: Innovation buzz network interest-intensity has higher tendency for co-evolution patterns than the interest-intensity in traditional product buzz networks.*

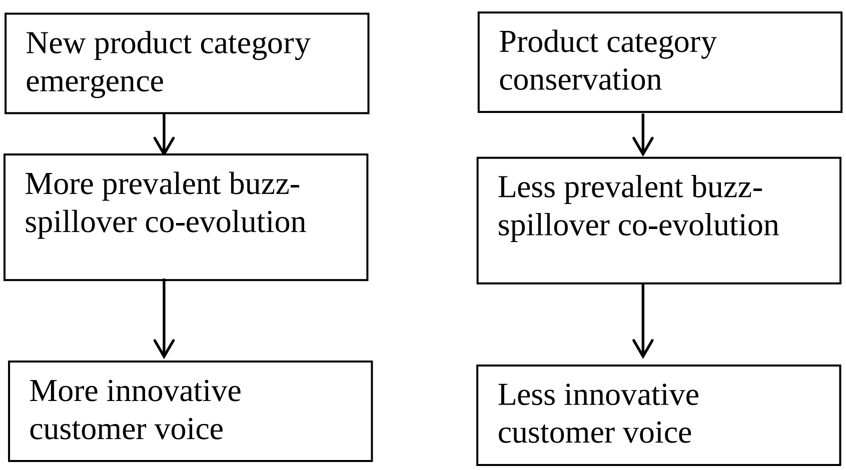


Figure 8: Innovation Predictor-Customer Voice relationship in hypothesis two

Indicators of interest-intensity behavioral co-evolution are: average similarity effects, total similarity effects and average alter effects.

These effects build on quadratic shape effects and mean more specific example evolutions.

Similarity effects mean that interest in buzz word tends to be on similar levels amongst the buzz words:

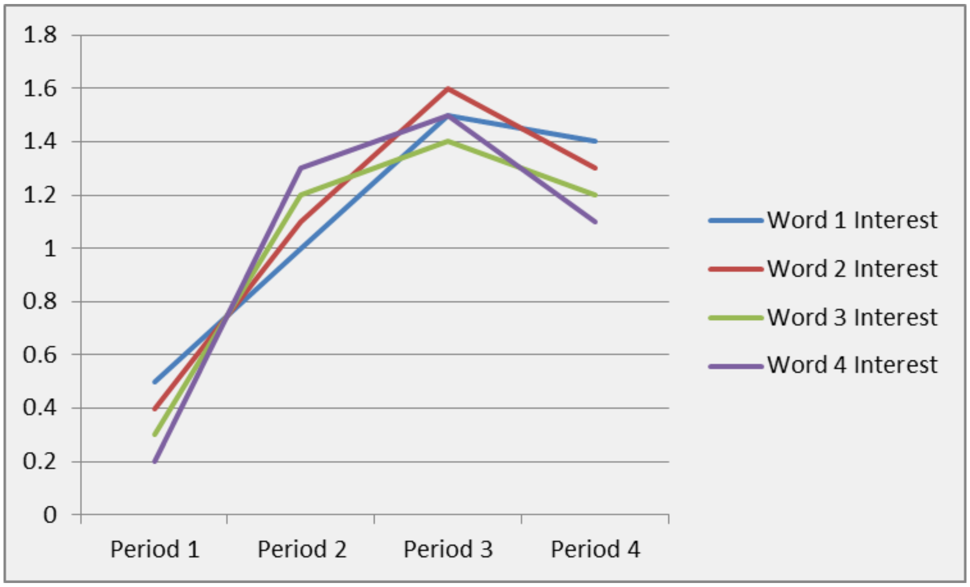


Figure 9: Similarity effects present in interest evolution

Here, one can see that interest in the buzz words is likely to be on similar levels for each buzz word at one period in time, i.e. they all spread somewhere around 0.4 at period 1 or around 1.1 at period 2. Again this is a dramatization, three of them being similar would be sufficient as well. Average alter effects mean that interest in a buzz word, which has a high average value of interest over the whole course also has a stronger tendency towards higher values of interest:

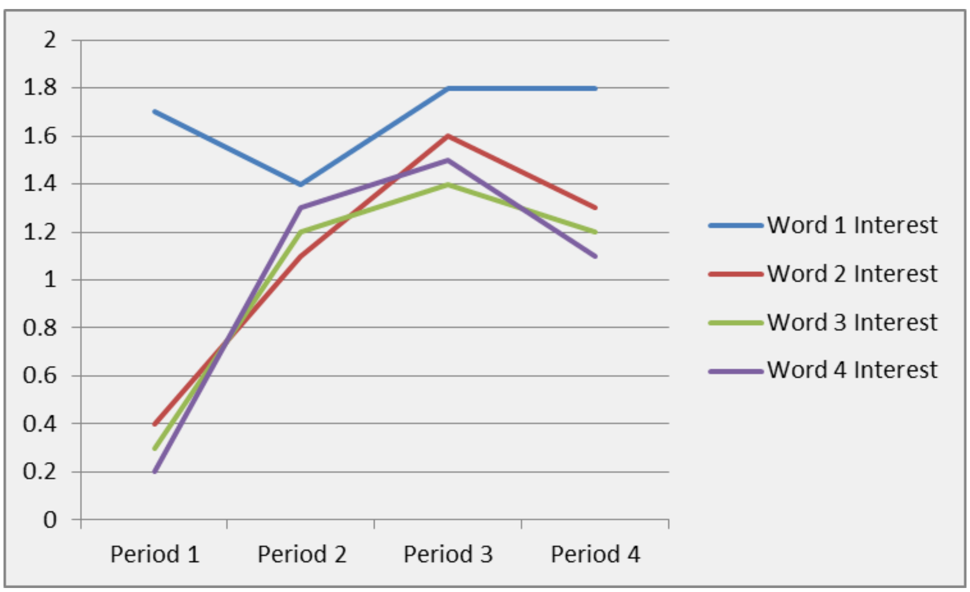


Figure 10: Average alter effects example evolution

Here, the interest in word 1, which is higher then the average interest over the whole time, also tends to has a stronger tendency towards a higher value of interest intensity at each period. This is again a dramatization, higher values at 3 of the periods would have been sufficient, too.
The reason to expect such cross-topic interest-intensity co-evolution to be more prevalent under innovative product buzz is, that for a radical innovation to arise, it needs to establish a new product category. The new product category emergence mechanism can be approximated by long term cross-topic buzz force fertilization, as reasoned upon in detail in the chapter on the new product category emergence mechanism. Cross-topic interest-intensity 'fertilization' has been called buzz-spillover in that chapter, which is shorter but less precise: Buzz-spillover could be confused with any of the other social media network metrics introduced below, buzz-force is only one of them, here buzz-force spillover is meant, not for example buzz-voltage spillover (also an interesting phenomenon).

5.4.4. Hypothesis 3:Average variance of innovation-network's interest-intensity rate changes (per period) is higher then the average variance of non-innovation network's interest-intensity rate changes.

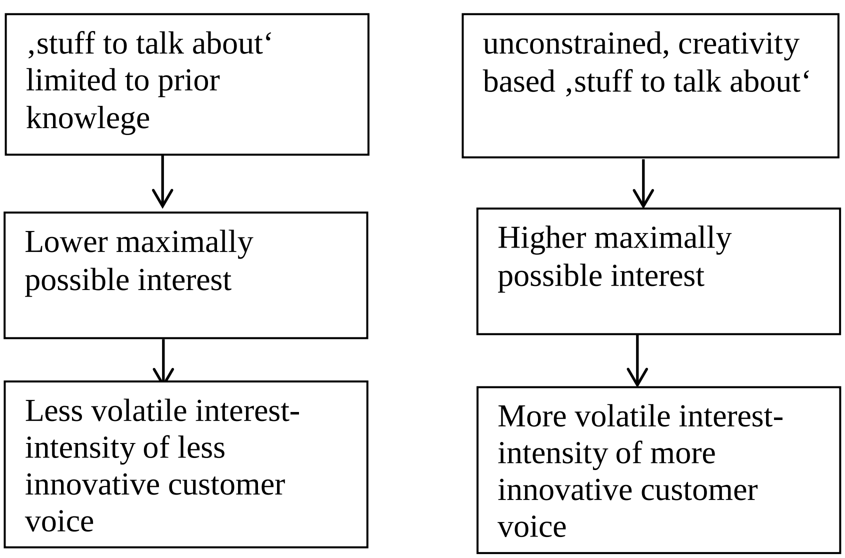


Figure 11 Innovation Predictor-Customer Voice relationship hypothesis three

This means that in innovation related buzz networks, buzz-topics tend to arouse more volatile interest-rate changes per period then in non-innovation related buzz-networks. The patterns abstraction does not make it a good example for other illustration then tabular depiction:

in-between period	rate changes Clipit Net	rate changes Coosto Net	rate changes Dropbox Net	period rate changes average	period rate changes standard deviation	**period rate changes variance**	**Cross-Period rate changes variance average**
Period 1	23.2971	3.4813	15.345	14.0411333	9.97203759	99.4415336	703.90918
Period 2	3.1959	8.0808	82.4251	31.2339333	44.4000813	1971.36722	
Period 3	2.4194	11.8939	14.6019	9.6384	6.3967794	40.9187868	

Table 1 cross-period rate changes variance average

To state it completely, the cross-period average of the variance of the cross-network rate changes has been compared between the innovation-buzz meta network and the traditional product-buzz meta network. This pattern is expected for the same reasons as the pattern of hypotheses 4 is expected, which theoretical discussion is better placed after that patterns description. In short, the here described variance average of the innovative meta-network is expected to be higher, because the pre-determinedness of innovative buzz is lower then that of traditional product buzz. Argumentation for the theory above is the same as the one discussed under hypotheses 4, which is a basic pattern for innovation-buzz.

5.4.5. Hypothesis 4: The interest-intensity rate changes per period across all innovation buzz-networks should tend more towards co-evolution then the interest-intensity rate changes per period across all networks of traditional products.

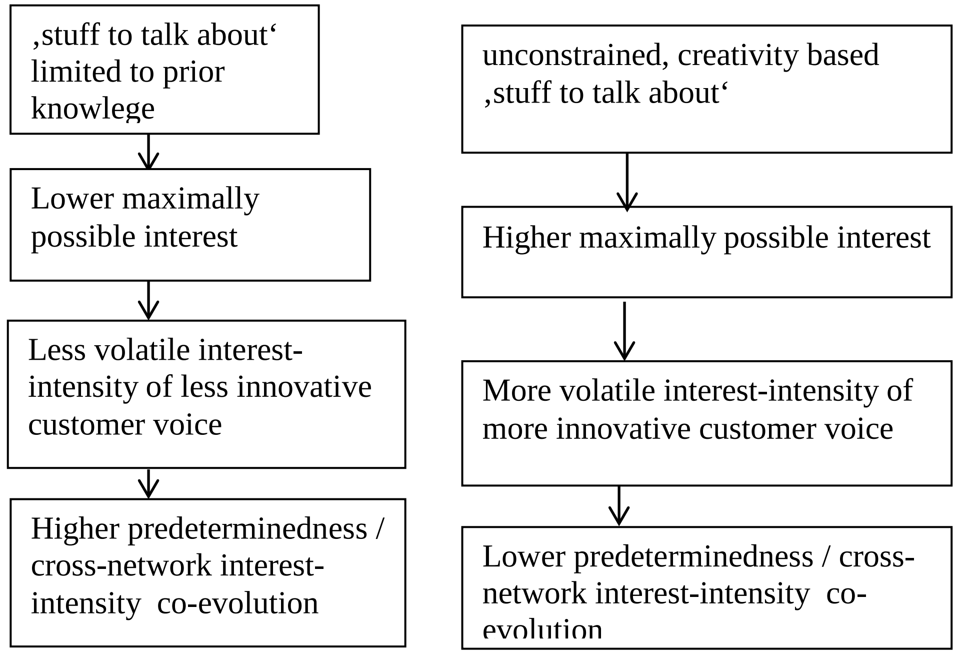


Figure 12 – Innovation predictor-customer voice relationship hypothesis four

Description of the pattern

With other words, in a meta-network of all studied innovation buzz networks, the normalized buzz-force rate changes is expected to tend to co-evolve more frequently then the normalized buzz-force rate changes of all studied traditional product buzz networks. Before giving the reasons for this expectation, a tabular depiction of what is meant seems more communicative then words. In-between period 1 refers to the rate changes from week 1 – week 2 (or observed period 1 – observed period 2). Each network refers to one of the studied innovation buzz networks or one of the studied traditional product networks. It is helpful to think of this analysis as applying network-analysis to do meta-analysis on all here undertaken 62 network-analysis studies (30 innovative, 32 traditional).

Table 2 Meta-network analysis data base: interest-intensity rate changes per network as new network behavior of the innovative and traditional meta-network

	Innovative normalized buzz-force rate changes					Traditional Product normalized buzz-force rate changes		
In-between Periods	Clipit Net	Coosto Net	Dropbox Net		In-between Periods	Amstel Net	Bacardi Net	Bavaria Net
Period 1	23.2971	3.4813	15.345		Period 1	1.17	32.5562	7.5247
Period 2	3.1959	8.0808	82.4251		Period 2	2.1524	24.416	26.0567
Period 3	2.4194	11.8939	14.6019		Period 3	1.3817	33.7477	62.9319
Period 4	2.6564	11.2427	106.6521		Period 4	1.0914	14.5366	53.0604
Period 5	9.3509	3.7502	3.8441		Period 5	5.7165	54.5254	14.8766

In hypotheses 4, it is proposed, that the meta-networks behavior (normalized buzz-force rate changes rate changes) shows higher tendency to co-evolve in the meta-network of innovative buzz than in the traditional product meta-network (each network becomes a node in this meta network, i.e. Clipit Network refers to Clipit node in the analysis of the innovation meta-network). Referring to the table above, this means that, by average, the rate changes of interest-intensity tend to co-vary between the networks (indicators: average similarity effects, total similarity effects, average alter effects).

