

Stolz, Simon Andreas

Targeting and Empowering Users in a Digital Era

Dissertation
for obtaining the degree of Doctor of Business and Economics
(Doctor rerum politicarum - Dr. rer. pol.)

at WHU – Otto Beisheim School of Management

Submitted
22 February 2021

Published
10 November 2021

First Advisor: *Professor Christian Schlereth*

Second Advisor: *Professor Dries Faems*

“If we have data, let’s look at data.

If all we have are opinions, let’s go with mine.”

- Jim Barksdale (former Netscape CEO)

Table of Contents

1	GENERAL INTRODUCTION.....	4
2	PREDICTING TIE STRENGTH WITH EGO NETWORK STRUCTURES	8
2.1	INTRODUCTION	8
2.2	THEORETICAL BACKGROUND.....	9
2.2.1	<i>Tie Strength in Marketing</i>	9
2.2.2	<i>Tie Strength Assessment</i>	10
2.2.3	<i>Ego Network Measures and Tie Strength</i>	12
2.2.4	<i>Ego- vs. Full Network Perspective</i>	15
2.3	EMPIRICAL STUDY	15
2.3.1	<i>Data Collection</i>	15
2.3.2	<i>Validation of the Ego Network Perspective</i>	17
2.3.3	<i>Prediction Model Formulation</i>	19
2.3.4	<i>Practical Demonstration</i>	22
2.4	DISPLAY ADVERTISEMENT EXPERIMENT: IMPACT OF TIE STRENGTH ON CLICKING BEHAVIOR.....	24
2.5	DISCUSSION	25
2.5.1	<i>Summary</i>	25
2.5.2	<i>Theoretical and Managerial Implications</i>	26
2.5.3	<i>Limitations and Future Research</i>	28
3	SOCIAL CAPITAL ACCUMULATION AND PREMIUM MEMBERSHIPS IN WORK-RELATED NETWORKS	29
3.1	INTRODUCTION	29
3.2	THEORETICAL BACKGROUND.....	33
3.2.1	<i>Social Capital</i>	33
3.2.2	<i>Networking</i>	34
3.3	ACCUMULATING SOCIAL CAPITAL: CAUSAL EVIDENCE (RQ1).....	36
3.3.1	<i>Experimental Set-Up</i>	36
3.3.2	<i>Data</i>	38
3.3.3	<i>Challenges and Estimation</i>	40
3.3.4	<i>Results</i>	42
3.4	FEATURE UTILIZATION: SUPPORTIVE EVIDENCE (RQ2).....	46
3.4.1	<i>Framework</i>	47
3.4.2	<i>Data</i>	48
3.4.3	<i>Challenges and Estimation</i>	50
3.4.4	<i>Results</i>	52
3.5	DISCUSSION	59
3.5.1	<i>Theoretical Implications</i>	60
3.5.2	<i>Practical Implications</i>	62

