Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 180 (2021) 172–179

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Industry 4.0 and Smart Manufacturing

The association between network centrality measures and supply chain performance: The case of distribution networks

Christian Wallmann[a,b]*, Markus Gerschberger[a,b]

[a]Josef Ressel Centre for Real-Time Value Network Visibility, Solarstr. 7, 4653 Eberstalzell, Austria.
[b]Institute/Department for Supply Chain Management, University of Applied Sciences Upper Austria, Wehrgrabengasse 1-3, 4400 Steyr, Austria.

Abstract

We analyze transport data on the worldwide distribution network for 1.2 million vehicles manufactured and distributed by a large German car manufacturer in half a year. To identify central nodes in this network, we calculate various centrality measures from social network analysis. We then analyze the association of these centrality scores and the key performance measures related to stay-times and inventory for ports, distribution centers, and plants. Our main result shows that nodes with high degree centrality perform worse than less central nodes. The main theoretical contribution of our research is to confirm for the very first time that network theory applies to distribution networks, i.e. that network structure influences network node performance.

Keywords: Social network analysis; performance; supply chain management; network theory

1. Introduction

Given the increase of value creation dispersion in supply chains, distribution networks are highly connected nowadays and consisting of a high number of network nodes. Multinational companies produce in many different countries and deliver into many different markets in many different countries. Shipments of finished goods pass through networks consisting of several plants, ports, hubs and other network nodes. In these distribution networks, multiple modes of transports and agents are involved. Therefore, it is reasonable to hypothesize that the company's

* Corresponding author. Tel.: +43 5 0804 33286; fax: +43 5 0804 33299.
E-mail address: christian.wallmann@fh-steyr.at

10.1016/j.procs.2021.01.153

distribution performance does not only depend on internal capabilities, but also on the structure of the distribution network as well as the performance of all involved network actors. The claim that structural embeddedness of nodes in networks is relevant to its performance is well known as “network theory” [1]. Network theory is now well confirmed and established for supply networks. However, there is no prior research whether the structural embeddedness of nodes is relevant to its performance in distribution networks too. We hypothesize that testing network theory in the context of distribution networks is largely lacking due to the availability of appropriate data. While for supply networks, there are some external data sources like Bloomberg, Mergent or Compustat [2,3], outbound transportation data is not publicly available. Lack of research in several interesting research areas may be explained by observing that companies are often reluctant to share data [4]. However, once available, combined company internal data is more reliable and complete than external data. In addition, company internal data on its inbound supply chain or its multi-tier network of parts supplier will often be more incomplete than distribution data. A more detailed assessment of network theory may therefore result from studying company internal data on the distribution network rather than studying company internal data on the supply network.

The purpose of this paper is to address this research gap by evaluating the hypothesis that more central nodes in distribution networks perform worse than less central nodes. To test the hypothesis that network centrality of nodes in distribution networks matters to their performance, we analyze data on the large distribution network of a major German car manufacturer (OEM) directly obtained by the OEM and apply social network analysis to it. The data contains network flow data for half a calendar year. Movement of 1.2 million shipments from over 15 plants to 2,000 dealers via 350 network nodes was gathered.

The main theoretical contribution of our paper is first confirming the applicability of network theory in the context of distribution networks. Since our result is based on large and high quality company internal data, we surpass earlier research on network theory by confirming it based on high quality data, thereby reducing bias.

The paper is organized as follows. In Section 2, we review the literature on network centrality measures, performance measures, describe our theoretical framework and formulate our hypotheses. In Section 3.1, we describe our dataset and the relevant variables. In Section 3.2, we present our research methodology and results. In Section 4, we conclude by highlighting theoretical implications and practical implications for managers’ decisions with regard to supply chains. We then highlight some limitations of our study.

