Business Model Clustering: A network-based approach in the field of e-mobility services

Christian Engel[a], Jan Haude[a], Niklas Kuehl[a]

[a] *Karlsruhe Institute of Technology*

Abstract

Empirical insights about business models in the field of e-mobility services are of high importance to academia, industry and politics. As basic clustering algorithms do not deliver semantically valuable findings on business model structures based on obtained empiric data, this paper proposes a similarity measure-based network approach of clustering the latter. On the basis of graph, social network and similarity measure theory, an approach is designed which compares every business model instances of a data set with each other. The paper comes up with a matching score in order to determine whether two business models are connected contentwise within a cluster or not. The plotting of the resulting matching scores leads to a visually based determination of a meaningful matching score which bonds two business models together or not. The elaborations result in four e-mobility service clusters: Data-and-software-driven-, brokering-, transportation- and energy supply-based business models. Additionally, further findings on current opportunities in clustering business models and future solution proposals are described.

Keywords: Energy and Mobility Services, E-mobility, Business Models, Clustering

Email addresses: christian.engel2@student.kit.edu (Christian Engel), jan.haude@student.kit.edu (Jan Haude), kuehl@kit.edu (Niklas Kuehl)

1. Introduction

In the last decade, the importance of an exit from nuclear and fossil-fuel energy concepts as well as sustainable mobility has become obvious among science, industry and politics. One important aspect of this shift is *e-mobility*, holistically defined as "*a highly connective industry which focuses on serving mobility needs under the aspect of sustainability with a vehicle using a portable energy source and an electric drive that can vary in the degree of electrification.*"(Scheurenbrand et al., 2015, p. 25) Besides traditional research areas of e-mobility–like battery technology, ICT, manufacturing, etc.–innovative business models and complementary mobility services are important for the success and acceptance of electric vehicles (Stryja et al., 2015a; Hinz et al., 2015). A framework for capturing and describing business models for e-mobility services was presented in Stryja et al. (2015b) and continued in Kuehl et al. (2015). This framework (cf. figure 1) delivers a scientifically created and practically validated tool for registering the essences of business models. As a next step, we would like to identify clusters of empiric data from existing projects, who could submit their business model via the web platform `e-mobility-atlas.de`. These clusters may support interested parties to find "types" of typical e-mobility service business models and identify gaps as well. This may be of help to researchers, who struggle with e-mobility services because of their high heterogeneity. It also delivers results for practitioners to see in which combination existing concepts occur–and what might be missing. A first attempt to cluster business models on the basis of a k-Means clustering algorithm was presented in Kuehl et al. (2015). One outcome was that when applying commonly used clustering algorithms the outcomes are not semantically meaningful. This may have the following reasons: As (Beyer et al., 1999, p.217) concludes: "[A]s dimensionality increases, the distance to the nearest data point approaches the distance to the farthest data point"–so typically-used distance measures are not meaningful anymore. Additionally, most clustering algorithms will assign all observations to a cluster–even outliers. One way to eliminate these outliers is outlier detection (cf. Aggarwal and Yu (2001)), but it would remove actual observations from the (already small) data set which are no errors. Also, the high number of features (characteristics of business models) makes the clustering more complex. In order to face this challenge a feature reduction is possible, but it only improved the meaningfulness of the clusters fractionally (as shown in Kuehl et al. (2015)). To continue with our research, the paper at