3.6	LIMITATIONS AND FUTURE RESEARCH.....	64
4	ONLINE LEAD GENERATION: LEARNINGS FROM AN EMERGING INDUSTRY	65
4.1	INTRODUCTION	65
4.2	THEORETICAL BACKGROUND.....	66
4.2.1	<i>Online Lead Generation Companies</i>	66
4.2.2	<i>Related Frameworks</i>	67
4.2.3	<i>Best Practices and Trends from the Literature</i>	67
4.3	RESEARCH CONTEXT AND METHODS.....	68
4.4	ONLINE LEAD GENERATION FUNNEL.....	68
4.5	FINDINGS AND RECOMMENDATIONS	71
4.5.1	<i>Lead Capturing: Attracting Leads Through Information</i>	71
4.5.2	<i>Lead Capturing: The Power of Multi-Step Forms</i>	71
4.5.3	<i>Lead Nurturing and Qualification: Prompt Responses Are Not Always the Best Option</i>	72
4.5.4	<i>Lead Scoring: Data-Rich Environments Enable Lead Scoring</i>	73
4.5.5	<i>Sales Closure and After-Sales: Trend Toward Business Model Expansion</i>	73
4.6	DISCUSSION AND FURTHER RESEARCH.....	74
5	GENERAL CONCLUSION.....	77
6	APPENDICES.....	79
A1	COMPUTATION OF CENTRALITY MEASURES	79
A2	COMPUTATION OF CENTRALITY MEASURES COMPUTATION OF CENTRALITY MEASURES	79
A3	KENDALL CORRELATION MATRIX – ALL NETWORK MEASURES FOR ADJACENT ALTERS OF EGO. 80	
A4	IMPACT OF CLOSE FRIENDS SURVEY ITEMS	81
A5	OVERVIEW OF FEATURES ON WORKSMN BY MEMBERSHIP TYPE	82
A6	EXPERIMENT: FLOW DIAGRAM	83
A7	EXPERIMENT: DESCRIPTIVE STATISTICS AND BIVARIATE CORRELATIONS	84
A8	EXPERIMENT: ELABORATION ON THE COMPLIER AVERAGE CAUSAL EFFECT (CACE)	85
A9	EXPERIMENT: COVARIATE BALANCE CHECK AFTER ATTRITION	88
A10	EXPERIMENT: FIRST-STAGE ESTIMATE AND IV VALIDATION	88
A11	EXPERIMENT: DISCUSSION OF POST-TREATMENT BIAS	89
A12	INDIVIDUAL-LEVEL PANEL DATA: ROBUSTNESS OF PERCENTAGE CHANGE COMPUTATION.....	90
A13	INDIVIDUAL-LEVEL PANEL DATA: DESCRIPTIVE STATISTICS AND BIVARIATE CORRELATIONS.....	92
A14	INDIVIDUAL-LEVEL PANEL DATA: PANEL VECTOR AUTOREGRESSION MODEL.....	93
A15	INDIVIDUAL-LEVEL PANEL DATA: PROPENSITY SCORE MATCHING	94
7	BIBLIOGRAPHY	96

1 General Introduction

The ongoing digitalization transforms our daily lives continuously. Market researchers estimate that in Germany the average internet user spends at least 5 hours and 10 minutes on digital devices daily, of which 1 hour and 22 minutes are dedicated to social media sites (eMarketer, 2020). Choice experiments underpin the important role of digital services: In a real-life setting, users equated losing access to the internet for one year with foregoing 5,000\$ in personal income (Brynjolfsson, Collis, & Eggers, 2019). Technological advancements enable and challenge existing businesses and give rise to new business models. At the time of writing this dissertation, the Corona pandemic even amplifies this shift: Practitioner surveys show that 70% of managers in the German-speaking industry believe that digital transformation will accelerate due to the pandemic (Malev, 2020).

In the digital domain, users generate a flood of data in various forms. The wealth of data allows us to assess various behavioral processes in detail that historically were difficult to quantify objectively. These new developments challenge researchers to develop new methods to analyze this data and to apply existing methods in novel settings (Moe & Ratchford, 2018). The three projects of my cumulative dissertation address this digitalization trend. In particular, they follow the overarching topic of how to target and empower users in those new digital settings.

The cumulative dissertation consists of three self-contained main sections, according to my three dissertation projects. In the following, I describe the different ways in which the projects add to the overarching theme of targeting and empowering users in a digital era and discuss the novel contributions with respect to existing work. Additionally, I use footnotes to identify sections that directly build on the work of others.

Firstly, in early 2018, within the first 12 months of my dissertation, the Cambridge Analytica investigations disclosed that voters in the US presidential election were “micro-targeted” by using individual-level Facebook data, obtained through readily accessible data interfaces (Cadwalladr & Graham-Harrison, 2018). The case highlighted that users often disclose sensitive information about themselves unexpectedly. For example, Facebook “Likes” allow predicting personality traits and even sexual orientations (Kosinski, Stillwell, & Graepel, 2013). Hence, the question of which sensitive (hidden) information those user data contain is of popular interest. In my first dissertation project, co-authored with Prof. Dr. Christian Schlereth, we provide new answers to this question. We derive predictions via data

abundantly available in networks: Ego network data. We show that it is possible to utilize the network data to predict tie strength, i.e., who a user perceived as her or his closest friend. We further present a review of prior work that show that stronger ties exert higher peer influence, so that this information enables granular targeting (for example via social advertisements).