A graphical representation may be more clear:

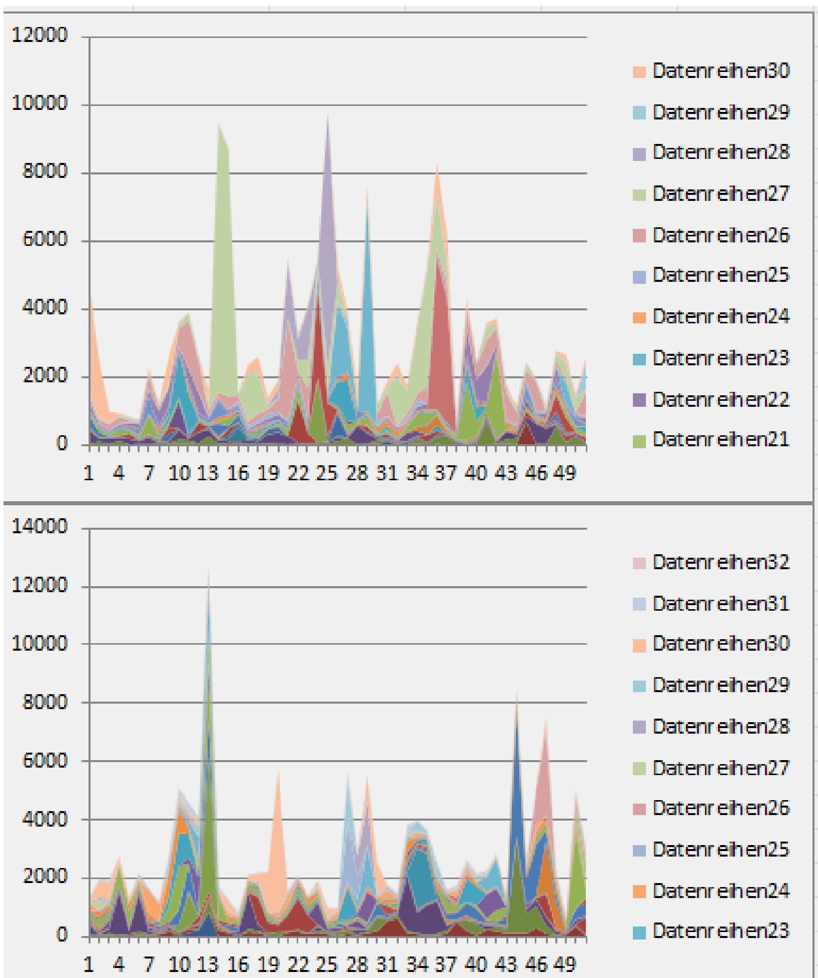


Figure 13 Meta-network interest-intensity rate changes (innovative above, traditional below)

Hypotheses 4 proposes, that by average (thus abstracted for all periods / beyond all 52 weeks observations), these lines have to put it in actor-oriented model of network evolution language a higher propensity to 'decide' to go up and down together from period to period in the innovative meta-network then in the traditional product network. Condensing the meaning of this in one sentence: It is proposed, that the ‘willingness’ of interest-intensity to have a co-evolving co-evolution pattern across a set of related networks is higher for innovative topics then for traditional product topics. To put it more simple, it is proposed, that innovation spheres, as a logical conglomerate of innovation networks, can be distinguished from traditional product spheres by a higher propensity of their normalized buzz-force rate changes to show co-evolution patterns.

The curious thing about this pattern is, that it no longer refers to a specific historical course of observed normalized buzz-force evolution. It does refer to a probabilistic curvature in historical buzz-force evolution, one which is maybe as hard to grasp as a curvature in spacetime.

One might add to the description of the pattern its immense, black-hole reminding potential to suck up information. The pattern, as describing the course of many different historical buzz-force evolutions amongst lots of associated networks (or of a larger scaled sphere), still describes a probabilistic commonality among them. One might suspect, that this kind of pattern, if inferred from a large numbers of networks, has a stronger power to predict an innovation then the other patterns studied here, because it presumably integrates larger amounts of information (information from single snapshot series of individual networks vs. information from multiple snapshot series of networked networks or spheres). The pattern is presumably ‘smarter’ then the other patterns studied here as it has more ‘knowledge’ of the web behind the world wide web.

Why to expect the patterns higher prevalence in innovation vs. traditional product networks

The answer is simple: interest-intensity of buzz on traditional products is more likely to follow a distinctive course of evolution then the interest-intensity of buzz on innovation topics, because by their very nature, traditional products mean a more limited range of ‘stuff to talk about’ then innovation topics: The amount of potential traditional products and all their associated categories is more limited then that of innovative products, because our collective memory is not above the phenomenon of forgetting. This sets a limit to the topics to talk about and by that to the interest-intensity people can have in traditional products (understanding interest-intensity as being fueled by the number of concepts being activated in mind, aka arousal). This limit is absent for innovative products, because they are not bounded to a limited set of categories to feed upon. Just recall that a radical innovation is something entirely new. Cultural forgetting does not limit the range of categories available for innovative products, because the process of setting up of Schumpeters ‘New Combinations’ knows to create an infinite repertoire of topics to talk about in a buzz and hence, a larger ‘fuel tank’ for the interest-intensity to draw upon: Not only the categories traditionally passed on are available to be activated, but also entirely new ones.

With regard to the course of historical interest-intensity evolution in buzz, the a priori more limited repertoire of interest-intensity ‘fuel’ of traditional product topics (and their associates) must mean a more clearly determined course of normalized buzz-force evolution. In other words, it is more likely, that a singular, typical line of average interest-intensity evolution could be depicted for traditional products then for innovative products. On the other hand, the potentially infinite complexity of innovative products allows to expect a much less clearly determined course of evolution in buzz. However, assuming that there is no magic in innovative thought processes, the innovative interest-intensity evolution still must be determined, somehow.

How? Not from nothing, but properly by other cognitions coming together, in a manner which may never repeat, but is determined from the mental categories which are there. As a reminder: The pattern described in hypotheses 4 observes how cross-topic interest-intensity co-evolution co-evolves differentially across innovative buzz vs. traditional product buzz. It seems likely, that such a pattern, which does not mean a distinct historical evolution but only the rules by which topics come together on the historical playground approximates the collective mental processes, which make an innovation a societal success (if feed with social media data).

5.4.6. *Hypotheses 5: The period bounded, cross-network average from the normalized buzz-force rate changes should differ between innovation meta-network and traditional product meta-network.*

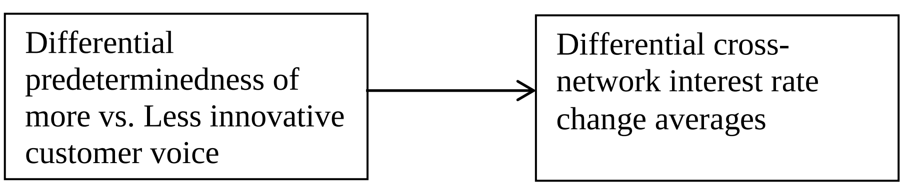


Figure 14 – Innovation predictor –customer-voice connection in hypothesis five

The reason for this difference is the same as the one for hypotheses 4, although here it is more simple: The cross-network co-evolution of normalized buzz-force rate change co-evolution is expected to be higher for the innovation meta-networks, as traditional product topics and talk about it allows a more clearly a priori determined path of buzz by their very nature. Average differences of the period bound innovative and traditional normalized buzz-force rate changes follow from a higher co-evolution in the innovative networks, because the average of the traditional product networks should follow a more clearly predetermined course then the one of the innovative networks. This means, that a difference in buzz-force rate change evolution indicates a period bounded innovation-traditional buzz distinction. For example, a statistically significant different average at period 13 means, that interest-intensity tends to differ for that period because of the pre-determinedness distinction between innovative and traditional product buzz, a difference from traditional product buzz on a distinct period is indicative for a period bound innovation typical difference. See under research design, that the innovative process has been observed for equivalent periods (necessary to attribute the difference to innovation buzz nature). In the following illustration, the cross-network average from the interest-intensity rate changes is shown to differ significantly between the two meta-networks for period 13 and period 49 at a $p < 0.05$.

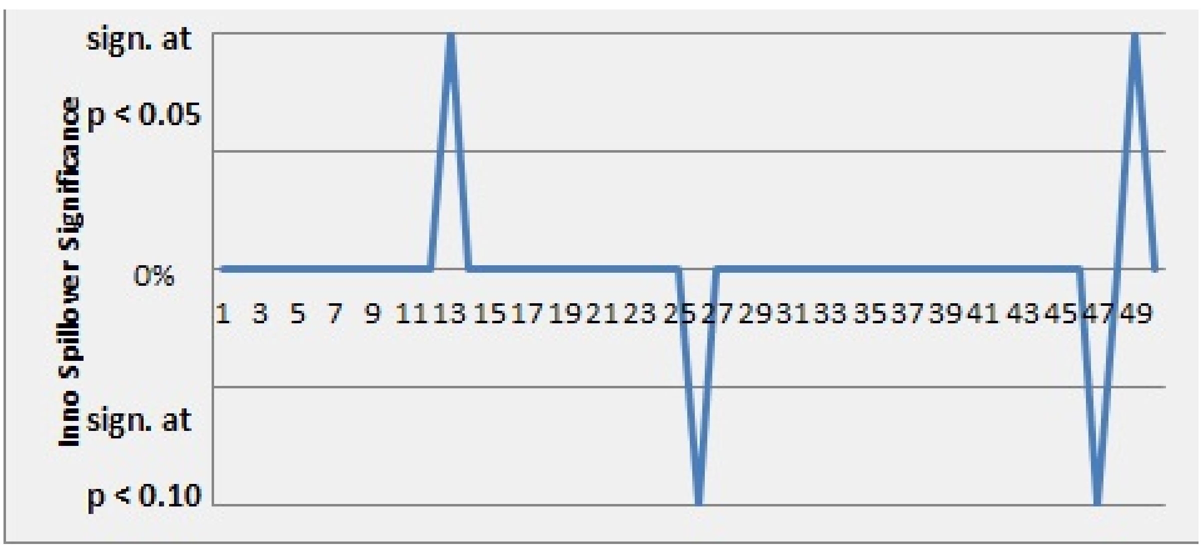


Figure 15 An example of the pattern of period bounded significant differences between the meta-network interest-intensity rate changes

5.5. Conclusion: The Suitability of Stochastic Patterns of Network Evolution to Predict Future Innovation

Some of the network-behavior co-evolution patterns capture trends of the buzz networks was over an observed time frame *at once*. These stochastic patterns describe buzz evolution beyond chronological time: The network evolution, as construed by the algorithm, is tested for effects by the average direction of changes in network and behavioral micro-steps (Burk, Steglich & Snijders, 2007). This average refers to the complete period of observation and simulation, the same average can refer to completely different factual network evolutions. For example, evidence for density effects (overall tendency of actors to have outgoing ties / degree of dyadic (two-sided) connection in the network) could come from a network evolution with lots of dyadic connections at the beginning and at a middle period and less in the end or from a factual evolution with lots of connections at the middle and end period and fewer in the beginning.

To reconcile, networks evolution patterns (i.e. network-attribute or behavior effects) are estimated as overall tendencies for the complete period at once. These tendencies apply likewise for different examples of chronological network evolutions. They group together similar courses of network evolution with different shaping: Evidence for the presence of an effect may come from a ‘real’ evolution, which has the effect in the beginning and middle period but not in the end. Likewise, the ‘real’ evolution, may have the effect at the middle and end period but not in the beginning. Both are characterized by the same change tendency.

For example, 30 products can be observed over a 52 week period. Network-behavioral micro-step average changes can be averaged for each of the 51 in-between periods. These periods can be compared for samples of different kinds of products and some periods may turn out to have different product type averaged micro-step averages for period 13, suggesting product typical average differences in the change of buzz between two physical points in time. These patterns may have very different factual evolution shapes in period 13. The tendency describes a mere statistical similarity beyond chronology, which is good for predicting the obviously multiple shaped process of innovation (or buzz preceding it).

Finding distinctive probability patterns for buzz network-attributes which factually co-evolved with an innovative product should mean, that these patterns predict the emergence of innovation, if these are found in comparison to co-evolution patterns of non-innovative products in a quasi-experimental research design.

Aboves mentioned patterns of social media dynamics could help to predict past, present and future innovation the like, because the change tendencies capture ‘evolution-dispositions’ of innovations (or another category of processes). Finding evidence for these innovation-evolution dispositions over longer periods around an innovations breakthrough means evidence for that disposition of a certain reliability. Seeking and finding these evolution-dispositions are a focal research goal of this study.

3. Methodology

3.1. Introduction - Data Analysis Overview

The goal of data analysis of this study was to find patterns, which arguably predict innovations and use those patterns to make predictions. The study can be seen as running through three phases. The first two phases are found conventional by Kalampokis, Tambouris & Tarabanis (2013) in big data studies, which try to find predictors for some event. The third phase is added, as the predictors are also used to forecast future innovations.

First, the 'raw' activity measures from the social monitoring service Coosto are transformed into longitudinal network data. Second, the results of inferential statistical analysis on network evolution of each individual network is meta-analyzed for predictive patterns. Third, the predictive pattern candidates are used to cast predictions for future innovations.

Phase one – data transformation to network data. The researcher assesses the activity on a particular topic with Coosto data over a defined time period in social media. In the first phase, the activity measures on a cluster of topics surrounding the buzz topic (i.e. innovation) were transformed into longitudinal network data with the software chain described below, because only by this transformation the buzz-networks evolution could be studied statistically. The transformation process works by first preparing the data in spreadsheet formatting complying with the expectancies of SonG software (see for detailed description of the software the extra section below), which delivers output readable for SoNIA. SoNIA in turn is the software which can format the data into a time series of adjacency matrices as the one presented in the part "2.2.2. Software implementation of the network evolution model". This time series of adjacency matrices is the data-basis for the inferential statistical analysis of each individual network.

This allowed to study the predictive variables: probability patterns in network evolution.

Phase two –data analysis for innovation prediction patterns. The data was analyzed for probability patterns in network evolution, which were distinctive for the innovative buzz-networks. To do the inferential statistics on network evolution with Visone software (see below), models have been employed, which had a fit with the theory above on the product category emergence mechanism (meaning an inclusion of average similarity effects, total similarity effects and average alter effects).

The results of these inferential statistical analyses of each network served as data-basis for the meta-analysis on predictive patterns of network evolution tendencies: The pattern of significant network and network behavior effects was compared between the innovation buzz-networks and the traditional product buzz-networks. Patterns which had a significantly higher proportion of emergences in innovation buzz-networks compared to traditional product buzz-networks were assumed to have predictive power for the emergence of a new innovation. Predictive patterns were always patterns of relative difference between the innovative and traditional sample and are only predictive by comparison with a traditional (i.e. the one here gathered) product sample. They are described in detail as Zeitgeist precognition metrics below, because they capture encryptions of future Zeitgeist.

Phase three –translation of innovation 'pregnant' social media buzz in a product positioning statement. Search queries on suspected fields of future innovation were done to find networks with the predictive patterns. For this, phases one and two were repeated for suspect networks (lucky guesses, since they were selected by hand). The predictive patterns of network evolution tendencies then served as screening criterion for buzz on innovation suspects. Some of the hand selection of innovation suspects returned as lucky guesses, meaning that pattern fit with the predictive models was found. The complete buzz (downloadable via the social media monitoring service) of these suspects served as data basis of phase three; forecasting of future innovations. When such a suspect network was identified, the source messages and discussion texts of these networks were downloaded as .csv files via Coosto for further analysis. Text mining software was employed to identify words, which belonged to the 50 most frequently mentioned words, normalized google distance algorithms served to estimate their semantic relatedness in the discussions. They are described in more detail as consumer voice precognition metrics below. These words and their semantic relatedness were then integrated in higher order categories and weighted for importance. The most important categories were then translated in a product positioning statement for the future innovation. This complex process is explained in detail under consumer voice forecasting.