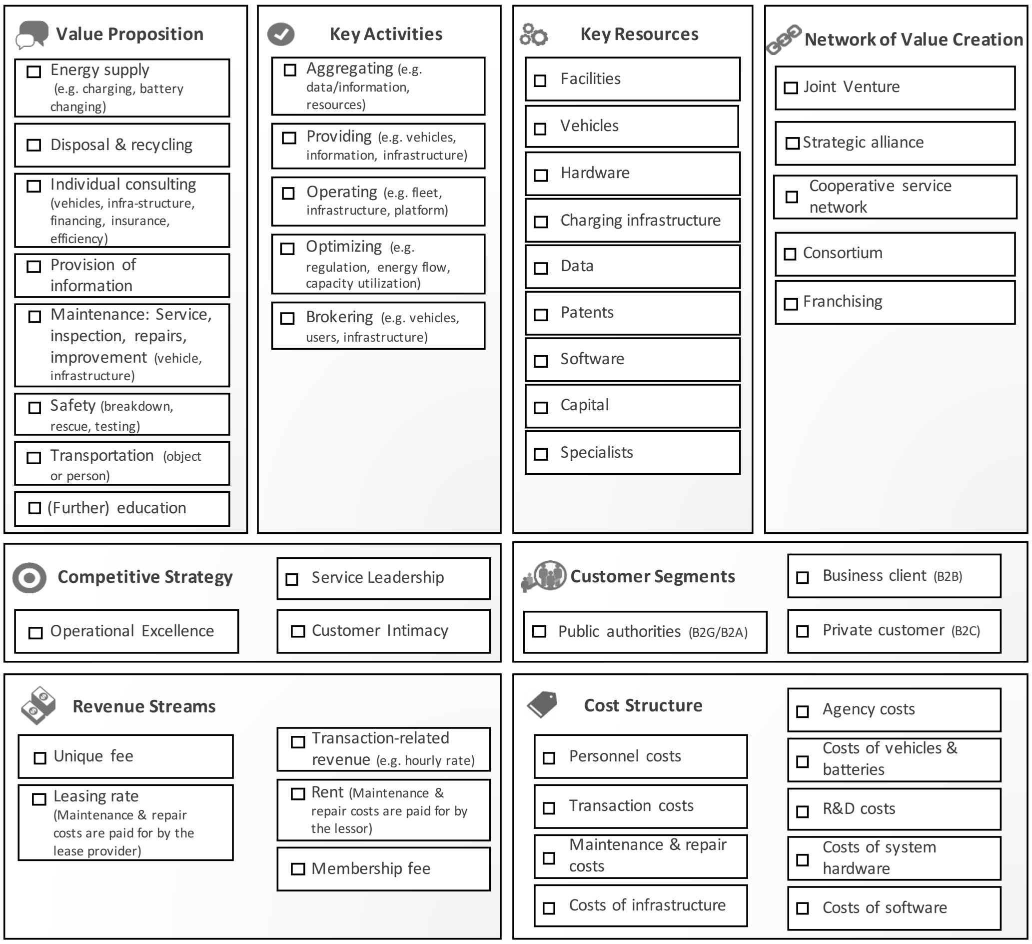

Figure 1: E-mobility specific service business model framework

hand aims at answering the following three research questions:

1. Which clusters are identified by applying a similarity network approach?
2. How do these results compare to a baseline of clusters from previous research?
3. What are relevant insights of e-mobility service business models for politics, academia and industry?

Its contribution is threefold: At first, we acquired a larger data set (n=40) of business model instances than in Kuehl et al. (2015). Secondly, we explain, apply and interpret a different approach of clustering and thirdly we are able to deliver new empiric insights of e-mobility service business models.

2. Methodology

As mentioned before, most commonly used algorithms for clustering have several restrictions in providing meaningful clusters for our problem. As meaningful and interpretable results are the key factor for our research, we need approaches which are customized to our specific needs. As a prerequisite we set the following definitions: The observations are called *Business Model Instances (BMIs)*, the attributes are called *characteristics*. Both dimensions are indexed as shown in figure 2. The indexing is based on the concept of

$$
\begin{aligned}
& C = \{c_1; c_2; \dots c_j \dots; c_p\}, && \triangleright \text{ Business Model characteristics} \\
& j \in J = \{1, \dots, p\} && \\
& B = \{b_1; b_2; \dots b_i \dots; b_n\} && \triangleright \text{ Business Model Instances} \\
& i \in I = \{1, \dots, n\} && \\
& x_{ij} \in X : I \times J \to \{0; 1\} && \triangleright \text{ Occurrence (0 or 1)}
\end{aligned}
$$

Figure 2: Prerequisites

Boolean matrix entries which are used in this work to present a bipartite network, also called "two-mode network" (Salton et al., 1983, 1975; Opsahl, 2013). For instance, the survey participants of the first BMI can characterize it along the dimensions $(x_{11} \dots x_{1p})$. This means that $x_{11} \dots x_{1p}$ can either take values of 1 or 0. Thus, all of the n business models can be represented through a Boolean vector. This transformation helps to make the business models comparable.