2. Literature review and hypothesis building

2.1. Network theory and social network analysis

We use network theory as the theoretical framework for our study. According to Borgatti and Halgin, Borgatti and Li network theory aims at answering the performance question, which is why some actors have better outcomes than others [1,5]. Network theory in general argues, first, that the structure of the network a company is embedded in has an effect on various attributes (independent of the attributes of the company itself) [1]. Second, network theory argues that the network flow influences performance. The goals of network theory are to explain choice and success, including performance and reward [1].

Several studies confirmed network theory for supply networks and have shown that there exists an impact of network position and ties on the performance of focal companies. These studies investigated various performance measures, such as innovation [6,7,8], tie-formation [9,10] and financial performance [2,11].

In this context, several supply chain management scholars applied social network analysis to investigate and analyze centrality in supply networks [1,3,10,12,13,14]. Identifying central nodes and edges in networks is one of the main aims of social network analysis. There are several different centrality measures, each capturing different aspects of embeddedness of a node in a network [13]. The three node centrality measures we consider here are degree-, betweenness-, and eigenvector-centrality. For example, edges and companies can be critical due to their ties with other critical companies (eigenvector centrality), they may control important material flows (betweenness centrality) or they can have a large operational load (degree centrality) [13]. Below, we quickly introduce these centrality measures.

Let us assume that the network has n nodes and let i, j, k denote these nodes. We set Xij if there is an edge between node i and j, and 0 else. Degree centrality corresponds to the number of direct ties a node has [15]. A node with high degree centrality is well connected in the network. Degree centrality measures where "the action is in the network" [16]. It is defined by:

$$\mathrm{DC}(i) = \sum_{j} Xij. \tag{1}$$

In distribution networks, a node with high degree centrality is subject to a high operational load, because it has to manage large inflows and outflows of material [13]. Degree centrality is a local network centrality measure. It only considers the direct neighborhood of the node.

Betweenness centrality is a global network centrality measures. To define it, we introduce the concept of path length

$$\mathrm{l}(m) = \sum_{(i,j)\epsilon N(m)} 1, \tag{2}$$

where $(i,j) \epsilon N(m)$ if the edge (i, j) lies on the path m. A geodesic for the pair i, j is a shortest path m from i to j. Betweenness centrality of a node measures the number of shortest paths that pass through this node [15]. A node with high betweenness centrality might control network flows, because it connects nodes that might be otherwise unconnected or that might be connected by much longer paths. If we denote by gkj the total number of geodesics linking the two nodes $k\, j$ and if $gkj(i)$ is the number of those geodesics that contain node i, then

$$\mathrm{BC}(i) = \sum_{j \neq k} \frac{gkj(i)}{gkj}. \tag{3}$$

In distribution networks, a node with high betweenness centrality may act as a bottleneck by controlling material flows [13].

Finally, a node receives high eigenvector centrality if it directly ties with many other nodes that receive a high eigenvector centrality themselves [17]. I.e., the scores $E = (E(1), \ldots, E(n))$ satisfy the equation:

$$\mathrm{EC}(i) = \frac{1}{\alpha} \sum_{j} E(j) Xij, \tag{4}$$

where α is a constant and $E(i) > 0$ for all i such that: $0 < i < n + 1$. A node with high eigenvector centrality may be important in distribution networks, because it ties with many other important nodes. Such a node may be critical, because if it has troubles, these may directly carry over to other critical nodes in the network.

2.2. Performance measures

There are several measures for assessing the performance of a supply chain, as for instance, total cycle time, delivery lead-time and delivery performance, accuracy of forecast and delivery reliability, capacity utilization, total inventory, product availability, quality, responsiveness, throughput, costs and many more [18,19]. In this paper, we are especially interested in time-related key performance indicators (KPIs) as the main purpose of distribution networks is to secure on-time delivery. We therefore collected time-stamp data for the OEM distribution network to evaluate the KPIs stay-time performance [19], stay-time variation [18,20,21] and inventory level [18,19,20]. Hypothesis building