This first project builds on Facebook network data, which I collected as part of my Master's thesis. Already in my Master's thesis, I noted the ability to predict tie strength from ego networks (Stolz, 2014, pp. 27-28, 45-47). However, I was unclear about the underlying explanation for these observed patterns, lacked a rigorous machine learning framework to address these observations, and did not provide an extensive positioning of this finding in the literature. We managed to address these aspects in section 2, which is published in the *Journal of Interactive Marketing* in Stolz and Schlereth (2021). Additionally, we contribute the anonymized network data to open research data portal Mendeley Data. The dataset is an own citable reference via Stolz and Schlereth (2020). It contains 41 anonymized ego networks and the corresponding total of 18,541 ego-to-alter dyad characteristics in the form of binary variables. Given the wide-ranging implications of our research, I hope that the dataset will be valuable in sociological, network, and managerial sciences.

Secondly, those social media networks do not only capture network data (i.e., who connects to who) but commonly all sorts of user activities are logged and can be used to target users. At the same time, in social media networks, network effects occur—meaning that every additional user enhances the product's value (Shapiro & Varian, 1999). For example, a network, like LinkedIn, would be worthless if no other users were using it. Hence, one of the most prominent business models in this setting is the freemium model. Freemium refers to offering a membership with reduced functionality for free and a premium membership with additional features to paying users (Shapiro & Varian, 1999). A common marketing measure is that users receive discounted or free premium memberships (Koch & Benlian, 2017). Thus, a primary managerial question in this setting is who these networks should target with those empowering premium memberships to foster networking activity and achieve the formation of a dense network. In my second dissertation project, co-authored with Dr. Michael Weiler, Prof. Dr. Andreas Lanz, Prof. Dr. Christian Schlereth, and Prof. Dr. Oliver Hinz, we investigate this question.

For this second project, we merged our insights from an observational dataset of detailed individual-level behavioral data with the experimental results described by Dr. Michael Weiler in his dissertation at Goethe University Frankfurt (2019a). As part of his dissertation,

Dr. Michael Weiler and Prof. Dr. Oliver Hinz conduct an experiment on a work-related social media network (“WorkSMN”) by donating premium memberships to random freelancers (2019a). We substantially extended the work of Weiler and Hinz (2019a) by providing a range of complementary analyses through an observational dataset, provided to us by WorkSMN, and by revising the entire manuscript. We provide robustness to their results by replicating their experimental analysis with observational data: We overcome challenges in the estimation of Weiler and Hinz (2019a) by replacing the pivotal subjective self-reported survey item (“strategic networking behavior”) with objectively measured user behavior and overcome the post-treatment bias inherent in the experiment by using pre-treatment observations. Our results converge, highlighting that Social Capital under premium membership increases most for those users who already exhibit high networking behavior. As our individual-level panel data confirms, particularly users who are actively networking already before their premium membership starts, increase their Social Capital even more when having access to the advanced networking features of a premium membership. In contrast, it is difficult to “wake up sleeper” accounts that are characterized by low networking activity. Further, we extend the prior work of Weiler and Hinz (2019a) by developing a novel framework of Social Capital creation in an online setting. Along the main dimensions “activity” and “saliency”, we identify key mechanisms how the users can utilize the premium features to accumulate Social Capital under premium membership. We investigate which premium features contribute most to the user’s own activity (active features) and the user’s saliency (i.e., visibility—passive features) in Social Capital accumulation. To adhere to the dissertation regulations at WHU and reflect the independence of my dissertation, I present an adapted version of the manuscript in which I added references to the dissertation of Weiler and Hinz (2019a). The revised manuscript is forthcoming at MIS Quarterly (Weiler, Stolz, Lanz, Schlereth, & Hinz, 2022).

Thirdly, given the abundance of information available on the Internet, users shift their attention to the digital space when searching for products and services. Emerging lead generation platform businesses (such as Check24, Verivox, or Enpal) offer a platform for those users who research their products online. In particular, they empower users by offering transparency (e.g., as price comparison services) or by acting as a reliable partner in complex undertakings (e.g., installing solar panels at home). In my third dissertation project, co-authored with Kilian Wisskirchen, Prof. Dr. Christian Schlereth, and Alexander Hoffmann, we investigate this emerging online lead generation industry.

The third dissertation project builds on Wisskirchen (2020) and the underlying data of 19 semi-structured practitioner interviews. In the project, we contribute a thorough reassessment of those interview statements. Among other things, we refine the online lead generation funnel and contrast it with the traditional “offline” sales conversation process. We find that the online lead generation process typically skips the initial stages of prospecting customers and derive novel practical recommendations along the online lead generation funnel. The manuscript is currently under review at St. Gallen Marketing Review.