This paper proposes an approach of clustering business models within a similarity measure-based node network. It combines two classic approaches which are often applied in literature: The concepts of *similarity measures* and the concept of *structuring relations within node networks.*

2.1. The concept of similarity measures

In order to compare the BMIs, this paper uses similarity measures (Choi et al., 2010). A binary similarity measure based on the Jaccard coefficient is used to say “how similar” the compared business models are (Choi et al., 2010; Cheetham and Hazel, 1969). If the business model instance Y and Z are compared, the similarity score between them is calculated as follows:

$$\text{Matching Score}(Y, Z) = \frac{\sum_{d \in D} \frac{|Y_d \cap Z_d|}{|Y_d \cup Z_d|}}{|D|}$$

The Jaccard value as a binary measure for similarity (Choi et al., 2010; Cheetham and Hazel, 1969) is calculated within each characterizing category of the vectors Y and Z of the business models and then normalized along the number of categories. This measure is called ”Matching Score” in this paper. D is the set of all high level categories of characteristics while d is one category within D. As shown in figure 1, d could examplarily be ”Key Activities”, and thus, the characteristics would be ”Aggregating”, ”Providing” etc.

In order to get an understanding of how each of the business models can be compared to the others and to find out the best-matches (the most similar business models) and the worst-matches (the most distinct business models), every business model vector has to be compared to all other business model vectors and the particular matching scores have to be calculated. If n is equal to the number of observed BMIs, this means that $\frac{n*(n-1)}{2}$ matching scores are calculated. We conduct this approach in order to find out how the similarity measures within the business model context behave and what a meaningful matching score which acts as a threshold should look like to speak of two business models being similar (Cha et al., 2005). In the following we speak of two BMIs being similar or related if their matching score is higher than or equal to the “meaningful matching score”. The meaningful matching score can be inferred by plotting the down-ranked matching scores within a figure and visually analyzing the curve. In this paper the meaningful matching score (=threshold) is set at the point bewteen disjunctive and non-disjunctive clusters. This matter-of-fact is explained in detail within the results section.

2.2. The concept of structuring relations within node networks

In order to apply distinct metrics to calculate valid business model clusters, this approach focuses on representing BMIs as nodes within a network. This approach originates from the research field of computer sciences to structure and view complex data sets and to retrieve new information out of it (Bondy and Murty, 1976; West et al., 2001). Nowadays, a vast modern research stream, called *social network analysis*, uses the network modeling approach to infer information about network structures like clusters and its participants (Scott, 2012; Hanneman et al., 2001). We try to apply the methods and tools of network analysis on our business model case: On the basis of the "meaningful matching score" which is illustrated within the results section, it can be inferred whether two nodes are connected or not: Only if two business models reach above the meaningful matching score, a non-directed relation is established between the particular nodes. One of the most-used metrics which can be derived from a network like this is the "degree" of a node which tells with how many other nodes the particular node is connected to (Bollobás, 1998). In the following we use the node degree to identify the most prominent nodes of the network and use them as the representatives of distinct clusters of the network (Scott, 2012; Kempe et al., 2003). This approach is based upon the so-called principle of "one-mode projection" which is highly prominent in Network Analysis literature (Zweig and Kaufmann, 2011). The tool to visualize the resulting business model networks which serves as the basis for the further analysis is called *NodeXL* and has been developed by scholars from the United States in cooperation with Microsoft Research (Smith et al., 2009).