Given network theory, an effect of social network position on performance in distribution networks is highly likely. For deriving more concrete and testable hypotheses, we invoke the interpretation of social network centrality in material flow networks provided by [13]. Nodes with high degree centrality are subject to large operational load and at the same time such a node would often be required to reconcile conflicting schedules or interests between others [13]. Degree centrality is also associated with difficulties to manage incoming material and outgoing material. The more nodes are downstream of the node, the more challenging it is for the firm to ensure on-time delivery, cost-effective inventory, and order-management for their customers. We therefore hypothesize that in nodes with large degree centrality high inventory, long and variable stay-times will result. Similar considerations, which we consider in an extended version of this paper, motivate that there is an effect of betweenness centrality and eigenvector centrality on performance. In aggregate, we hypothesize that:

Nodes with high degree centrality, high betweenness centrality or high eigenvector centrality will suffer from higher inventory, higher stay-times, and higher stay-time variation.

3. Methodology

3.1. Data and variables

Our dataset describes the distribution network of the German OEM for half a calendar year. The distribution network is globally dispersed with several different players involved. The OEM ships 1.2 million vehicles per half year from 15 plants to over 2,000 dealer via 370 nodes (ports, vehicle distribution centers, airports, and others). The distribution network reaches into almost all geographic areas worldwide. Plants are located in 13 countries on five continents. Dealers are located in 124 countries and 212 delivery areas on all continents. A large part of the transport is therefore intercontinental, adding to complexity. Over 290 different logistic service providers carry out the transports. Several different transportation modes are involved: road, rail, ship and to minor extend air travel. Figure 1 displays a typical network path of the distribution network.

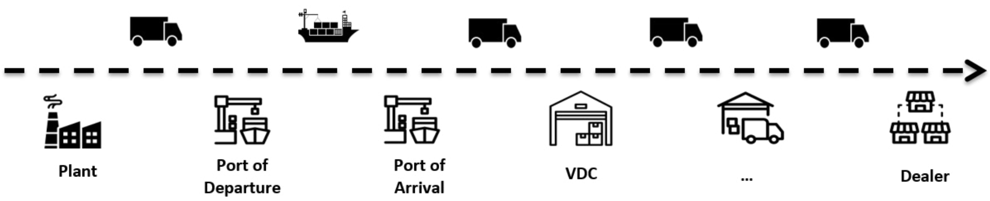

Figure 1. A typical network path of a distribution network.

For each vehicle, timestamp data for the complete path through its network was available, i.e., when did the vehicle enter and leave each node on its path to the dealer (Table 1).

Table 1. An example of timestamp data.

Vehicle_Id	Source	Target	Leaving	Arriving
5378	VDC 10	Dealer 325	20.03.2020 10:00	20.03.2020 15:00
5378	POA 27	VDC 10	17.03.2020 09:00	19.03.2020 14:00

As independent variables, we use our three main node centrality measures. We calculated these variables using the R-package igraph. For each node, as dependent variables we used and calculated the following performance measures:

- standard deviation of stay-time per vehicle (planning uncertainty)
- mean stay-time per vehicle
- mean inventory per day.

Since the transport ends at the dealer's sites, no stay-times and inventories exist for dealers and plants. Consequently, we removed dealers and plants from our analysis leaving us with 352 data points. However, we calculated social network centrality for all the remaining nodes in the complete network containing all dealers. There are several nodes passed by only few vehicles. For our dependent variables to being reliable, we removed all nodes passed by less than 30 vehicles from the data set. We then calculated all the dependent variables for the remaining 238 data points.

To reduce bias, we use volume passing a node and type of node (port of arrival, port of departure, vehicle distribution center, rail hub, airport of arrival. airport of departure and others) as control variables in our OLS-regression. After all, it is reasonable to assume that at least some of the centrality measures are affected by the type of node and affect the volume passing, which in turn affect the dependent variables. For instance, being a vehicle distribution center may be a common cause for high degree centrality and high standard deviation.

3.2. Data analysis and results

To assess the effect of centrality measures on performance, we consider ordinary least square (OLS) regression with controls type of location and volume passing.