3.2. Design

The main study compared patterns of social media networks of innovations with patterns of networks of more traditional, less innovative products (i.e. products which do not change substantially and are a commercial success). The patterns studied in this (meta-) meta-analysis of buzz networks studies were: Theoretically expected and non-expected significant effects in probability models on the evolution of buzz networks. Sphere patterns were compared per group as fraction of empirically emergent patterns of significant effects ($p>0.05$), which complied with the expectation. The design is a quasi-experimental multiple-time series design (Campbell, Stanley & Gage, 1963). One conventional way to diagram a quasi-experimental multiple-time series design (Campbell, Standley & Gage, 1963) is depicted below: The line above the --- represents measurements (O) in the natural group of innovative buzz, the line below --- means the measurements in the natural group of traditional product buzz, a full depiction would include 52 O. The X represents an intervention or treatment, here it is the critical period of radical innovation breakthrough, which is controlled by selecting a period of observation, which places the innovation breakthrough in the middle. This research design diagram illustrates that the breakthrough period has been placed in the middle of the complete time frame:

O O O OXO O O O

— — — — — — — — — —

O O O O O O O O

Effect significance patterns were compared between data of products and services, which underwent the process of innovation and products and services, which are less innovative by nature. Two types of data representing naturally assembled collectives of higher vs. lower innovativeness data were drawn from a pool of equivalent ‘sample-topics’, namely products; evidence for differential innovativeness of both natural groups is provided in the results section. The innovation process was identified in the process of qualitative, desk research on online articles with indicators on a breakthrough. The strength of this qualitative research was, that it allowed case-specific, idiosyncratic indicators of innovation breakthrough. This case-sensitive analysis did *not work by checking criteria, which were chosen in advance* but by appraising the individual stories of innovative products and firms.

For example, the Taleo recruitment software, a big data software selecting job applicants instead of HRM managers, was found to be innovative based on the evidence in Appendix B. Online journalists reported that the Taleo corporation became a market leader in integrated talent management, showing that it commercialized its recruitment-invention successfully (commercializing inventions = innovation, US Department of Commerce, 1967). Later on, Taleo was acquired by Oracle Corp. for $ 1.9 billion, hinting for exceptional return on equity expectations, which are likely to come from a breakthrough innovativeness in talent management. Even later on, a first user European user conference has been announced, were users can learn from each other how to make better use of the novel software, the need for knowledge sharing on a novel product indicates innovativeness as well.

The time series of each individual process of radical innovation was chosen so that the critical time period of the radical innovations breakthrough (X) was indicated to have occurred in the middle of the 52 weekly measurements (being evidenced by multiple types of indices found in qualitative research, as seen in the result section). The time series of each less innovative, ‘traditional’ products measurements’ was picked by convenience, as such products naturally lack a distinct period of radical innovation (there is no X).

3.3. Unit of Analysis

Buzz-networks are the unit of analysis of this study. Buzz-networks are interwoven topic nests in social media, which are collectively connected by n-links.

To make the difference between innovation and traditional product more extreme, radically innovative products (innovations with evidence for their novelty, uniqueness and technological impact, Dahlin & Behrens, 2005) were studied. Traditional products were taken as products, which have an ongoing commercial success due to their propensity to stay the same way as they are. The topics were observed over a 52 week period, on a weekly basis (as there are limitations in the number time periods process-able by Visone software).

The buzz on the following innovations were monitored, note that as shown under results only those were admitted with evidence for relatively higher radical innovativeness. For these examples, there was neither a practical possibility nor necessity to exclusively include ideal type examples of radical innovations, only the relative, statistically significant *group difference* on radical innovativeness of the two sample-groups is what is important for the validity of the conclusions drawn (evidence: see result section). Clipit, Coosto, Dropbox, Geldvoorelkaar.nl, Hootsuite, E-books, Iphone 4s, Irobot Roomba, Gamification, Kinect, League of Legends, Nissan Leaf, Taleo Recruitment Software, Samsung Galaxy, Smart Tv, 'Zombies, Run!' Game, Tesla Model S, Zynga social games, Twitter, Mercedes S-Klasse, Android, Apple App Store, Military Drones, Dacia Duster, Google Apps, Instagram, Ipad, Opel Ampera, PlanetSide2, WhatsApp.

The many criteria for their selection (radicalness difference test (see results), innovation historically taking place during Coosto recording time window (2009-2013), finding evidence for the innovation process (see results), and the presence of buzz about the product) made the screening of innovative products disproportionally time consuming. Therefore, for the products, which were assumed to be relatively non-innovative or being commercialized due to their non-changing nature, products were searched for which due to their socializing connected nature there was at least no problem with the absence of social media activity: Amstel, Bacardi, Bavaria, Beck's, Berentzen Apfelkorn, Bitburger, Dalmore, Dry Martini, Erdinger, Grolsch, Guiness, Gulden Draak, Heineken, Hendrick's Gin, Hertog Jan, Hoegaarden, Jack Daniel's, Jägermeister, Jameson Irish, Jever, Johnnie Walker, Keizer Karel, La Trappe, Oettinger, Paulaner, Rebel Yell, Springbank, Stella, Stolichnaya, Talisker, Veltins, Warsteiner.

Coosto.nl, a social media monitoring service helped to identify associated topics around the buzz topics. The top ten topics associated with the buzz topic have been used to identify the clusters. In principle, Coosto.nl suggest more topics to be included in the cluster. However, hardware limitations suggested the inclusion of ten topics only, otherwise, the algorithms of the software constructing the network layouts to use up all available RAM, resulting in system crash.

3.4.Basic Instruments

Coosto.nl social media webcrawling services served as primary instrument to capture the buzz in social media contents. Coosto gathers data from social media (i.e. Facebook or Youtube) in real-time. The measurements can be downloaded as .csv files. Coosto has a web archive since 2009 providing time sliced data. Coosto quantifies activity on an arbitrary search term (i.e. innovation term) and suggests associated topics, which co-occur with the search term (trending topics).

The main elements of the open source software chain are (some reformatting steps are omitted):

1. SonG (SoniaGetter helper application) to reformat Coostos .csv input into a formatting interpretable for SoNIA, comes with the software package of SoNIA. Theoretical output: timer ordered network patterns, sequential buzz-network cartography
2. SoNIA (Social Network Image Animator) is a software primary for visual simulation of dynamic, attribute-rich networks by sophisticated layout algorithms working with time-sliced network event data (see Bender-deMoll & McFarland, 2006). However, besides visual animation, SoNIA also allows to export the so simulated dynamic networks as time-series of adjacency matrices (represents the network over time) interpretable for Visone. Theoretical output: network dynamics, spheres over time
3. Visone (Visual Social Networks) to do inferential statistics on the networks evolution (see Indlekofer & Nagel, 2010). An open source code basis for dynamic network analysis, which was used several times by the open source community to develop programs for statistical analysis of dynamic network data is the SIENA (short for Simulation Investigation for Empirical Network Analysis, see Snijders, van de Bunt & Steglich, 2010) program, which is based on repeated measures models for the dynamic actor-oriented model of network evolution (Burk, Steglich, & Snijders, 2007). Recently, the command line based, less user-friendly RSIENA (using the mathematical program called R) was made accessible for GUI software for analyzing dynamic network data: Visone. To put it simple, Visone is 'remote-controlling' RSIENA within the command line based R environment from Visone GUI. Theoretical output: statistical patterns of buzz-network evolution beyond time

The use of this dynamic network analysis (DNA) toolchain is justified by constituting the simulation and computation backbone to a) capture social mediological networks and b) to do inferential statistics on their evolution.

4. TreeCloud (see Gambette & Véronis, 2010) served as a tool to analyze the social media discussion data, to which Coosto's algorithms pointed as constituting the cluster. Coosto affords to download these messages as a .csv file (limited to 10.000 messages). TreeCloud is a text mining software, which yields a graphic depiction of the semantic

proximity of words. TreeCloud can be used to group together words, which are in close association in the social media discussions. A more detailed account on how this was done can be found under consumer voice precognition metrics.

5. SUMO (the Suggested Upper Merged Ontology, see Sevcenko, 2003) is a formalized ontology, which was used to integrate the meaning of the word groups found by TreeCloud in higher order categories. SUMO can be used to categorize a lot of concepts, as it has been mapped out to all of the WordNet lexicon. It was used as a big but consistent coding scheme, as a means of meaning integration and to provide information of the weighting of the categories by procedures explained below under consumer voice forecasting.

The usage of these two programs is justified, because they were used to craft future innovation predictions from network data with predictive patterns.

3.5.Phase specific metrics and details on the data-flow in innovation forecasting

3.5.1. Buzz-network metrics – Data-transformation phase metrics

First the buzz measurements from Coosto.nl are transformed into longitudinal network data like this. As argued in the introduction under theoretical focus: buzz-networks, the definition of the metrics should be research question specific and is tailored for it. However, the current study, as the first of its kind and with a research question, which did not specify in advance any specific *relationship* to be studied in the data, the definition of the network was guided by the interest of the researcher. The influential metric for this research is normalized buzz-force, also called interest-intensity. The second influential metric is the trending topics algorithm of Coosto. Influential metrics are metrics, which define the network in a way, that their definition affects the outcomes of the inferential statistics employed. With other words, a dynamic network analysis was done on interest-intensity of sets of associated buzz-words, which are related to each other by being each others trending topics, which means that they are mentioned together in social media. This allows to observe, how interest-intensities change for the main associates of i.e. an innovation, it is a way to study interest-intensity of a major cutout of the buzz surrounding a particular topic / an innovation.

All other metrics did not influence the outcomes with the settings taken in this study, but can very easily be studied in future research, as they are all readily computed for all data. They are the outcome of explorative research.

1. The ego-node (i.e. innovation topic) buzz (aka buzz-focus) was estimated by Coostos activity measure for the centralized topic. This estimate was taken from both responses and postings, because it is a cross-discussion activity measure.
2. Alter-node buzz (aka buzz-sparks) was estimated by Coostos response and posting based activity measure for the trending topics, because trending topics are topics which Coostos monitoring software identified as being mentioned together with the ego-node topic (i.e. innovation).
3. Ego-alter-node link length / distance (aka buzz-ampere) was estimated by the activity measured by Coosto for the innovation topic in 'AND' conjunction with the trending topic. This serves as an estimate of link strength between the ego- and alter-node, because this measure feeds on peoples responses associating both topics (explicitly or implicitly) within discussions. Both responses and posts are included in this measure, because posts are an important anchor for within discussion linkages.
4. Ego-alter-node link strength or weighting (aka buzz-voltage) was estimated as the share of the nodes activity from the summed activity of each alter-nodes activity related to the ego-nodes base-line activity (alter-node activity divided by ego-node activity), because the conglomerate of the found alter-nodes served as a proxy for the most important parts of the ego-node emergence explanatory sub-node constituency.
5. Node specific buzz-spill over potential (aka buzz-force, the *network-behavior variable*) was estimated by node specific response-only activity assessed by Coosto, because the degree the topic sparks off discussions both depends upon its endogenous spreading potential (i.e. to what extend the topic is priming related topics in debating minds) and its exogenous spreading potential (i.e. others being drawn to the topics discussion by indirect priming not from within the discussion but by discussions about discussions (meta-discussions)). RSIENA affords to capture behavioral variables with values of whole numbers from 0-250 only, so the amount of reply messages was rescaled to this range for each individual network. The rescaling factors were picked for each network individually to save as much of the relative proportions between the single nodes reply behaviors (or

network-relationships) as possible: Rescale factors were chosen which resulted in minimized cut-off at both the low (below 1) and high levels (above 250). The factually observed interval variable could be described as topic tendency to arouse reply-messages (aka normalized buzz-force or interest-intensity of the crowd).

The metrics are corrected for distortions from fluctuations of overall buzz-extent, the activity measures of each metric have been divided by the complete number of messages published on each measured period (week). This way, a buzz-metric with higher information load has been construed from Coostos activity measure, this metric is called buzz-intensity. B*uzz-intensity* is corrected for the period specific 'thickness' of the 'buzz-ether', i.e. the probable relative buzz-impact of one message on another message on a day were only ten messages are send in total is larger than the probable buzz-impact of a message send on a day on which 10.000 messages are send. In other words, the lower the overall activity, the higher the average 'ballistic' effects of a single message on 'buzz-matter'. This correction works by the plausible assumption, that the lion's share of messages send in social media does not tend to have a continuously self-maintaining tail of responses but one with an end or interruptions. Put simple, it assumes that most social media buzz is not sourced by unpaused storytelling (i.e. a never-ending Tweet) but from limited episodes or storytelling with moments of silence in between (i.e. a post in a forum having a final answer or being answered after a year of pause). If this plausible assumption would not be true, then the average 'ballistic' effects of one message on a day with 10.000 messages would be higher.

Furthermore, only messages from Facebook were admitted, to avoid distortions of throwing together activity from different conglomerates of social media users, as the average demographics differ per medium (see Royal Pingdom, 2012 and Comscore Data Mine, 2013). The only exception was done in study 2 for measures on buzz of recent (known) inventions, where all social media monitored by Coosto were admitted as source, because the buzz was very weak on these emergent topics.

3.5.2. Innovation prediction models - Predictive analysis of phase two

Next the longitudinal network data is studied for patterns, which arguably predict innovations, the quasi-experimental design makes this arguable.

The predictive analysis seeks to assess buzz-network evolution patterns which are distinctive for innovations. Buzz networks were studied within the following statistical model, which components are depicted conceptually only. The distribution of significant effects was compared between innovation networks and traditional product networks to identify significant effect patterns with innovation prediction power (holding these models constant). A few pictures are helpful in explaining what is meant here. The second row from the right is the result of the effect-specific significance test, two example cases are shown.

behavior spreading linear shape	-0.036	0.143	0.801	-0.17
behavior spreading quadratic shape	0.001	2.918	< 0.001	0.075
behavior spreading average similarity	-1.654	6.870	0.809	0.036
behavior spreading total similarity	-1.876	11.33	0.868	0.03
behavior spreading average alter	0.003	0.006	0.661	-0.057
behavior spreading linear shape	-0.424	0.035	< 0.001	18.118
behavior spreading quadratic shape	-0.003	5.353	< 0.001	-0.635
behavior spreading average similarity	-7.356	4.906	0.133	2.492
behavior spreading total similarity	-10.293	7.950	0.195	2.521
behavior spreading average alter	-0.013	0.002	< 0.001	7.43

Figure 16. Partial model of buzz-dynamic tendencies with significance tests

One model was chosen in advance and held constant (see below), which serves to capture tendencies of network dynamics, it was specified upon theoretical reasons (see 2.3.5.), here the model tests for i.e. quadratic shape effects in all networks. The frequency of significant effects per category was counted for the more innovative sample and the traditional products sample. For example, in this subsample, two quadratic shape effects are significant, p was chosen as $p < 0.05$. Then the proportion of significant effects per effect category from all tests of the innovative or traditional product category was computed. Effect emergence-proportions of the innovative sample, which differed significantly from the traditional product sample, became part of the innovation prediction model. Differential prevalence of an effect in the samples is a signal of archetypical-evolution patterns of innovative buzz (compared to less innovative buzz).