2.3. The similarity network methodology at one sight

Summarizing the above explained methodology leads to the following structured order:

1. Indexing: Representing business models within a Boolean vector
2. Calculation of the matching score for every business model pair
3. Visual identification of the "meaningful matching score"
4. Representation of the business models within a node network according to the distinct matching scores
5. Calculation of the node degree in order to identify clusters and their particular representatives

3. Results

The methodology from the previous chapter is now applied on the data set. The data set consists of 40 invited e-mobility projects, which were chosen for

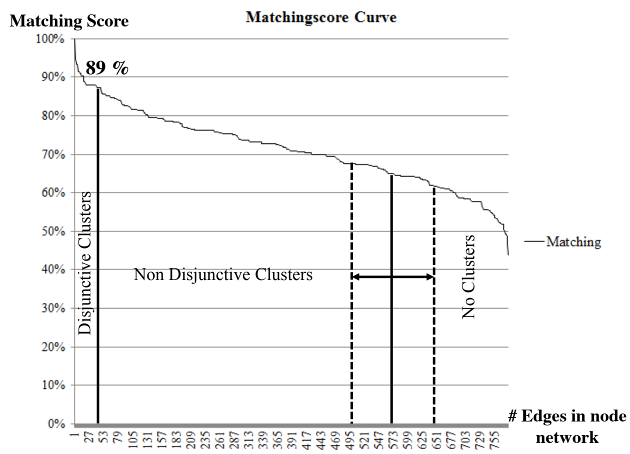

Figure 3: Matching score curve

their diversity and their service focus to submit their business model through an online tool (`e-mobility-atlas.de`).

3.1. The Matching Score Curve

As n equals 40, $\frac{40*(39)}{2} = 780$ comparisons are conducted, the particular matching scores are calculated and ordered decreasingly. Figure 3 shows the resulting matching score curve. The number of comparisons can be interpreted as the number of edges in a resulting node network consisting of BMIs. As figure 3 depicts, three areas of distinct cluster types can be derived from the analysis: A disjunctive clustering area, a non-disjunctive clustering area and an area of no clusters. The iterative visualization of the node network with a changing number of nodes according to the matching score shows that below a value of 89 percent in this data set no more disjunctive clusters can be found by performing the methodology proposed in section 2. Below a value of 89 percent all nodes of the network are connected to one big component (cf. figure 5).

Therefore, this part is called non-disjunctive clustering area. Identifying the lower boundary below no more clusters (disjunctive and non-disjunctive) can be found will be a future research focus, but is not of interest for this paper as the focus is put upon finding disjunctive clusters. Therefore, only business model pairs which reach above a ”meaningful matching score” of 89 percent are considered.

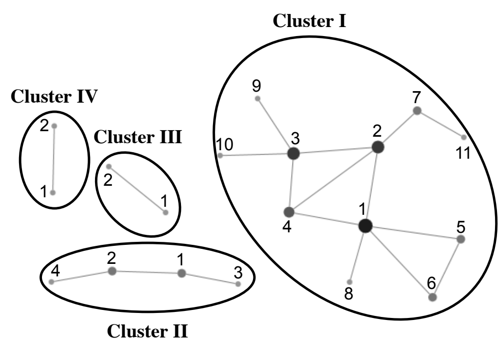

Figure 4: Disjunctive e-mobility clusters based on the chosen approach

3.2. Resulting Clusters

From the data set of 40 business models four disjunctive clusters of 19 BMIs can be found. Figure 4 illustrates the node network. The node size depends on the particular node degree. The higher the degree, the bigger the node. The nodes in each cluster are numbered and matched to the particular business model characteristics which are depicted in the appendix in table 1. The business model node with the highest node degree can be seen as a representative for the cluster. The four e-mobility business model clusters which are found in this paper can be named as following:

- Cluster I: Data-and-software-driven services
- Cluster II: Brokering vehicles or specialists
- Cluster III: Energy services
- Cluster IV: Transportation services