We analyzed the 238 data points for relationships between network centrality measures and mean stay-times, standard deviation of stay-times and mean inventory. Considering descriptive statistics, we observe highly significant correlations ($p<0.01$) between all centrality measures and all dependent variables (Table 2-

Table 3). Correlations range from 0.20 to 0.53. It seems therefore that there is an effect of all centrality measures on all performance variables.

Table 2. Summary statistics.

Variable	Mean	Median	SD	Min	Max
1) Betweenness centrality	5,037.00	359.00	13,128.88	0.00	110,905.00
2) Degree centrality	21.12	5.00	63.54	2.00	516.00
3) Eigenvector centrality	0.08	0.03	0.12	0.00	1.00
4) Standard-deviation stay-times	5.44	1.39	7.77	0.00	36.31
5) Mean stay-time	6.55	4.40	7.39	0.00	51.50
6) Mean inventory per day	304.58	10.49	968.80	0.00	9,231.13
7) Volume	1,120.00	695.00	15,174.14	32.00	180,240.00

Table 3. Pairwise correlations, * p<0.05; **p<0.01; *** p<0.001, p-values have been adjusted by Holm's method, n=238.

Variable	1)	2)	3)	4)	5)	6)	7)
1) Betweenness centrality	1						
2) Degree centrality	0.61***	1					
3) Eigenvector centrality	0.27***	0.45***	1				
4) Standard-deviation stay-times	0.38***	0.44***	0.21**	1			
5) Mean stay-time	0.20**	0.35***	0.22**	0.78***	1		
6) Mean inventory per day	0.53***	0.52***	0.37***	0.78***	0.66***	1	
7) Volume	0.60***	0.43***	0.17*	0.42**	0.15*	0.71***	1

We consider now OLS-regression for the three dependent performance variables. To increase interpretability of effect-sizes, we normalize numerical variables by z-transforming them. We first enter the control variables volume and type of location (baseline model) and then we enter all centrality measures individually as independent variables into the model. Finally, we consider the full model consisting of all centrality variables and the control variables. Tables Table 4-Table 6 display the result of the regressions.

Table 4. Regression results (DV Standard-deviation stay-times) * $p<0.05$; ** $p<0.01$; *** $p<0.001$, standard error in parentheses, $n=238$.

Variable	1)	2)	3)	4)	5)
Degree centrality		0.19 (0.06)**			0.18 (0.06) **
Eigenvector centrality			0.05 (0.06)		0.45 (0.06)
Betweenness centrality				0.08 (0.06)	0.03 (0.06)
Volume	0.11 (0.05) *	0.07 (0.05)	0.10 (0.06)	0.07 (0.06)	0.04 (0.06)
Type of location	YES***	YES***	YES***	YES***	YES***
Multiple R2	0.40	0.42	0.40	0.40	0.42
Adjusted R2	0.38	0.40	0.38	0.38	0.40

Table 5. Regression results (DV Mean stay-time) * $p<0.05$; ** $p<0.01$; *** $p<0.001$, standard error in parentheses, $n=238$.

Variable	1)	2)	3)	4)	5)
Degree centrality		0.14 (0.07)*			0.15 (0.07) *
Eigenvector centrality			0.05 (0.06)		0.06 (0.06)
Betweenness centrality				0.01 (0.07)	-0.03(0.07)
Volume	0.07 (0.06)	0.04 (0.06)	0.05 (0.06)	0.07 (0.07)	0.03 (0.07)
Type of location	YES ***	YES***	YES***	YES***	YES***
Multiple R2	0.29	0.30	0.29	0.29	0.30
Adjusted R2	0.26	0.28	0.26	0.26	0.28

Table 6. Regression results (DV Mean inventory per day) * $p<0.05$; ** $p<0.01$; *** $p<0.001$, standard error in parentheses, $n=238$.