Also the estimated behavioral rate change series were compared to identify innovation specific behavior rate change patterns. Finally, in a meta-meta analysis of the original data (or a meta-network analysis), innovation specific behavior rate changes of the meta-network with the individual networks behavior rate changes as behavior were studied as well. The statistical model for network evolution was construed following the advice of Ripley, Snijders & Preciado (2011); starting off with a simple model and adding and deleting effects so that significant effects start to emerge. For details on the meaning of the model, see Ripley, Snijders & Preciado (2011) and Snijders, Van de Bunt & Steglich (2010)):

Tested model of network dynamics and network-behavior co-evolution (Table 3):

Network Evolution	Effects on Behavior Evolution
Constant network rate function	constant behavior rate function
Network Evaluation Function, Structural Effects	Behavior (normalized buzz-force, i.e. response message arousal) Evolution Evaluation Function
Density Effect	Shape effect Quadratic shape effect Average similarity effect Total similarity effect Average alter effect

Tested model of network dynamics and network-behavior co-evolution in the meta-network analysis (which does not take into account network rate changes as multiple networks are compared by means of a dummy network to observe behavior-evolution differences only) (Table 4):

Network Evolution	Effects on Behavior Evolution
Network Evaluation Function, Structural Effects	constant behavior rate function
Density Effect	Behavior (normalized buzz-force, i.e. response message arousal) Evolution Evaluation Function
	Shape effect Quadratic shape effect Average similarity effect Total similarity effect Average alter effect

Table 4 Meta-network model of dynamic tendencies

3.5.3. Consumer voice metrics – Innovation forecasting phase preparations

Having found some predictive patterns, the raw data from the social media monitoring has to be content analyzed in the case of formulating a prediction of innovation of unknown inventions, because the associated topics suggested by Coosto do not represent what the customers are talking about but only its trending topics skeleton. The patterns identified with Coosto are a computerized gist of the buzz-network detached from the human language of thought, interpreting them in human language makes an extra analysis of the raw message and discussion data necessary. For this analysis, text-mining the raw message data was conducted. TreeCloud served to infer these metrics. Note that the metrics are a-chronological, just as the predictive patterns are. Both the settings of TreeCloud and SUMO software used for the forecasts is explained in detail in appendix A. Studying appendix A is advised to enhance understanding of the complex forecasting process, because a complete depiction is out of the conventional scope of a methodology section.

1. Bottom level semantic word clusters (aka micro buzz-sparks). Tree cloud identifies semantic relationships between the most frequently mentioned words of the analyzed messages. The bottom level relationships serve to identify word clusters, which are treated as conveying a common meaning in the subsequent analysis. The ellipse in figure 7 shows a cluster in TreeCloud output.
2. Bottom level isolated words (aka isolates aka micro buzz-foci). These are by not words without semantic association to the rest of the words, they only do not have a direct associate at the bottom level of the text-mined relationships. Isolates survive the first meaning integration of the subsequent analysis and contribute meaning to the next level of meaning integration directly. Isolates convey relatively specific information and do contribute as much to the forecasting as the integrated micro buzz-sparks (have equal weighting), because their isolated relatedness in the messages means that they contribute more to the meaning of the buzz-network then individual words of the micro buzz-sparks. The arrow in figure 17 shows an isolate in TreeCloud output.

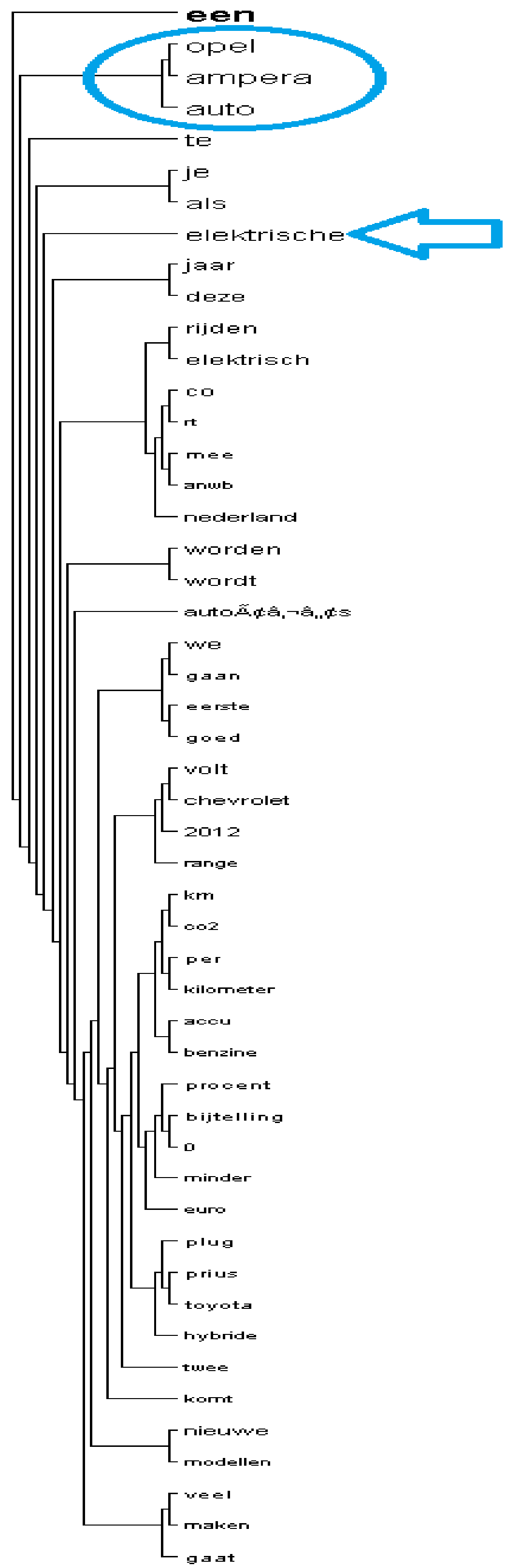


Figure 17. TreeCloud metrics for consumer voice precognition

3.5.4. Consumer voice forecasting – Innovation forecasting phase

Applying the consumer voice metrics on the complete textual data of a buzz-network with dynamics which predict innovations allows to forecast what the innovative buzz should deal with in the (nearby) future and this implies the prediction of a future innovation candidate.

The prediction of future consumer voice contents from its past contents is possible after scouting consumer voice with innovative tendencies in its dynamics: The innovatively changing consumer voice of the now and past implies the innovation itself, the innovative dynamics of the given contents are what constitute the future buzz on the then acknowledged innovation, because the innovative dynamics were observed in a buzz-network with a stable set of topics (stable set of trending topics). With other words, the future buzz on the innovation as well as the innovation itself is somehow already encrypted in the present and past buzz. Using the forecasting process of this study is a first step towards finding procedures, which help to crack that encryption. For that, the forecasting process must be improved on and on in future research to more closely match the words found for the innovation by the people.

To cast the first categorical predictions, the precognition metrics are subjected to iterative meaning integration and category prominence estimation. SUMO serves as a large yet consistent coding scheme for the consumer voice metrics (called consumer murmur in appendix A). Once coded, the common categories in SUMO are identified for each pair of semantically associated codes (the linkages are indicated by the merging lines in TreeCloud like the ones above). The precise process is outlined in appendix A. This process is repeated, until a reasonable number of relatively specific categories emerges (5-7), which yields an understandable product positioning statement. The categories are weighted as more important for the product by looking at the relative proportion which they have to the total number of categorizations. Only the specific categories are used for the product positioning, because they survived the powerful meaning integration process. The more general categories are too abstract to provide information, which is meaningful for human interpretation. For example, the abstract SUMO category weights give indication in how far the product is physical vs. abstract. This information is too broad to point towards any specific kind of product and can only be used later on to guide the precise formulation of the positioning statement.

This procedure ensures that categorizations, which came forth repeatedly and which managed to circumvent the meaning integration processes become very influential on the positioning. These features were interpreted as the defining features. The weighting process emphasizes categories, which escaped meaning integration by being dispersed in the metrics but still having congruent meaning. These categories are the gist of what the consumer voice is about in social media and were categorized like this:

- Upper group of the weighted categories: Core features of the product interest
- Mid group of the weighted categories: Expected features of the product interest
- Bottom group of the weighted categories: Augmented features of the product interest

A standardized product positioning statement looks like this and runs from high weight categories to low weight categories, its length depends on the number of found categories:
A … with … (core features) that is expected to … and to … (expected features) and is augmented by being … and … (augmented features).

4. Analysis of Data - Study One – Detection of Predictive Pattern Candidates

4.1. Introduction

First the results of the data analysis in support of the studies assumptions, i.e. differential innovativeness of the groups, are presented. After that the results concerning the main hypotheses are presented. Finally, conclusions are drawn on their implications to predict innovations in the subsequent study. Research question one is answered by that, the data is studied for patterns, which arguably predict innovations, considering the evidence of a quasi-experiment.

4.2. Evidence for the natural compound of collectives of more vs. less radically innovative products

An empirically validated definition of an inventions radicalness consists out of the dimensions of novelty, uniqueness and having a future impact on technology Dahlin & Behrens (2005). The natural radicalness of the two groups was estimated by measuring the normalized Google distance (NGD) for each individual product. NGD (aka Google Similarity distance, see Cilibrasi & Vitanyi, 2007), indicates symmetric conditional probability of co-occurrence of two words. This means, that if one of the words occurs on a web-page, NGD (x, y) measures the probability of that web-page also containing the other word. As the content of the internet is fed by the minds of people using it (or virtually everybody), the NGD maybe taken as a quickly assessable proxy of the semantic distance between arbitrary concepts within the average mind of the general population. A lower Google *similarity distance* means a *closer association* between concepts. To collect evidence for naturally higher radicalness in the natural group of highly innovative products, the average NGD between the three dimensions (or Dutch or English synonyms for them (i.e. new for novelty or innovative for technological impact)) of radicalness was computed for each of the 62 individual products. These NGD estimates of each products radicalness were computed with the website-hosted software Mechanical Cinderella (2013). For example, the Tesla Model S innovation scored a radicalness average of 0.1029 based on the NGDs between the words Tesla Model S and the words new (0.0739), unique (0.1688) and innovative (0.0660). Comparing the radicalness of the natural groups of more vs. less innovative products revealed a significantly lower Google Similarity distance average score (M = 0.31; SD = 0.16) for the natural Group of higher innovativeness products then for the natural group of less-innovative products (M = 0.62; SD = 0.13; t(29) = 8.53; p < 0.001), which indicates a closer association between radicalness and the more innovative products then between radicalness and the less-innovative or traditional products.

4.3. Evidence for the temporal position of radical innovation during the midst-periods of the 52-time series of more innovative product data

Indication for the critical time period for each (more innovative) products innovation process was gathered in a qualitative study, namely a desk research and online investigation. An open minded, data-inspired approach was taken to find case-specific indicators for the critical time period of the innovative process (commercialization), the links to the evidence for each cases time window is depicted in appendix B (as can be seen, not all candidates passed the probing for both invention radicalness and the presence of an innovation process).

4.4. Evidence for the buzz-network structures in social media meta data based on trending topic relationships

A network analysis was conducted for the example of the Twitter network. Gathering the top ten trending topics for ‘Twitter’ from Coosto, which are the ten topics, which are most frequently mentioned together with ‘Twitter’ and the trending topics of those, the relationship of trending topics to be each others trending topics was studied. Next to the egocentric layout resultant from the ten topics being the trending topics of 'Twitter', ties between lots of the trending topics being each others trending topics as well emerged and are depicted graphically in the following figure:

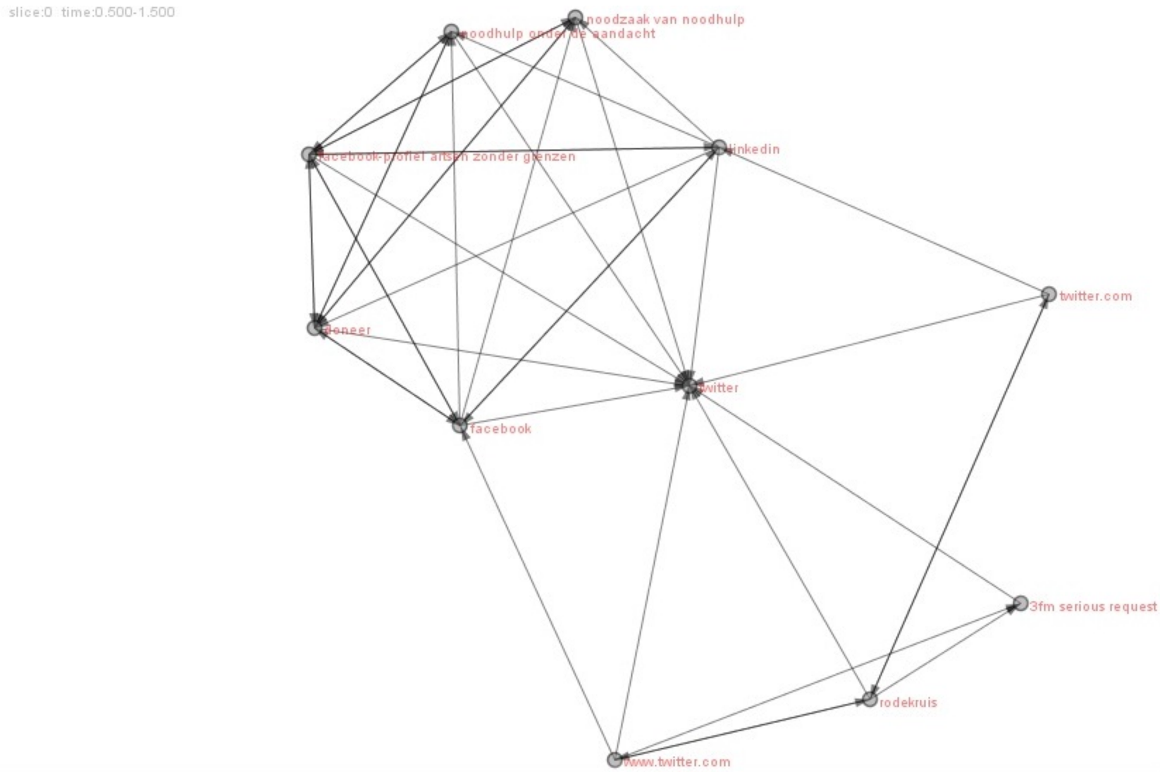


Figure 18 Twitters Trending Topics and the trending topics relationships between its trending topics

The network structure is not only apparent in this visual evidence but also in the descriptive statistics of this network: The network density, which shows how many of the possible links in a network are actually realized (van Wegberg, 2003), is considerable: 41/110 = 0.3727. Also the average degree of 3.7273, which is the average number of relations per actor (=trending topics per trending topic) (Boer, et al., 2003), is an indication for the presence of a network structure. To present even stronger evidence for the presence of a network structure, an extra analysis was conducted. By means of the exponential random graph models (Robins et al. 2007), it is possible to do inferential statistics even on single observations of network data with StOCNET software (Boer, et al., 2003). Utilization of these relatively recently developed exponential random graph models for analysis is understood as 'substantive network science' (Cranmer & Desmarais, 2011;

Keegan, 2013). For aboves Twitter network, an estimation with the following effects was run, convergence of the algorithm with the observed values was good. The structural meaning of these effects is depicted in bellows pictures, which are self-explanatory:

Estimates and standard errors

1. reciprocity 1.7501 (0.9467)
2. transitive triplets 0.4950 (0.0813)
3. 3-cycles -0.8547 (0.3210)

Network structure meaning:

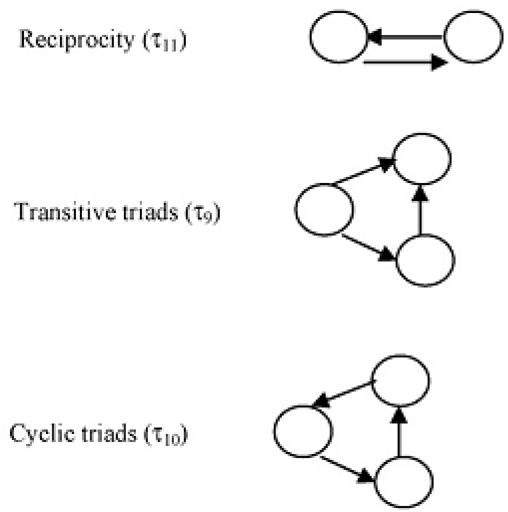


Figure 19: Network structures, pictures from Robins et al. 2007

The parameters, which were estimated from 1460 iterations of exponential random graph model simulations based on the real network observation indicate the presence of a network structure: The simulations yield evidence, that transitive triplet emerged more frequently from the estimations (as the estimate is positive and the estimate/standard error ratio exceeded 2 which is evidence for non-zero effects, see Robins et al., 2007), whereas 3-cycles emerged less frequently from estimations and the emergence of reciprocity structures was not statistically significant. There is statistical reason to assume, that there are real transitive triplet network structures in the observed layout, from which the random graph sample layouts are simulated. This is evidence for the presence of network structures between trending topic relationships of the buzz on Twitter and the trending topics of the Twitter buzz. Trending topics of Twitter are not only mentioned together with Twitter but also with each other, resulting in the emergence of a network of topics associated with each other in social media. By means of better fitting model specification in StOCNET, it would be possible to search for evidence for other network structures, yet here, the goal was just to demonstrate their presence, at all.