3.3. Non-Disjunctive Clusters

As elaborated above, we only analyze disjunctive clusters in this paper. Nonetheless, a short wrap-up of the non-disjunctive clustering problem shall be given. As figure 5 shows, below a value of 89 percent all nodes within the network are connected to one component and with a decreasing matching score the number of edges within the resulting business model networks is rapidly increasing. When the matching score reaches a value of 44 percent, all nodes are connected with each other. Due to reasons of visualization the

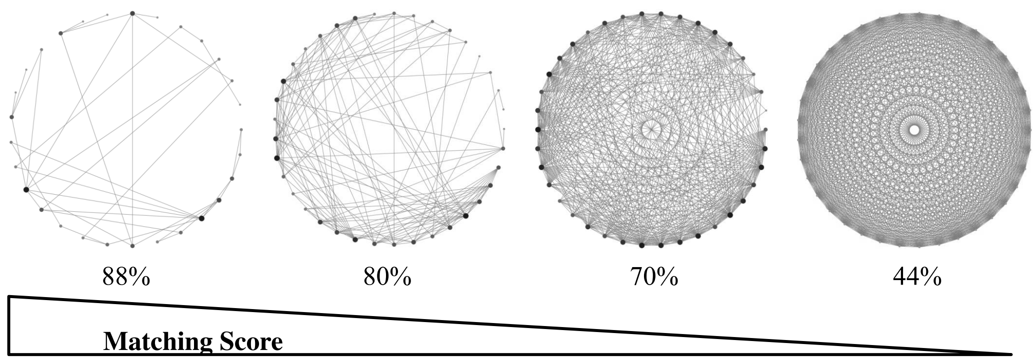

Figure 5: Business model networks in dependence of distinct matching scores

networks are depicted as circles. This makes it easier to see how the network changes with a decreasing matching score. This consequently means that the setting of the matching score requirement in the non-disjunctive clustering area has got huge impact on the clustering results. Thus, it could be of further scholarly interest to develop ways of appropriately setting the matching score and clustering in the non-disjunctive clustering area. This could lead to more clusters than only studying the disjunctive clustering area and consequently reveal further insights concerning business model research.

3.4. Comparison to baseline

As a baseline we use the clustering approach from Kuehl et al. (2015) with the larger data set (n=40) with a minimum number of desired BMIs per cluster of 4 ($k_{min} = 4$) and a minimum number of characteristics per cluster of 3 ($g = 3$). The approach results in two clusters, which are almost identical to the results from the smaller data set:

- Cluster I: In this cluster we identify four different BMIs and corresponding projects after five iterations, all of them therefore sharing at least five commonalities. Their value proposition is the *Provision of information*, the key resources are *Data* and *Software*, and the key activities are *Providing* and *Aggregating*.
- Cluster II: In this cluster we identify five different BMIs and corresponding projects after three iterations, all of them sharing at least

three commonalities. Their key activity is *Operating*, their value proposition is *Transportation* and their key resources are *Vehicles*.

The main structural difference between the two approaches is the following: The frequences-algorithm (Kuehl et al., 2015) faces the prerequisite of the characteristics being disjunctive while the similarity measure-based approach is working on the basis of disjunctive BMIs. The results shown in table 1 from the approach presented in this paper find two more clusters than the baseline and also assign almost twice as many BMIs over all clusters. Moreover, cluster 1 in the baseline is a proper subset of cluster I presented in this paper in table 1 and the second cluster in the baseline is a subset of cluster IV in this paper. So even though two elementary different approaches were chosen, both methods come to similar results, which shows that there is no contradiction, but that these results support each others hypothesis. The approach at hand in this paper is more detailed, so depending on the interest of the researcher, (s)he may choose between an easy approach with restrictive input parameters or a more complex approach with more sophisticated clusters.