Variable	1)	2)	3)	4)	5)
Degree centrality		0.26 (0.04)***			0.28 (0.04)***
Eigenvector centrality			-0.05 (0.04)		-0.04 (0.04)
Betweenness centrality				0.00 (0.04)	-0.06 (0.04)
Volume	0.79 (0.04)***	0.73 (0.03)***	0.81 (0.04)***	0.79 (0.04)***	0.78(0.04)***
Type of location	YES**	YES	YES**	YES**	YES
Multiple R2	0.71	0.76	0.71	0.71	0.76
Adjusted R2	0.70	0.75	0.70	0.70	0.75

We highlight the effect of the control variables first. Type of location has a significant positive effect on mean and standard deviation of stay-times in the full model including all relevant variables. Such effects are reasonable, because different kinds of nodes have different functions in the network. A vehicle distribution center's main function is to store vehicles for different dealers, leading to higher and more difficult to plan stay-times than in ports for instance. Volume has a significant effect on mean inventory. Such an effect is to be expected, because higher volume leads to a higher turnover and hence to higher mean inventory.

Regarding the effect of centrality measures on performance, when we carry out OLS-regression with controls type of node and volume passing, we observe a significant positive effect of degree centrality on the standard deviation of stay-times, on mean stay-time and on mean inventory only. All other centrality measures have no significant effect on the dependent variables.

In summary, we confirm the hypothesis that degree centrality is significantly associated with performance issues, while we refute the hypotheses that betweenness centrality and eigenvector centrality are significantly associated with performance issues.

4. Conclusion and limitations

Network theory claims that the position of a node in its network is highly relevant for its performance. Several studies confirm network theory for inter-company networks. However, there has been no prior study on whether network theory also applies to distribution networks. Our study shows that this is indeed the case.

Our paper has several theoretical and practical implications. Our theoretical contributions are threefold. First, our key result is that degree centrality is significantly associated with different performance measures in the case of distribution networks. Our study therefore is first to confirm the applicability of network theory in the context of distribution networks. Second, however, that the effect is only significant for degree centrality provides a more nuanced picture, is somehow surprising and calls for further investigation. Finally, our result is based on large and high quality company internal data. Earlier work on network theory considered smaller datasets or data obtained from external data sources [4]. However, coverage and quality of external data is normally lower than coverage and quality of company internal data. Hence, we contribute to the research on network theory by confirming it based on high quality data, thereby reducing bias.

As a practical contribution, invoking our results, managers may focus on and monitor nodes that are likely to suffer from performance issues by considering network centrality measures. Interestingly, the global structure of the network has an impact on local performance of nodes. Hence, global network decisions should be aligned with knowledge of local properties of nodes. For instance, when a node is known to be resilient and effective, it may be reasonable to try to increase its degree centrality. After all, this node may effectively deal with the performance issues resulting from high degree centrality. Finally, social network analysis needs only network edges as input data. It does not need any time-stamp data of the transports carried out in a company's distribution network. Indeed, companies will often have difficulties to provide such transaction data. Social network analysis is therefore much easier applicable than performance analysis. It may therefore be an additional way to assist greatly in performance related decision-making.

We considered the two control variables type of location and volume. Future work could control for additional variables such as shipment frequencies and geographical location of the node. In addition, instead of using volume as a control variable, mediation models studying the direct and the effect of centrality measures mediated by volume may be considered. In this case, betweenness centrality and eigenvector centrality may also be associated with performance.

Acknowledgments

The financial support by the Austrian Federal Ministry for Digital and Economic Affairs and the National Foundation for Research, Technology and Development and the Christian Doppler Research Association is gratefully acknowledged.

References

[1] Borgatti, Stephen P., and Xun Li. (2009) "On social network analysis in a supply chain context." *Journal of Supply Chain Management* **45** (**2**): 5–22.

[2] Lu, Guanyi, and Guangzhi Shang. (2017) "Impact of supply base structural complexity on financial performance: Roles of visible and not-so-visible characteristics." *Journal of Operations Management* **53-56** (**1**): 23–44.