Because this set of trending topic-trending topics was limited to the top ten, it is reasonable to assume more relationships to emerge if all would have been admitted. Furthermore, it is plausible to assume, that all other studied, egocentric network layouts are factually part of much larger trending topic structures, which would yield statistical evidence for the presence of network structures as well if studied like the Twitter network.

4.5.*Hypothesis 1: Innovation related buzz networks should exhibit a lower tendency towards interest-intensity behavior subjecting itself to a negative feedback then non-innovative buzz networks (the proportion of the statistical models exhibiting significant negative quadratic behavioral shape effects is lower in the innovative networks compared to the less-innovative networks).*

Hypothesis corroborated. The proportion of significant negative quadratic behavioral shape effects is significantly lower in the more innovative-product networks (0 %) then in the traditional product networks (12.5 %; $z = 2{,}00$, $P < 0.05$).

4.6.Hypothesis 2: Innovation buzz network interest-intensity has higher tendency for co-evolution patterns than the interest-intensity in traditional product buzz networks (the proportion of statistical models exhibiting significant total similarity, average similarity or average alter effects for the normalized buzz-force behavior of the network is higher in the more innovative product networks than in the traditional product networks).

Hypothesis not supported. No significant differences ($P = 0.05$) were found for the occurrence of any of these three effects operationalizing inter-topic buzz-force influences (analogue to their operationalization of social influence on some behavior attributed to each person in a social network, i.e. smoking frequency).

4.7.Hypothesis 3: Average variance of innovation-network's interest-intensity rate changes (per period) is higher then the average variance of non-innovation network's interest-intensity rate changes.

Hypothesis corroborated. The average of the variance of (normalized) buzz-force change rates (averaged across the innovative products networks) is larger ($M = 220821.09$; $SD = 469491.739$) for the change of networks on more innovative products then the variance of buzz-force change rate averages (averaged across the traditional product networks) ($M = 95084.55$; $SD = 182211.87$; $t(29) = 1.78$; $p < 0.05$).

4.8. Hypothesis 4: The interest-intensity rate changes per period across all innovation buzz-networks should tend more towards co-evolution then the interest-intensity rate changes per period across all networks of traditional products.

Hypothesis not supported. No indicator of behavioral co-evolution (average similarity effects, total similarity effects and average alter effects) was significant ($p = 0.05$) in either of both meta-networks. There is no evidence of any co-evolution of period bound interest-intensity rate changes and hence there is no difference between the tendency of both networks to co-evolve on period bound interest-intensity rate changes.

4.9. Hypotheses 5: The period bounded, cross-network average from the normalized buzz-force rate changes should differ between innovation meta-network and traditional product meta-network.

By means of a time-series of independent samples t-tests on these averages, an innovation-traditional product distinguishing pattern of significant differences emerged. This pattern shows an innovation-distinctive course of rate change average evolution of significant difference arisal, because all data was sampled in such a way, that the period critical for the innovation process was in the middle of the 52 period. Curiously, there appeared an evolution of effect significances, which accounts for innovations even observed at different points in physical time, as long as they are observed in the same durations of time (52 weeks and a controlled critical period of the innovation process). In a 51 rate change period (as 52 week changes are observed), averages of period 13 and 49 differed significantly ($p < 0.05$) between innovation meta-networks and traditional product meta-networks (period 26 and period 47 differ with $p < 0.1$)

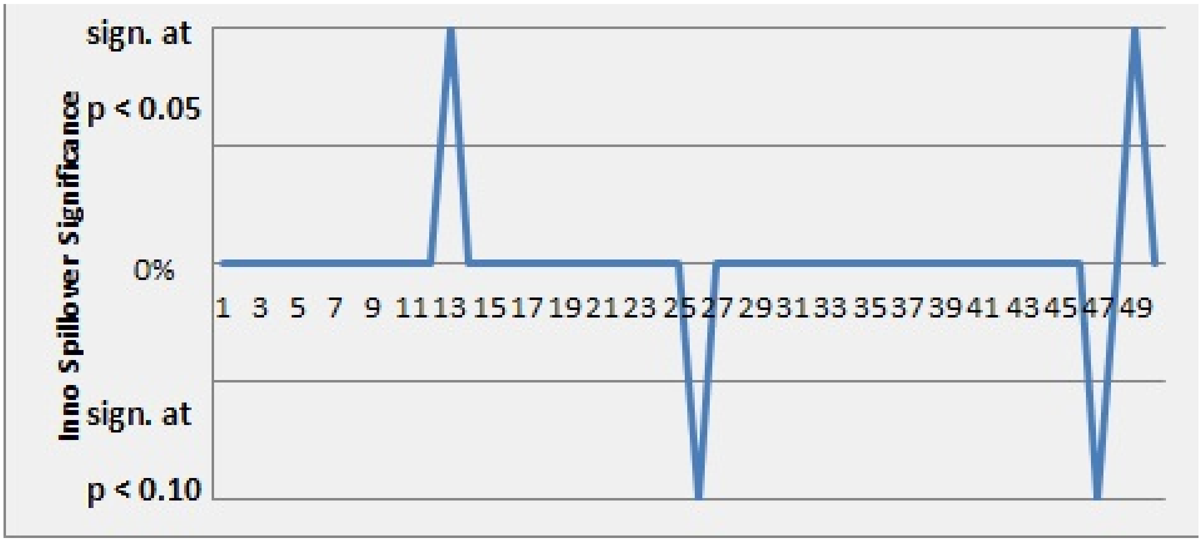


Figure 20. Innovation distinctive significance evolution of normalized-buzz force change rate average differences, a heartbeat pattern of innovation buzz dynamics compared to less innovative buzz dynamics

4.10. Conclusions for the innovation prediction studies – Predictive patterns

The research question on the identification of patterns, which arguably predict innovations, was answered with the found patterns like this.

Finding a tendency of exhibiting negative quadratic shape effects of interest-intensity in a new sample, which is significantly lower than the one found in the traditional product sample means that the topics of the new sample have higher innovation potential.

Finding a variance of interest-intensity change rates in a new sample, which is by average significantly larger than the variance of the traditional product sample means, that the new network samples topics have innovation prediction power. The variance of the new sample can come from the buzz-force change rate averages of the variance of the buzz-force change rate estimations for a single network, because the here handled cross-network averaged change rates construe yet another, equivalent estimation: An estimation of a generalized 30 node innovation buzz-network (a logically existent higher order network). A new sample network is seen as predicting innovation based on this variance pattern if the variance calculated for the new sample networks estimations is significantly larger than the variance calculated for the 30 node innovation buzz-network.

Finding new samples of buzz-networks, which differ significantly from the traditional products network on a proportion of differential average change rate periods (13, 49 and possibly on 26 and 47, too), which is higher than a proportion of 0 means that there is predictive significance effect model fit in the new sample (the new sample is predictive of an innovation).

In conclusion, initial evidence for the following probabilistic archetypes of innovative customer voice dynamics was found:

1. Higher probability of interest-intensity-oscillation in more innovative buzz
2. Higher volatility of interest-intensity rate changes (short name: variance pattern)
3. Period specific differences between cross-network interest rate change averages over a 52 weeks observation, placing the innovation breakthrough time in the middle (short: proportional time series significance pattern)

The patterns have been described and visualized in exhaustive manner in the theoretical chapters introducing the hypotheses.

5. Analysis of Data - Study Two – 'Product-of-Customer-Interest' Oracle Methodology

5.1. Introduction - Prediction of future innovations

Using the statistical patterns above, a number of buzz-networks has been found which have a systematically higher innovation prediction power then a random sample of buzz-networks.

Two kinds of predictions have been made:

1. Predictions on the likeliness of a known invention to become a radical innovation
2. Predictions on the likeliness of yet to be created inventions to become radical innovations

The second pattern works by a self-fulfilling prophecy of an to be created invention based on an interpretation of the consumer voice in buzz-networks with innovation prediction power but without an explicated invention. The invention is steered by an interpretation of the consumer voice and, as a self-fulfilling prophecy regarding future innovation, it may be regarded as scientifically reasonable consultation of a 'big data oracle' by a query and content analysis of Coosto data.

5.2. Prediction of radical innovation of known inventions

The buzz on the following not-yet innovated (commercialized) inventions have shown patterns motivating to assume systematically higher likelihood for a radical innovation of those:

Table 5 Innovation predictions for existent inventions

	Description	Predictive Pattern
Coin Betaalkaart	One programmable credit card unifying all others on one card	Variance pattern fit ($p < 0.05$) AND proportional time series significance pattern model fit ($p < 0.1$ for period 13 and 49 model, $p < 0.05$ for periods 13, 26, 47, 49 model)
Oil fabricated from water	A company has developed a microorganism, which digests water, sun and CO2 into diesel fuel or ethanol	Variance pattern fit ($p < 0.05$) AND proportional time series significance pattern model fit ($p < 0.1$ for period 13 and 49 model)
Del-i-cious	New retail concept: Warehouse for fine food founded on geldvoorelkaar.nl	Variance pattern fit ($p < 0.05$) AND Proportional time series significance pattern model fit ($p < 0.1$ for period 13 and 49 model, $p < 0.05$ for periods 13, 47, 49 model)
Foldable tablets	New technology allows the design of tablets or digital newspapers, which are foldable	Variance pattern fit ($p < 0.05$) AND proportional time series significance pattern model fit ($p < 0.05$ for periods 26, 47, 49 model)

5.3. Prediction of radical innovation of unkown inventions – Inventions from Self-Fulfilling Social Media Innovation Prophecies

The following buzz-networks exhibit innovation predictive patterns for not yet invented products. By applying the big data oracle methodology described in appendix A, the following interpretations of a consumer interest in an to be developed invention is explicated in table 6:

Table 6 Innovation predictions for non-existent inventions

Topic	Predictive Pattern	Big Data Oracle Methodology Self-Fulfilling Invention Prophecy
Debt concerns (national and private)	Variance pattern fit ($p < 0.10$; $p = 0.06$) AND proportional time series significance pattern model fit ($p < 0.1$ for period 13 and 49 model, $p < 0.05$ for periods 13, 26, 49 model)	An interactive social service dedicated for noticing and acting out jobs (core features) that is expected to decrease payments per month and to take care of family finances (expected features) and is augmented by being new, euro-currency based, having a proximate location, and dealing equally and truthfully (augmented features). Shorter statement: interpretation: A private replacement (i.e. insurance with widespread, local offices) of job-centers, which frees from the necessity to pay social security taxes (the deal: no social security will be received from the state) and has lower monthly charges than social security taxes for an corporation with a positive /ethical corporate identity.
Sustainable enterprises	proportional time series significance pattern model fit ($p < 0.1$ for period 13 and 49 model, $p < 0.05$ for periods 13, 26, 47, 49 model)	A large, new, internet-related corporation (core features) that is expected to promote inter-firm cooperation to increase corporate integrity and that is augmented by green and regional image products (augmented features). Interpretation: An innovation-brokerage firm with focus upon creating opportunities to increase business (moral, public) integrity by synergy of cross-sectional efforts to be perceived as socially responsible business
Shale gas	Variance pattern fit ($p < 0.05$)	An enterprise Drilling for a physical substance (Schaliegas) (core features) that is expected of organizational integration of national and environmental wants is augmented by distancing itself from unsustainable devices (drilling technology) and governance structures.
Sun panels	Variance pattern fit ($p < 0.05$) AND proportional time series significance pattern model fit ($p < 0.1$ for period 26 and 47 model)	Affordable sun panels from the region (core features) that is expected to meet certain quantity criteria (i.e. VAT on installation) and to be of Dutch origin (expected features) and is augmented by being novel on popular dimensions of sustainability. Interpretation: Solar panels being produced regionally are likely to gain competitive advantage, if they are financially and technically superior (i.e. by dramatic efficiency gain of some radically different technology) and have an even greener image than competitor products (i.e. by being industrially biodegradable (not in normal use of course)). Local solar panel production may have a future, if solar panels are re-invented fundamentally.

5.4. Conclusion – Why should one study the patterns in this dual set of studies?

Regarding the evaluation of the prediction strength, it is important to point out the distinction of predicting the innovation of known inventions vs. unknown inventions: Only predictions on known inventions, which have not been innovated yet can be used in future research to evaluate the prediction strength of the found predictive patterns alone. This is the case, because the strength of the predictions on unknown, social media based inventions is altered by self-fulfilling prophecy effects and the quality of the methodology of 'reading' the invention from the buzz. Self-fulfilling prophecy effects can be expected when reconciling the essence reasoned upon in the part on societal and scientific relevance: Self-prophecy effects (Sprangenberg et al. 2003) within a co-produced networked narrative (Kozinets et al., 2010), which is likely to emerge on a scientifically sound expectation on future innvation are likely to result in megamarketing (Humphreys, 2010), so the prediction of the innovation is boosted by putting it in words with the big data oracle methodology (writing it down the first time is already kind of a nano-buzz). Also the quality of the guess (compatibility with the real societal breeding ground for innovation) due to the sophistication level of the big data oracle methodology influences the prediction strength. The prediction of a known invention, if prevented from causing extra-buzz by secrecy, can be used to evaluate the predictive patterns strength without these distortions. The 'distortions' however are very welcome in a research line investigating the outcome prediction strength in combination with the here developed oracle methodology, an improved version of that or alternative oracle techniques.

6. Conclusions and Implications

6.1. Introduction

In short, the research problem of advancing prognostic market intelligence has been successfully addressed by the following findings.