4. Conclusion and Outlook

This paper proposes an approach for clustering business models within an e-mobility service context. Summarizing the steps which have been taken to achieve this goal, the following can be stated: On the basis of treating business models as Boolean vectors which describe the characteristics of distinct business model instances (BMIs) and make them comparable, a pairwise comparison grounded on network, graph and similarity measure theory is conducted. This approach leads to four clusters of e-mobility services: Data-and-software-driven services, brokering vehicles or specialists, energy services and transportation services. The fundamental insights which are gained during the taking of the methodological steps are that disjunctive and non disjunctive clustering areas exist within the network-based approach. These areas have to be treated in different ways whilst this paper concentrates on elaborating the former. Concerning a wider matter of interest, this work delivers politically and economically relevant insights into the structure of current efforts of deploying e-mobility within a socially accepted and profit gathering manner. It is indisputably clear that in the e-mobility projects which are analyzed in this paper the main focus is put upon leveraging data, software and specialists in order to transfer the idea of e-mobility to common sense. Additionally, services concerning the e-mobility-based transportation and energy supply are in focus of current national study efforts in a business model context. These clusters should be taken as a starting point for both, further studies in this field and for business people for entering the so-called "white spots" which are not displayed in the current e-mobility service spectrum.

As limiting aspects of this work several points have to be adressed. The first restrictive point is that the Jaccard coefficient, which is part of the matching score formula, overestimates the non-selected (=0 or false) characteristics within the business model vectors because there are more characteristics which are not "fulfilled" in the particular business models than characteristics being applied in the business model. This is the reason for 44 percent being the matching score which connects every node with each other (cf. figure 5). It could be of future scholarly use to adjust the existing and study other kinds of matching score calculations and compare the results to the findings in this paper. As this phenomenon occurs in each BMI vector comparison, the results of this approach are nonetheless valuable. Concerning alternative similarity measures, the similiarity measure presented by Sohn (2001) could be an interesting starting point for further research, for example. Furthermore, dealing with disjunctive networks and the setting of a lower boundary (below it no more clusters are found) should be in the focus of further research in this field in order to eliminate the above named inaccuracy and to broaden the spectrum of identifiable business model clusters. At the moment, the node degree is the only graph theoretical network measure which is in focus of the elaborations. In future works it could be enlarged by adding further measures like network centrality measures and distinct clustering coefficients for networks.

At the moment the data set which was retrieved by the project DELFIN consists of 40 projects dealing with e-mobility business models. This number of analyzable projects will be increased in near future and offer the opportunity to apply the above mentioned prospective proposals in order to substantiate the findings of this paper and to receive new results concerning the fields of business model clustering and e-mobility services.

Acknowledgements

This paper has been written in the context of the research project *DELFIN*. The project is funded by the German Federal Ministry of Education and Research (BMBF) under the promotion sign 01FE13002. We also thank the project management agency German Aerospace Center (PT-DLR) for the project support.

References

Aggarwal, C. C. and Yu, P. S. (2001). Outlier detection for high dimensional data. *ACM SIGMOD Record*, 30(2):37–46.

Beyer, K., Goldstein, J., Ramakrishnan, R., and Shaft, U. (1999). When is "nearest neighbor" meaningful? *Database Theory—ICDT'99*, pages 217–235.

Bollobás, B. (1998). *Modern graph theory*, volume 184. Springer Science & Business Media.

Bondy, J. A. and Murty, U. S. R. (1976). *Graph theory with applications*, volume 290. Macmillan London.

Cha, S.-H., Yoon, S., and Tappert, C. C. (2005). Enhancing binary feature vector similarity measures.

Cheetham, A. H. and Hazel, J. E. (1969). Binary (presence-absence) similarity coefficients. *Journal of Paleontology*, pages 1130–1136.

Choi, S.-S., Cha, S.-H., and Tappert, C. C. (2010). A survey of binary similarity and distance measures. *Journal of Systemics, Cybernetics and Informatics*, 8(1):43–48.

Hanneman, R. A., Riddle, M., and Robert, A. (2001). Social network analysis. *Riverside: University of California*, pages 1–154.

Hinz, O., Schlereth, C., and Zhou, W. (2015). Fostering the adoption of electric vehicles by providing complementary mobility services: a two-step approach using Best–Worst Scaling and Dual Response. *Journal of Business Economics*, pages 1–31.