[3] Shao, Benjamin B.M., Zhan Shi, Thomas Y. Choi, and Sangho Chae. (2018) "A data-analytics approach to identifying hidden critical suppliers in supply networks: Development of nexus supplier index." *Decision Support Systems* **114**: 37–48.

[4] Swift, Caroline, V. Daniel R. Guide, and Suresh Muthulingam. (2019) "Does supply chain visibility affect operating performance? Evidence from conflict minerals disclosures." *Journal of Operations Management* **65** (**5**): 406–429.

[5] Borgatti, Stephen P., and Daniel S. Halgin. (2011) "On Network Theory." *Organization Science* **22** (**5**): 1168–1181.

[6] Wei, Xuan, and Wei Chen. (2019) "How Does A Firm's Previous Social Network Position Affect Innovation? Evidence from Chinese Listed Companies." *Sustainability* **11** (**4**): 1–20.

[7] Kim, Myung Kyo, Sriram Narayanan, and Ram Narasimhan. (2020) "Supply network architecture and its contingent impact on innovation performance: A field study." *International Journal of Production Economics* **224**.

[8] Bellamy, Marcus A., Soumen Ghosh, and Manpreet Hora. (2014) "The influence of supply network structure on firm innovation." *Journal of Operations Management* **32** (**6**): 357–373.

[9] Gulati, Ranjay. (1999) "Network location and learning: the influence of network resources and firm capabilities on alliance formation." *Strategic Management Journal* **20** (**5**): 397–420.

[10] Ahuja, Gautam, Francisco Polidoro, and Will Mitchell. (2009) "Structural homophily or social asymmetry? The formation of alliances by poorly embedded firms." *Strategic Management Journal* **30** (**9**): 941–958.

[11] Shia, Jinyan, Jiangheng Yang, and Yanxi Li. (2019) "SUPPLY NETWORK POSITION AND FIRM PERFORMANCE: EVIDENCE FROM CHINESE LISTED MANUFACTURING COMPANIES." *Journal of Business Economics and Management* **20** (**6**): 1258–1277.

[12] Carter, Craig R., Lisa M. Ellram, and Wendy Tate. (2007) "The Use of Social Network Analysis in Logistical Research." *Journal of Business Logistics* **28** (**1**): 137–168.

[13] Kim, Yusoon, Thomas Y. Choi, Tingting Yan, and Kevin Dooley. (2011) "Structural investigation of supply networks: A social network analysis approach." *Journal of Operations Management* **29** (**3**): 194–211.

[14] Yan, Tingting, Thomas Y. Choi, Yusoon Kim, and Yang Yang. (2015) "A Theory of the Nexus Supplier: A Critical Supplier From A Network Perspective." *Journal of Supply Chain Management* **51** (**1**): 52–66.

[15] Freeman, Linton C. (1977) "A Set of Measures of Centrality Based on Betweenness." *Sociometry* **40** (**1**): 35.

[16] Wasserman, Stanley, and Katherine Faust. (1994) *Social Network Analysis. Methods and Applications.* Cambridge: Cambridge University Press.

[17] Bonacich, Phillip. (1972) "Factoring and weighting approaches to status scores and clique identification." *The Journal of Mathematical Sociology* **2** (**1**): 113–120.

[18] van der Vorst, Jack G.A.J. (2000) "Effective food supply chains; generating, modelling and evaluating supply chain scenarios." Dissertation. Wageningen University.

[19] Gunasekaran, Angappa, C. Patel, and E. Tirtiroglu. (2001) "Performance measures and metrics in a supply chain environment." *International Journal of Operations & Production Management* **21** (**1/2**): 71–87.

[20] Dong, Ming, and F. Frank Chen. (2005) "Performance modeling and analysis of integrated logistic chains: An analytic framework." *European Journal of Operational Research* **162** (**1**): 83–98.

[21] Raghavan, N. R. Srinivasa, and N. Viswanadham. (2001) "Generalized queueing network analysis of integrated supply chains." *International Journal of Production Research* **39** (**2**): 205–224.