Patterns in buzz-network evolution were shown to be linked to degree of innovativeness. Technically speaking, probability patterns amongst models, which describe network evolution stochastically, have been linked to sets of buzz topics, which were related in a network structure by being trending topics of a central buzz term.

6.2. Conclusions about research question one

Regarding research question one, specific patterns in buzz-network evolution have been found, which are useful in predicting innovations by means of computer-supported analysis.

Relatively more radically innovative buzz showed a tendency towards network-evolutions, which fit the following probability patterns among models of network evolution including the rate changes of the network ties, density effects, interest-intensity or normalized-buzz force (NBF) behavior, NBF behavior shape effects, NBF behavior quadratic shape effects, NBF behavior total similarity effects, NBF behavior average similarity effects and NBF behavior average alter effects:

1. Lower tendency towards negative NBF behavior quadratic shape effects then buzz on less innovative topics
2. The average of NBF change rate variances tends to be larger for the change of buzz-networks on more innovative products then the average of the variance of NBF change rates
3. In a 51 NBF rate change period, averages of change rates of period 13 and 49 differ significantly ($p < 0.05$) between innovation meta-networks and traditional product meta-networks (period 26 and period 47 differ with $p < 0.1$), if the 51 NBF period is sampled in a way, which positions the breakthrough period of the radical innovation in the middle of all observed periods.

Because of the quasi-experimental design of the study, relatively *strong evidence is found* for the differential nature of the two kinds of buzz (radically innovative vs. traditional product related, which difference is also grounded in empirical evidence (see first two points under results)) to be the cause for the observed differences.

6.3. Conclusions about research question two

Charged with these insights, which need further replication in similar studies and longitudinal studies upon those studies to be *confirmed* by the scientific community to be on causal ground, an attempt was undertaken to predict future innovations, addressing research question two. This was done by repeating this analysis on a number topics, which were suspected by the researcher to entail innovativeness as well, a prediction on increased chance for innovation of nine buzz topics (the predictive subsample, the lucky guesses) was made. Four of these topics were on topics of known inventions, which were not yet innovated and five on topics, for which there was no known invention yet. Using the here developed method, the buzz was ‘read’ to come up with a social media based invention idea, which just like its original buzz has presumably a higher innovation potential as well.

A few notes on how the lucky guesses were found seems to be justified by the fact, that the search for them did not take as large amounts of time as was worried. As this may indicate that the here handled intuitions performed well, the rules of thumb which were used here may be of worth for future research searching such networks by hand as well:

- To find buzz on topics of recent inventions, trending topics on the words of ‘uitvinding’ (invention) were browsed as well as their trending topics. Checking out a crowdfunding platform named Geldvoorelkaar.nl also turned out to be helpful.
- To find buzz on topics of non-yet invented future innovations, crisis-loaded themes which were under discussion in Dutch online media over the last year were studied, following the Chinese and Greek linguistic wisdom that the word crisis is also related to the upcoming of new chances (Zeit, 2003). Crisis-loaded buzz may be well suited to be transformed into innovation buzz of the future.

6.4. Conclusions about the research problem: Design of innovation forecasting software

All analysis steps were either entirely software based (finding predictive patterns and testing them on known inventions) or are heavily supported by software, with potential for complete automation (testing and using predictive patterns for unknown inventions). This research identified toolchains, which can become the backbone for the design of an innovation forecasting software. Much of the programming work is saved in this design, because most tools are open source software, the source code of some of these tools may flow directly in a project to develop the software. Major elements of the software could draw back on the following functionalities:

1. A social media monitoring service, i.e. Coosto or free alternatives like Social Mention
2. Transformation of social media activity measures into longitudinal network format: SonG and SoNIA source code
3. Inferential statistical analysis of network evolution with the source code for RSIENA or SIENA
4. Meta-analysis for predictive patterns, i.e. based on source code from open source alternatives for SPSS, like PSPP

5. Product-of-customer-interest oracle ‘reading’ of raw buzz. First element: finding main themes in the buzz by means of TreeCloud alike text-mining. Second: automatic weighting their importance, making use of a SUMO based coding scheme. Final element: Interpretation as product positioning statement by automated common sense. A rough conception of automated common sense casting positioning statements: Connecting the important themes by a software with access to the already existent SUMO-WordNet interface (ontology-lexicon connection to retrieve words) and then spicing the stochastic selection process by super-positioning the common sense biased DOLCE ontology semantics on the SUMO ontology.

These elements may be a fruitful start-off for a commercial project to develop an innovation forecasting software, which has the power to deliver the necessary quantities of evidence to confirm predictive patterns, help to refine them and do robust predictions.

6.5. Implications for theory: Innovative buzz and mediological origins of innovation

Reconciling the key points of the theory behind the found predictive patterns, innovation related buzz seems to capture peoples interest like a hype or temporary trends and fashions and less like cognitive habits, which are deeply rooted in culture. Also, interest on innovation buzz appears to follow trajectories, which are relatively less pre-determined in advance when observed in isolation and are more determined in terms of co-evolution: There seems to be no single typical curvature of interest-intensity in innovation buzz, but there appear to be typical tendencies of interest co-evolution in innovation buzz. This weaker degree of determination is likely to come from the fuzziness of the innovation concept itself, the categorical bondage of innovations is relatively low and so is the development of interest-intensity on its corresponding buzz.

Another theoretical implication of this research is the perspective one takes upon concepts such as the here proposed buzz-networks. Buzz-networks and especially metrics like interest-intensity are examples of hybrid concepts. Interest-intensity is a good example of a bastard entity somewhere in between of a social media technology blurred idea of a socio-psychological or collectio-cognitive state. Interest-intensity assesses the degree of fascination of the dynamically changing currently active crowd in social media on a particular topic. As such, it is a sociological, technological and psychological mix. Accordingly, mediology is “the discipline that treats of the higher social functions in their relations with the technical structures of transmission”, (Debray, 1996) and

„In the word "mediology," "medio" says not media nor medium but mediations, namely the dynamic combination of intermediary procedures and bodies that interpose themselves between a producing of signs and a producing of events. These intermediates are allied with "hybrids" (Bruno Latour's term), mediations at once technological, cultural and social.” (Debray, 1996)

Studying innovation from such a mediological point of view on social media may help to develop more advanced terminology on the unit of analysis and to observe ‘social mediological’ intelligence processes as the here proposed category learning mechanism.

Regarding the lack of support for the category learning mechanism found in this study, it may be more suitable to study a more advanced definition of buzz-networks in future research, one which is more adjusted towards relationships in buzz-processing instead of buzz-states like being a trending topic over the complete period of observation, a topic that was frequently mentioned with the main topic. A less static relationship is being a trending topic for a portion of the observation period. For example, one could observe and capture weekly trending topic changes, resulting in more dynamic network structures and processes. Intelligence processes like category emergencies may arise in these social mediological constellations and processes for the emergence of innovation can be continued to be studied in this manner. A social mediological perspective can help to observe and describe, how the collective intelligence of the crowd thinks and affords a breeding ground for innovations.

6.6. Implications for policy and practice – crowd governed economy

Innovation forecasting software may reform policies both on the firm level as well as on a systemic level.

On the firm level, innovation forecasting from social media can exert a top down control on integrated marketing strategies, as implied by the E-marketing concept of Constantinides (2013):

Figure 21 The position of social media marketing in integrated marketing strategy (Constantinides, 2013)

Marketing strategy includes market orientation as reflected in quality products at the bottom level, organizational resources for effective marketing of these products on the next level and an online presence to channel both product and organization to the customer.

The position of social media marketing strategy is at the top and a passive approach of using social media as tools to generate market intelligence can guide bottom levels of marketing strategy. Innovation forecasting software could be used to inform marketing strategy towards promising fields, tailored to the available resources of the firm, as forecasting innovations implies prognostic market intelligence, meaning that customer needs and competitive movements would be known in advance with a limited certainty. Marketing policies on all lower levels could become anticipatory instead of reactive with prognostic market intelligence.

On the systemic level, the policy implications are even more far reaching and provocative. The systems of innovation approach sees innovation not as the result of isolated firms but as the result of continuous interactions of firms with other organizations of the system (Chaminade & Edquist, 2008). Public policy interventions attempt to counterwork systemic problems. However, the promise of a working innovation forecasting software would allow public policy interventions to prevent systemic problems: By monitoring forecasts on systems or clusters of innovations of societal relevance (i.e. those being connected to a lot of workplaces), systemic problems arising from regional restructurings of economy could be predicted. Public innovation policies could then be made, which direct the local system of innovation in trajectories, which are anticipated as promising by the software. Costs of such state financed economical restructurings by means of subsidized R & D in promising trajectories are then likely to be overcompensated by taxes gathered from the healthy innovation system. Another interesting consequence is the prevented increase of social security expenses due to shorter or absent periods of economic downturn. In a provocative way, systemic governance through prognostic market intelligence of the crowd in social media has the look and feel of a planned economy. Big data fed market intelligence may avoid the failure of such a planned market economy: One of the main drawbacks of the idea of planned economy is insufficient centralized information on individual needs to do efficient planning (von Hayek, 1996). Big data on customer voice used for prognostic market intelligence may eliminate this drawback. The development of new, computer-supported plans for systems of innovation would then seem to be a viable perspective.

6.7. Implications for private sector managers – Recommendations for Innovation Managers and Marketing Managers

Taking the state of the research as it is, innovation managers, marketing managers and R & D staff can benefit from the methodology for inspiration to anticipate or create trends. The methodology at its current state of development allows to screen potentially innovation related customer voice by quantitative means. Furthermore, the methodology of ‘reading’ the main themes in large databases on customer voice is a novel approach towards creating possibly counterintuitive ideas for innovation and marketing strategy as well as for new product development. Even without looking at customer voices with the predictive patterns, the ‘product-of-customer-interest’ oracle methodology allows to generate product ideas, for which there already is a community of an engaged online crowd. Regardless of the endogenous innovativeness of the crowds talk, anything new generated from its talk may be especially suited to innovate with, because the interest of the engaged crowd could be transformed in product interest more easily than of a random sample of people.

Assuming an innovation forecasting software is developed, marketing managers, innovation managers and R & D staff would be well advised to consult and browse the prognostic intelligence frequently for their own analyses.

6.8. Implications for public sector managers – Recommendations for governance representatives

Considering the promising potentials of a working forecasting software for societal welfare, representatives may be interested in co-innovating the software to generate automated inferences for governing systems of innovation by the collective intelligence of the crowd, as suggested above. Great regional, cross-sectional competitive advantages could be realized, if the development of the software is supported by R & D incentives.

6.9.Further research and advancing limitations of the study

A number of topics should be discussed with regard to future research. These regard the value and way of usage of the different findings of this study in research on the prediction strength, other interesting foci of research with the here gathered data and line of research as well as research using other data bases, as well as other interesting metrics to be tested in future research.

Prediction Strength Research Lines

With regard to other foci of research using the here gathered and transformed data, the fit of the probabilistic network evolution model as well as other definitions of the networks as well as other kinds of proposed patterns could be studied.

Probabilistic Network evolution model fit – A limitation

The probabilistic network evolution model (density effects, average similarity effects, etc.) of this research was held constant and picked upon both theoretical grounds and explorative research. However, it is not claimed that it is the model with the best fit to the data. It is possible to do meta-analysis with the here gathered data on the convergence of the simulated evolutions with the real, observed values. A search could be undertaken, to identify a network evolution model, which has the some kind of optimum average on the convergence statistics among all studied network evolutions. This should increase predictive power, as the sophistication and number of pattern fit indicators / predictive variables of the resultant predictive models would increase.

Alternative buzz-network perspectives

With the research question having a focus on NBF behavior co-evolution, the network was mainly construed with NBF behavior the node definition from the trending topics algorithm.

An alternative research focus could be placed on the relative proximity of the trending topics in social media, by defining a network behavior capturing the distance between the trending topics in social media. The innovation term-trending topics combined activity as distance (buzz-ampere) and proportional distance (buzz-ampere) between the trending topics buzz-intensities readily available in the data set for explorative research, only some reformatting of the data is necessary for that. Buzz-ampere above some innovation indicative threshold (which could be searched amongst the here gathered data) could be used to define ties emerging and vanishing to study network structural predictors.

A alternative measure for this could be the number of trending topic queries in-between the trending topics to establish a trending topic algorithm based trace amongst the buzz. Observing this behavior co-evolution might give indication, in how far a newly emergent trending topic is a function of other trending topics proximity. This type of analysis could be used to design an alternative, automated oracle methodology.

Furthermore, the buzz-network perspective could be given higher structural sophistication. For example, the trending topics could be observed period wise as the activity data as well, allowing to map out networks with emerging and vanishing ties between the buzz topics. This would have the great advantage of allowing to observe network structural network evolution patterns as predictive. Also structural-behavioral co-evolution could be studied. Here only network-behavior co-evolution has been observed, structural analysis was not undertaken. Broader structures then the here undertaken egocentric, ‘buzz-lead-term-focussed’ layout could be studied by that, relationships between the trending topics could be entered as well. Only these structures have a chance of being incorporated on spherical map constellations on some day, because using Levin’s (1972) techniques requires to observe more comprehensive network structures (there must be more network complexity to be simplified by Levin’s techniques).

Alternative probabilistic network evolution patterns

Related to different network perspectives on the data is the inclusion of other kinds of effects to capture the probabilistic network evolution patterns. Besides network-behavior co-evolution, structural tendencies alone, network-structure – network-behavior co-evolutions as well as alternative behavioral or attribute variables, observed next to the network could be included. For example, data from alternative methodologies could be combined, that includes qualitative research methods. For example, the prediction results of the big data oracle methodology adjusted to yield predictions for the business and science park of Enschede from a longitudinal database of articles and publications related to their businesses (semi-automatically googled and downloaded with the Ghostmouse software) searched for network-evolution patterns in their time series Tree-cloud networks allows to do highly specific big data studies without the availability of a data gatherer like Coosto (as Enschede buzz is properly too specific to be found on Coosto in big quantity). Combining such a research with interview data on the guesses of entrepreneurs of the region on the likability of the predictions outcomes allows to observe co-evolution patterns, which could help to find patterns to predict, when self-prophecy effects set in.

Alternative oracle procedures – Another limitation

Buzz, which is suspected to predict innovation, is 'read' in the current research by taking the complete record of the buzz. However, it is likely, that only a very distinctive fraction of predictive buzz is actually constituent of the encrypted innovation. The current study did not study particular dynamics of the complete buzz. Only fractions of buzz-network dynamics of predictive buzz were studied. More advanced procedures would look for the complete picture of buzz transformations, until the innovation has been labeled within the buzz. This may lower the predictive power of the forecasts, future research could improve on this by studying more extensive buzz networks then the 10 node networks studied here. In addition, particular or detail dynamics should be identified, which lead to the labeling of the innovation. Using these dynamics in the forecasting process may increase predictive power of the methodology.

Suitability of the buzz-blitz metric for innovation prediction

As an example for alternative metrics, a metric to be called buzz-blitz could be studied for its use to predict future (innovation) buzz itself, as here the focus was placed on innovation, an event outside of the social media.