Kempe, D., Kleinberg, J., and Tardos, É. (2003). Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM.

Kuehl, N., Walk, J., Stryja, C., and Satzger, G. (2015). Towards a service-oriented business model framework for e-mobility. In *Proceedings of the European Battery, Hybrid and Fuel Cell Electric Vehicle Congress*, Brussels, Belgium.

Opsahl, T. (2013). Triadic closure in two-mode networks: Redefining the global and local clustering coefficients. *Social Networks*, 35(2):159–167.

Salton, G., Fox, E. A., and Wu, H. (1983). Extended boolean information retrieval. *Communications of the ACM*, 26(11):1022–1036.

Salton, G., Wong, A., and Yang, C.-S. (1975). A vector space model for automatic indexing. *Communications of the ACM*, 18(11):613–620.

Scheurenbrand, J., Engel, C., Peters, F., and Kuehl, N. (2015). Holistically Defining E-Mobility: A Modern Approach to Systematic Literature Reviews. In *5th Karlsruhe Service Summit*, Karlsruhe, Germany.

Scott, J. (2012). *Social network analysis*. Sage.

Smith, M. A., Shneiderman, B., Milic-Frayling, N., Mendes Rodrigues, E., Barash, V., Dunne, C., Capone, T., Perer, A., and Gleave, E. (2009). Analyzing (social media) networks with nodexl. In *Proceedings of the fourth international conference on Communities and technologies*, pages 255–264. ACM.

Sohn, M.-W. (2001). Distance and cosine measures of niche overlap. *Social Networks*, 23(2):141–165.

Stryja, C., Fromm, H., Ried, S., Jochem, P., and Fichtner, W. (2015a). On the Necessity and Nature of E-Mobility Services – Towards a Service Description Framework. *Proceedings of the International Conference on Exploring Service Science 1.5.*, 201(Lecture Notes in Business Information Processing):109–122.

Stryja, C., Schueritz, R., Kuehl, N., Hottum, P., and Satzger, G. (2015b). Entwicklung eines Frameworks zur Beschreibung von Geschaeftsmodellen fuer Elektromobilitaetsdienstleistungen. In *Konferenzband der 9. Internationalen Energiewirtschaftstagung (IEWT)*, Wien, Austria.

West, D. B. et al. (2001). *Introduction to graph theory*, volume 2. Prentice hall Upper Saddle River.

Zweig, K. A. and Kaufmann, M. (2011). A systematic approach to the one-mode projection of bipartite graphs. *Social Network Analysis and Mining*, 1(3):187–218.

Appendix

Cluster	BMI	Node Degree	Value Proposition	Key Activities	Key Resources
I	1	5	Provision of information	Aggregating	Data
	2	4	Provision of information	Aggregating, operating	Data, software
	3	4	Individual consulting, provision of information	Aggregating, providing, operating	Data, software
	4	3	Provision of information	Aggregating, providing	Data, software
	5	2	(Further) Education	Aggregating	Data, specialists
	6	2	(Further) Education, provision of information, safety	Aggregating	Data
	7	2	Provision of information	Operating	Software
	8	1	Provision of information	Brokering	Data
	9	1	Individual consulting, provision of information, transportation	Aggregating, providing, operating	Data, vehicles, charging infrastructure, software
	10	1	Individual consulting, provision of information	Aggregating, providing, operating, optimizing	Data, software
	11	1	Transportation	Operating	Software
II	1	2	Transportation	Brokering	Vehicles
	2	2	(Further) Education	Brokering	Specialists
	3	1	Transportation	Providing, brokering	Vehicles
	4	1	Safety	Brokering	Specialists
III	1	1	Energy supply, provision of information	Operating, brokering	Data, software
	2	1	Energy supply, provision of information	Operating, brokering	Supplying customers, software
IV	1	1	Transportation	Providing, Operating	Vehicles
	2	1	Transportation	Providing, Operating	Vehicles

Table 1: Overview of the resulting disjunctive business model clusters and their characteristics (BMI enumeration can be compared with figure 4)