Future research could study, if interest-intensity as measured by NBF predicts the structural evolution of buzz-networks. Assuming some network definition could be found for this to be true, it could successively studied, if there is some predictive probability distribution among the NBF to predict distinct structural evolutions of buzz-networks (social media talk discourse conventions). Next, networks could be searched, which copy these conventions from each other, or which persuade each other to buzz in similar manner. Predictors could be developed from comparative studies like this, which help to predict if certain buzz is about to hijack another

buzz. Putting these predictors together could help to construe a metric estimating the likelihood of an arbitrary buzz to predatorily capture other buzz, which could be called buzz-blitz.
Buzz-blitz patterns in turn could be studied for their connection with interest-intensity or buzz-force, searching for patterns. Again, assuming patterns found, the co-evolution of buzz-force and topics, which are found to be a great societal issue by some other method (Zeitgeist) could be searched.
Such a chain of patterns would allow to predict Zeitgeist-change from change of social media buzz-blitz, which captured peoples online interest (buzz-force) by some pattern. Events like the Arabic spring could then become predictable from the mere structural change of social media buzz, not even looking at their contents.

Improved Research Designs

In principle, true experimentation instead of quasi-experimentation is possible for research like this one: The topic 'assignment' could be randomized by a random topic picking software and then categorizing the topic as fitting the to be contrasted categories or not (here: innovative vs. traditional products). However, a larger data-base may be desirable, because these random words would properly result in a lot of necessary random-'assignment' retries, if no buzz was found for the random topic. Using Coosto database with a theoretical maximum reach of roundabout 20 million people still yielded lots of no-buzz measurements on properly too specific innovation topics. A larger data base (i.e. factual global Facebook reach: one billion) could yield activity measurements on these highly specific topics: the more buzz is monitored, the higher the chance that highly specific, niche discussions are observed, too.

6.10. Conclusion

It is concluded, that the application of dynamic network analysis on big data is useful to predict radical innovations, the research problem of advancing prognostic market intelligence has been successfully addressed. The study of network-patterns in big data and their role in making predictions for the future can only be regarded as at its very start. Particularly, the combination of different tools to generate network structures over time from more and more flexibly integrate able textual data and statistically analyzing this data with the here handled tools or others could yield a whole new bunch of methodologies to study big data for various kinds of research questions in quantitative (Study 1) or qualitative (Study 2) manner.

Further development of the here suggested methodologies for prognostic market intelligence would benefit entrepreneurs in various ways, such first mover competitive advantages, increased chance of adoption of the innovation by the target groups or in creating new fields of business, which offer additional chances for long-term growth by complementary product lines.

References - Articles

Bender-deMoll, S. & McFarland, Daniel A. (2006) "The Art and Science of Dynamic Network Visualization." Journal of Social Structure. Volume 7, Number 2.

Boer, P., Huisman, M., Snijders, T. A., & Zeggelink, E. P. (2003). StOCNET: An open software system for the advanced statistical analysis of social networks. ICS/SciencePlus, Universiy of Groningen, Groningen, Netherlands, version, 1.

Burk, W. J., Steglich, C. E., & Snijders, T. A. (2007). Beyond dyadic interdependence: Actor-oriented models for co-evolving social networks and individual behaviors. International journal of behavioral development, 31(4), 397-404.

Chaminade, C., & Edquist, C. (2008). Rationales for public policy intervention in the innovation process: A systems of innovation approach. Innovation Policy-Theory and Practice. An International Handbook.

Choi, H. and Varian, H. (2012), "Predicting the Present with Google Trends", The Economic Record, Vol. 88, pp. 2-9.

Cilibrasi, R. L., & Vitanyi, P. M. (2007). The google similarity distance. Knowledge and Data Engineering, IEEE Transactions on, 19(3), 370-383.

Collins,A. M.& E. F. Loftus(1975). A Spreading Activation Theory of Semantic Processing, *Psychological Review*, 82,407-28 .

Constantinides, E., & Fountain, S. J. (2008). Web 2.0: Conceptual foundations and marketing issues. *Journal of Direct, Data and Digital Marketing Practice*, *9*(3), 231-244.

Cranmer, S. J., & Desmarais, B. A. (2011). Inferential network analysis with exponential random graph models. *Political Analysis*, *19*(1), 66-86.

Dahlin, K. B., & Behrens, D. M. (2005). When is an invention really radical?: Defining and measuring technological radicalness. *Research Policy*, *34*(5), 717-737.

Geroski, P., Machin, S., & Van Reenen, J. (1993). The profitability of innovating firms. The RAND Journal of Economics, 198-211.

Gligorov, R., ten Kate, W., Aleksovski, Z., & van Harmelen, F. (2007, May). Using Google distance to weight approximate ontology matches. In *Proceedings of the 16th international conference on World Wide Web* (pp. 767-776). ACM.

Goel, S., Hofman, J.M., Lahaie, S., Pennock, D.M. and Watts, D.J. (2010), "Predicting consumer behaviour with Web search", Proceedings of the National Academy of Sciences (PNAS), Vol. 107, No. 41, pp. 17486-17490.

Gruhl, D., Guha, R., Kumar, R., Novak, J. and Tomkins, A. (2005), "The Predictive Power of Online Chatter", in Eleventh ACM SIGKDD international conference on Knowledge discovery in data mining, ACM Press, pp. 78-87.

Gambette, P., & Véronis, J. (2010). Visualising a text with a tree cloud. In Classification as a Tool for Research (pp. 561-569). Springer Berlin Heidelberg.

Guzman, G. (2011), "Internet search behavior as an economic forecasting tool: The case of inflation expectations", Journal of Economic and Social Measurement, Vol. 36, No. 3, pp. 119-167.

Haythornthwaite, C. (1996). Social network analysis: An approach and technique for the study of information exchange. Library & Information Science Research, 18(4), 323-342.

Kalampokis, E., Tambouris, E., & Tarabanis, K. (2013). Understanding the predictive power of social media. *Internet Research*, *23*(5), 544-559.

Kirchhoff B. 1991. Entrepreneurship's contribution to economics. Entrepreneurship Theory and Practice 16(2): 93–112.

Kozinets, R. V., De Valck, K., Wojnicki, A. C., & Wilner, S. J. (2010). Networked narratives: Understanding word-of-mouth marketing in online communities. *Journal of marketing*, *74*(2), 71-89.

Levine, J. H. (1972). The sphere of influence. *American Sociological Review*, 14-27.

Nelson, R. R., & Winter, S. G. (1977). In search of useful theory of innovation. *Research policy*, *6*(1), 36-76.

Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph models for social networks. Social networks, 29(2), 173-191.

Sevcenko, M. (2003). Online Presentation of an Upper Ontology. In Proceedings of Znalosti 2003, Ostrava, Czech Republic, February 19-21, 2003.

Simon, H. (1959). Theories of decision making in economics and behavioural science. American Economic Review, Vol. 49, No. 3, S. 253 - 283.

Snijders, T.A.B., van de Bunt, G.G., and Steglich, C.E.G. (2010). Introduction to actor-based models for network dynamics. Social Networks, 32, 44-60

Spangenberg, E. R., Sprott, D. E., Grohmann, B., & Smith, R. J. (2003). Mass-communicated prediction requests: Practical application and a cognitive dissonance explanation for self-prophecy. *Journal of Marketing*, 47-62.

Stremersch, S., Tellis, G. J., Franses, P. H., & Binken, J. L. (2007). Indirect Network Effects in New Product Growth. *Journal of Marketing*, *71*(3), 52-74.

Thij, M. (2013). Modelling trends in social media-Application and analysis using random graphs. Masterthesis, University of Twente.

van Wegberg, M. (2003). A note on network structure: functional relationships between network indicators. METEOR, Maastricht research school of Economics of Technology and Organizations.

Wu, L. and Brynjolfsson, E. (2009), "The Future of Prediction: How Google Searches Foreshadow Housing Prices and Quantities" in 30th International Conference on Information Systems, AISE, http://aisel.aisnet.org/icis2009/147

Zhuge, H. (2009). Communities and emerging semantics in semantic link network: Discovery and learning. *Knowledge and Data Engineering, IEEE Transactions on*, *21*(6), 785-799.

References – Online Sources

Comscore Data Mine (2013). Retrieved on 22.07.2013 from http://www.comscoredatamine.com/2013/06/who-uses-social-networks-in-the-netherlands/

Indlekofer & Nagel, 2010. Visone Manual. Retrieved on 21.11.2013 from http://visone.info/lib/exe/fetch.php?media=docs:visonesiena_man.pdf

Keegan, B. (2013). https://sites.google.com/site/statnetonec2/home

Mechanical Cinderella (2013). All queries retrieved between 31.07.2013 and 31.11.2013 from http://www.mechanicalcinderella.com/

Network theory. Wikipedia, 2013 retrieved on 08.12.2013 from http://en.wikipedia.org/wiki/Network_theory

Neville, J. (2011). "Modeling Complex Social Networks: Challenges & Opportunities for Statistical Learning & Inference". Retrieved from http://www.youtube.com/watch?v=1xLjYc7EUEU

Royal Pingdom (2012). Retrieved 22.07.2013 from http://royal.pingdom.com/2012/08/21/report-social-network-demographics-in-2012/

Snijders, T. A., Steglich, C. E., Schweinberger, M., & Huisman, M. (2007). Manual for SIENA version 3. Groningen: University of Groningen, ICS.

Zeit (2003). Krisenzeichen, retrieved 10.12.2013 from http://www.zeit.de/2003/36/Stimmts_Chin__Schriftzeichen

References - Books

Andrews, D. (1984). The IRG Solution -Hierarchical Incompetence and How to Overcome It. Souvenir Press: London.

Ashcraft, M. H. (2005). Cognition. Pearson Prentice Hall.

Campbell, D. T., Stanley, J. C., & Gage, N. L. (1963). *Experimental and quasi-experimental designs for research* (pp. 171-246). Boston: Houghton Mifflin.

Christensen, Clayton M. (1997), The innovator's dilemma: when new technologies cause great firms to fail, Boston, Massachusetts, USA: Harvard Business School Press, ISBN 978-0-87584-585-2.

Constantinides, E. (2013) Social Media Marketing: Challenges and Opportunities in the Web 2.0 Marketplace. In: Consumer Information Systems and Relationship Management: Design, Implementation, and Use. IGI Global, 51 - 73. ISBN 9781466640825

Debray, R., Carnegie, G., Wolnizer, P. W., & Gibson, R. W. (1996). Media manifestos: on the technological transmission of cultural forms. New York: Verso.99–417. Retrievable from: http://monoskop.org/images/d/d4/Debray_Regis_Media_Manifestos_On_the_Technological_Transmission_of_Cultural_Forms.pdf

von Hayek, F. A. (1996). Die Anmaßung von Wissen: neue Freiburger Studien (Vol. 32). JBC Mohr (P. Siebeck).

Schumpeter JA. 1975. Capitalism, Socialism and Democracy. Harper & Row: New York. (Originally published in 1942 by Harper and Brothers.).

US Department of Commerce 1967. Technological Innovation: Its Environment and Management. Washington, DC: USGPO.

Appendix A

Interpreting the Consumer Voice from Consumer Murmur – A Methodological Sketch

1. Requirements

A number of software tools need to be available on a local computer to perform the analysis, for a Windows computer, these include:

- TreeCloud (TreeCloud, 2013), of use for the necessary Textmining of social media discussions on the topic of interest (i.e. an innovation). It can be used to identify the semantic relatedness between prominent words within discussions
- A translation program (i.e. Google Translator), if the discussions are in another language than English
- Access to the SUMO (Suggested Upper Merged Ontology) through the Sigma knowledge engineering environment (Ontologyportal, 2013)
- Table calculation software or other means to compute and assign concept weights

2. Defining Consumer Voice and Consumer Murmur

A consumer voice is an inter-subjective commonality of expressed thoughts and interests of a group of people with a directed buying interest, which is recorded in some (linguistically behavioral) data-base. Put more simple, denoting the consumer voice means answering the question: What do (certain) people want to have?

Any database where large numbers of consumers habitually denote their thoughts on arbitrary topics is reflecting what consumers (or regular citizens) think and say. Meta-data on social media data is especially useful to identify relatively self-consistent consumer voices, because social media have a self-organizing nature: Consumers with similar interests are aggregated within social media, leading to directed or self-consistent talk of the consumer voice by auto-segmented groups of consumers.

The consumer voice of such auto-segmented groups is to be found in the thought patterns between the outerrances of individuals. The common interest of these groups can be interpreted from their accumulated output.

The raw data from which the consumer voice has to be interpreted is called consumer murmur in this paper. Here, several distinctions of gradients of murmur (towards voice) are made, because a metaphoric understanding seems useful to develop a language to use during data analysis.

- Confused consumer murmur over time (chronological)
- Meaningful consumer murmur beyond time (a-chronological)
- Separated low level-phonemes
- Synthesized high level-phonemes
- Weighted phrases on distinct product features of interest
- Fully verbalized consumer voice statements on an interpreted product positioning of interest

I. Procedure

1. Gathering and recycling confused consumer murmur over time

By access to some large scale database directly reflecting or implying expressions of what consumers think, textual data has to be allocated in a single .txt file. Coosto.nl offers .csv downloads of discussions on a particular topic (i.e. an innovation), which can convey up to 10.000 messages (sometimes shortened versions). Coosto messages are ordered chronologically and are not semantically connected with each other in the chronological order, since they come from several websites. This lack of semantic structure may hold true for many big data bases. This lack of mapped semantic relatedness is why the murmur can be called confused at this stage.

Systematic patterns of meaningless or meaning distorting contents of the messages should be identified at first. In the here handled example of Coosto messages, hyperlinks such as “Read on” which direct towards full contents of a shortened message can be deleted. After that, the data is considered recycled. However, the necessity of this step can be evaluated by starting with the next step and looking for suspicious contents of the meaningful consumer murmur, because TreeCloud appears to be rather robust with regard to these distortions.

2. Mapping out meaningful consumer murmur beyond time

Converting the textual output to a .txt file is mandatory. With TreeCloud, the … most frequently mentioned words (i.e. top 50) can be quantified and the semantic distance between the words can be estimated with one of several formulas based on the co-occurance of the words (i.e. normalized google distance). The analysis is performed within sliding windows, which are superimposed upon the text and 'scan' through all of the document for the co-occurances. Which sliding width and steps to choose depends on the nature of the message data and the research goal. The data here consists of multi-sourced messages, which may motivate to choose a small (message specific) sliding window. The current research had the goal to look for inter-subjective semantics, meaning that the boundary of a single message was deliberately omitted to capture the collective intelligence active in the social media data of Coosto. The sliding window has to be small enough to allow the co-occurrances to emerge whilst also choosing a setting that does not result in computational overload and system crash. A width of 100 words and a sliding step of 1 window after each other seemed to yield results with apparently satisfactory efficiency and precision.

It is desirable to experiment with the settings and compare the resulting tree clouds and scout for settings with high mapping-consistency for each new text database.

The TreeCloud software works in interaction with another software, originally of use in bioinformatics to map out phylogenetic trees: SplitsTree. The tree which is send from TreeCloud to SplitsTree to display the calculated TreeCloud. In order to identify meaningfully associated clusters of words in the consumer murmur, certain settings in SplitsTree help out to give a better quick visual overview of these semantically associated topic clusters. For example, a tree cloud can be reorganized to a Phylogram, which as a Cladogram maps out sets of semantically associated words in the text. These sets are meaning full clusters of words or relatively isolated words (which have justified higher weight then the clusters words, because they have semantic relatedness by themselves at a higher level than the cluster words). The bottom level connectedness serves to identify the clusters, because it identifies the most direct relationships between the words. Also this yields a number of relations, which are manageable during the following hand-made computations. However, much more relation information is actually displayed by these Phylograms and can be of use in future more automated analysis procedures.

A crucial characteristic of this meaningful display of consumer murmur is crafted from all messages regardless of their chronological order. This allows to capture the social mediological / collective intelligence, which is at work between the minds of the individual contributors and which, as a deterministic chaos, works beyond chronological time.
Analyzing the data achronological is necessary to derive at a valid prediction of the consumer voice, which should be based on timeless patterns in the Zeitgeist captured by Coosto.

3. Labeling separated low-level phonemes of the voice

The next goal is meaning integration of the meaningful murmur to come to more and more clear wording of what the murmur implies to voice. Each of the clusters and isolated words is translated in a phoneme of what phonemes may have been expressed by the consumer voice through these murmurs.
For that, an integer or entirely consistent set of interpretative semantic relationships is necessary to minimize translation-ambiguity. Using a formalized ontology such as the SUMO as a coding scheme to summarize the clusters meanings and to translate the isolates meanings has the effect of employing a half-automated machine-‘mind’ for the translation process, which with more programming work is fully automatable.
The SUMO coding scheme is quite extensive, because the SUMO has been mapped out to WordNet, which is a very exhaustive lexical-semantic network (or put more simple, a highly sophisticated thesaurus (although it actually is not a thesaurus)). Coding the cluster words or isolated words (isolates) in SUMO can be done by doing a query in the online Sigma knowledge base engineering environment:

- First the word is queried as an English word in Wordnet. Sigma suggests SUMO mappings of the WordNet word. The murmur-word to WordNet-word drop-out rate seems low, so that murmur-words without WordNet word mapping are excluded from further analysis. The low drop-out rate means that the picture emergent in the successive analysis should be based on minimal meaning loss. Besides, some of the drop-outs are linguistically meaningless and are validly excluded from analysis by this mechanism (i.e. sign-sequences found by the TreeCloud co-occurrence formula).
- Next one of the mappings has to be selected. Several choice heuristics guide the selection here: First of all, if a cluster of words is to be integrated, then common-sense similarity of

the suggested mappings with the clusters words is directing the selection towards a mapping, which fits more or less all of the clusters words. Note that this subjective choice can be automated as well in the future, by programming a similarity based selection with a common-sense biased ontology like the DOLCE (Gangemi et al., 2002).

If an isolate is to be categorized, then a mapping which is suggested repeatedly by independent relationships is a promising coding candidate, because a frequently fitting mapping means that SUMO is highly likely to 'think' in terms of this high frequency mapping. In the SUMO ontology, the isolate should have strong association to this mapping.

If several mappings out of the same part of a SUMO hierarchy are suggested (i.e. transportation device or automobile), then the lower order, more concrete mapping is preferred as it conveys more specific meaning. The subsequent iterative meaning integration process employed here is very powerful in reducing the meaning to very few categories. Selection mechanisms which conserve specific meanings for the upper level integrations enjoy preference, because there is the danger of deriving at too broad categories. Specific meanings, which are not shared are automatically abstracted into high order categories by which their influence on the specifics of the consumer voice is diminished. Only few specifics survive the integration process below. Common-sense proves to be a helpful selection heuristic in itself too, if the mappings still are ambiguous. In these cases, common-sense is used to do a probability selection of what seems intuitively reasonable out of the context. Remember that this IS automatable and less subjective than it may sound: Common-sense can be formalized and the SUMO-WordNet mapping context is clearly recorded. This can be the basis for a machine-'intuition' or an educated guess of a to be written software.

- Meaning integration works like this: As all SUMO concepts, the selected mapping is an integer part of the formal SUMO ontology. By clicking on 'graph' in the Sigma knowledge engineering environment and setting the relation to subclass with a high number of classes observed above the selected mapping, it is possible to map out all higher level categories of the selected mapping. A cluster of murmur-words can be integrated into one SUMO category by simultaneously mapping out the higher order categories for all words of the cluster. The cluster is summarized by the single category, which fits for all of the cluster words at once and is at the lowest, most concrete level possible. SUMO is a consistent ontology, which means that there is no cluster without a (high order) summarizing category, even if that is the all summarizing category named ‘entity’. SUMOs consistency and high abstraction potential is why the meaning integration process employed here is very powerful.

4. Synthesizing high-level phonemes of the consumer voice

- Meaning integration of the now summarized clusters and isolates can be repeated over and over again to come at as abstract categories as is wished. However, because using SUMO the way as described here is a very powerful integration mechanism, one may be interested in a summary of labels which is concrete enough to be understandable. If things go bad, then the resultant category is ‘entity’ which is not really a practical prediction category. The other extreme is that too many isolate specifics survived the integration process, which makes re-integration of the isolates meaning according to the TreeCloud relationships advisable. However, keeping the positioning statement relatively lengthy (ending up at 4-7 categories) seems to help to derive at an output which can be interpreted by humans as a product positioning statement. Shorter positioning statements can be too abstract to be interpretable.

5. Finding weighted phrases on distinct product features

With the meaning integration process finished at the phoneme level (or product features level), the next step is to find out what features are more important constituents of the positioning statement predicting the (innovation) category. Identifying the weights and classifying phrases of the product features is the last preparatory step for the innovation prediction.

In a nutshell, the weights can be estimated as proportion of the feature category (high-level phoneme) from the total number of categorizations (the number of all implied ‘entity’ categorizations). To compute these weights, it is advisable to start with the highest categorizations and work downwards to more specific ones. First one can count all direct categorizations in the highest level categories. The highest three levels seem to cover a lot in many cases. Then one can go to the lower ordered categorizations and add them to the higher levels one by one and count the surviving specific categories at their lowest non-overlapping level. This has to be repeated until all isolate categorizations are partitioned and the higher level overlap has been quantified.

The result of this process is usually a lot of summarizing abstract categorizations and some more specific surviving isolate categories. The total weighting of the surviving isolates has defining influence on the innovation category, because the higher-level, pre-partitioning categories all exert equal influence on the weights of the surviving isolates. In a way, the higher level, pre-partitioning categories collapse towards the direction of semantic meaning by the small but decisive pushings and pullings of the surviving isolate categories. The reason for this is, that although the higher-level pre-partitioning categories do have higher weight then the isolates, they do not exert the critical bias of the innovation concept towards a specific direction in semantic space. Although they have higher proportional weight in the integrated concept, they do have less defining influence, because of their highly integrated nature (i.e. a lot of things can be classified as physical entities). Some surviving isolate categories overlap, increasing their definitional weight over-proportionally. Most stay isolated.

In the end, a list of weighted product feature categories assorted around more abstract categories emerges:

Entity	21		
Physical	8	Abstract	6
	0.38095238		0.28571429
Object	4	Quantity	4
Automobile	2	time duration	1
electric		positive real	
plugin	1	number	1
object	1	number	1
Process	4	Attribute	1
Radiating	2	relational attribute	1
Process	1		
Motion	1		
		Physical Quantity	1
		length measure	

Total Weights per subcategory

of objects	0.19047619	of quantity	0.19047619
of process	0.19047619	of attribute	0.04761905
		of physical quantity	0.04761905

6. Verbalized consumer voice product positioning statement

The constitutive features above have been weighted. In the last step, the features are grouped together according to their weighting. The most important features are grouped together as core product features, those with mid-level importance are expected features and lower-level importance features are augmented product features. In each of these groupings, the most important features are named first and the latter features are conceptual completions. Bringing these features together in one sentence is the positioning statement interpreted as the consumer voice. It has the following standard format:

Innovation Positioning Statement:

A … with … (core features) that is expected to … and to … (expected features) and is augmented by being … and … (augmented features).

Appendix A References

Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., & Schneider, L. (2002). Sweetening ontologies with DOLCE. In Knowledge engineering and knowledge management: Ontologies and the semantic Web (pp. 166-181). Springer Berlin Heidelberg.

Ontologyportal (2013). http://www.ontologyportal.org/ access to Sigma via http://sigma-01.cim3.net:8080/sigma/Browse.jsp?kb=SUMO&lang=EnglishLanguage

TreeCloud (2013). http://www2.lirmm.fr/~gambette/treecloud/

Appendix B

Innovation Name	Radicalness rationale	Innovation evidence
ipad	Another Apple example for transforming an existant radical innovation to a discontinuous innovation taking large market shares and delivering for a large segments	http://www.tuaw.com/2011/04/20/ipad-held-85-market-share-in-2010-acco
Chevrolet Volt = Opel Ampera	Radicalness was estimated for the Chevrolet Volt (which is the same car), because only the Chevy Volt seems to be present on the web in degrees allowing to compute GSD. The Opel Ampera is the first and most successfully marketed serial-plugin hybrid in the Netherlands qualifying it for a discontinuous innovation	http://media.opel.com/media/intl/en/opel/news.detail.html/content/Pages/
General Atromics Military Drones	Innovation in military strategy radically changing the way war is prevented or conducted by the US world police army	http://en.wikipedia.org/wiki/Drone_attacks_in_Pakistan

Whats app	Smartphone App that made SMS obsolete and replaced it	http://gigaom.com/2013/04/29/chat-apps-have-overtaken-sms-by-message-
Dacia Duster	First car with a low cost-high value strategy in the 'premiumish' SUV segment bearing market success with radical focus on technological necessities (having an impact on future technology trajectories by stimulating the market to focus on core technologies)	http://www.autoevolution.com/news/dacia-eu-s-fastest-growing-brand-in-m http://www.google.de/url?sa=t&rct=j&q=&esrc=s&source=web&cd=4&ved=(d.bGE
PlanetSide 2	One of the first commercially succesful free-to-play games, radical in the sense that it is one of the first examples of commercially viable implementation of the free-to-play technology	http://www.vg247.com/2012/12/13/planetside-2-doing-better-than-other-ti
Google Apps	starts to replace locally installed, MS office by taking market share of MS	http://rcpmag.com/articles/2013/04/23/google-apps-vs-microsoft-office.asp:
Apple App store / Mac app store average	Both scores are averaged, the app store technology was unpredecendet and a direct commercial hit	http://news.cnet.com/8301-13579_3-20032012-37.html

Instagram video sharing	first mobile device compatible app (technology) for instant picture and video sharing	http://blog.appboy.com/2010/10/5-things-instagram-got-right-that-others-b
android operating system	first for free smartphone operating system (charged with Googles enhanced rights on information) took large market shares in a few years	http://arstechnica.com/gadgets/2011/01/android-beats-nokia-apple-rim-in-2

coosto: september 21 2012 uitbreiding naar GB				
http://www.frankwatching.com/archive/2012/07/30/social-media-tools-van-monitoring-via-dashboards-tot-socially-engaged/				
social media monitoring service	0.46303553	0.12566013	0.03549502	
tracebuzz best reputation nederland	http://www.topseos.co.nl/tracebuzzcom			
Hootsuite	http://de.slideshare.net/UpstreamStrategies/webcare-in-nederland-een-quickscan			
buzzcapture				
Clipit				
tesla s	0.07386927	0.16875704	0.06602438	
http://nl.wikipedia.org/wiki/Tesla_Model_S	https://twitter.com/teslatrends/status/365429053370544129	in usa öfter verkauft als sklasse	http://www.autozine.nl/overzicht/autoverkopen.php?mok=1666	http://www.telegraaf.nl/autovisie/autovisie_nieu el_S_per_week_geassembleerd_in_Tilburg___.htr
zombies, run!	0.02561394	0.07277598	0.1484864	
http://en.wikipedia.org/wiki/Zombies,_Run!				
e-book	0.19260007	0.17914206	0.3149165	

http://www.digitaltrends.com/mobile/sales-of-digital-goods-growing-fast-good-news-for-amazon/	http://www.publishersweekly.com/binary-data/ARTICLE_ATTACHMENT/file/000/000/522-1.pdf	http://www.futurebook.net/content/predictions-dutch-ebook-market-2012	
geldvoorelkaar.nl	0.59575905	0.78791482	0.37909174
http://www.crowdfunding.nl/links-test/	https://www.graydon.nl/blog/article/2013/10/03/crowdfunding-wordt-in-snel-tempo-volwassen		
hydraulic fracturing bad example	0.62260235	0.38454109	0.23821185
cloud software as service for companies or governments			
amazon as IAAS			
salesforce as PAAS			

s klasse 2013	0.23762784	0.01069091	0.13876654	
s-klasse as first autonomous vehicle				
new materials used in products (				
automation of knowledge work	http://dupress.com/articles/the-future-of-knowledge-work/			

iphone 4 siri iphone 4s	http://www.redmondpie.com/most-successful-iphone-launch-ever-iphone-4s-was-pre-ordered-over-a-million-times-in-first-24-hours/ 0.52084723	0.21859914	0.60667514
dropbox software	0.51052991	0.32170872	0.62944885
shell offshore technology	0.34738843	0.39151275	0.39150664
twitter	0.14024859	0.24764632	0.3708217

zynga games	0.0209635	0.02924507	0.00033834

taleo corporation	0.15056166	0.58000615	0.05605699
correlation ventures	0.35066822	0.31300556	0.31421419
blue prism	0.06287107	0.1994005	0.11730557
Measurement Incorporated	0.66310921	0.4511731	0.31997613
Blackstone Electronic Discovery & Legal Support Services	0.58222231	0.31940069	0.15670355

Ernst & Jung ediscovery	0.54776017	0.26325861	0.08714072
iRobot Packbot	0.48043313	0.33017774	0.21636994
Nissan Leaf	0.58778265	0.32845901	0.16792726

iRobot Roomba	0.68444325	0.48592826	0.3630395
L3 Communications bodyscanner	0.67933641	0.4768115	0.35222573

Wonderbook: Book of Spells	0.59416308	0.33784436	0.17986184
Gambitious crowdfunding	0.55749015	0.42555717	0.29936223
League of Legends	0.52090706	0.21832126	0.03182194
Brick Force	0.59166136	0.3337626	0.17480622
S4 League	0.59836267	0.34469632	0.1883486
IBM Connections	0.22015181	0.36333702	0.21143675
Kudos Badges	0.43903894	0.54203505	0.43277002
Huawei Ascend D	0.61154291	0.36620095	0.21498398
Huawei Ascend G615	0.43461965	0.53842717	0.42830134
samsung galaxy	0.46133304	0.03971655	0.60221225
smart tv			

Samsung Smart Hub	0.13309329	0.29176844	0.12532753
Smart Viera	0.2637381	0.39850055	0.25714266
LG Smart TV	0.04808084	0.22231633	0.03955354
Philips Smart TV	0.2016509	0.34777754	0.19449929
sunis indoor wirefree	0.22032349	0.65372586	0.1654496
kinect	0.49863172	0.54270699	0.56357913
gamification	0.5488368	0.70015119	0.60145071