Further, I was happy to be involved in another practitioner cooperation with project partner Facebook. In various business workshops, we assessed the question of whether social media advertisements on Facebook are capable of influencing brand equity on a representative scale. Our approach and method in this cooperation are described in Hein, Schlereth, and Mueller-Klockmann (2019), who acknowledged my support.

In the following main sections 2, 3, and 4, I present the dissertation projects and provide a general conclusion in section 5. As a final remark, the terminology used in the first two projects differs: Section 2 refers to online social networks (OSN), whereas section 3 refers to social media networks (SMN). These terms are, however, largely relating to the same concept and the different wording of SMN has emerged in adaptation to prior work of my co-authors.

2 Predicting Tie Strength with Ego Network Structures¹

2.1 Introduction

In an online social network (OSN), not every connection is equally strong and important. With our closest friends, we spend the most time, perceive emotional intensity, intimacy, and confide in each other. OSNs often do not distinguish between best friends and mere acquaintances, so that all relationships are by default uniformly labeled, e.g., as friends on Facebook or as contacts on LinkedIn (Gilbert & Karahalios, 2009; Jones et al., 2013). The latent metric tie strength differentiates strong ties (i.e., closest friends) from weak ties (Granovetter, 1973) and is an important variable in online and offline marketing. Studies in word-of-mouth (e.g., Brown & Reingen, 1987), viral marketing (e.g., Hayes, King, & Ramirez Jr, 2016), and social advertisement (e.g., Bakshy, Eckles, Yan, & Rosenn, 2012) agree that tie strength has a major positive influence on referral processes and peer influence.

In practice, survey-based assessments of tie strength for all users of an OSN are often infeasible due to scale, implied costs, and privacy concerns. Hence, various studies employ readily available online user data, called “revealed preferences,” such as interaction frequency, as a proxy for tie strength (e.g., Bakshy et al., 2012). Yet, relying on such established online measures can be problematic. For example, as one uses multiple online and offline communication channels, interaction logs are incomplete (Wiese, Min, Hong, & Zimmerman, 2015). This discrepancy between online and offline worlds is also evidenced in R. Bapna, Gupta, Rice, and Sundararajan (2017), who find a limited association between online proxies of tie strength and actual trust. With this research, we propose an approach that predicts real-world tie strength via online predictors by combining network measures with non-network measures.

While most of the network structure studies in marketing use “macroscopic” sampled, full or socio-centric networks (e.g., Ebbes, Papies, & van Heerde, 2016; Hinz, Skiera, Barrot, & Becker, 2011; Katona, Zubcsek, & Sarvary, 2011; Schlereth, Barrot, Skiera, & Takac, 2013), we propose to take a “microscopic” perspective to predict tie strength. That means, we look only at ego network structures, which contain all first-degree connections and the interlinkage

¹ I presented this project at the 2018 EMAC Conference in Glasgow, Scotland. The work is published as Stolz and Schlereth (2021) “Predicting Tie Strength with Ego Network Structures” in *Journal of Interactive Marketing*, Volume 54

among them. On the level of the ego networks, we can observe common contacts and social circles - structures that are not apparent in the full network. We argue theoretically and demonstrate empirically that bridging positions between social circles in ego networks have high individual predictive power. Research on tie strength prediction that uses ego networks is scarce. The only exception, known to the authors, is Backstrom and Kleinberg (2014), who identify strong ties for the special case of romantic partnerships. Yet, we apply ego centrality measures to predict real-world tie strength beyond this special case. Moreover, emphasizing practicability, we limit ourselves to data retrieved via questionnaires and application programming interfaces (APIs).

We make the following contributions: Firstly, we motivate why bridging ties in ego networks may indicate real-world tie strength. Secondly, we empirically validate the resulting assumption that structural measures from an ego network have high a predictive ability and compare them to a sampled perspective. Thirdly, we extend the portfolio of revealed preference measures of tie strength by highlighting the predictive power of a network-based measure, namely bridging positions in ego networks. Our analyses concur that bridging ties in ego networks provide predictive power in identifying the rare 1.4% of strong ties among the 18,541 real-world connections. Lastly, we demonstrate the relevancy of our applied real-world tie strength measure of identifying the few “closest friends” in a display experiment of social advertisements.

We structure the remainder of this paper as follows: Section 2.2 discusses the key features of OSN, compares prior studies of tie strength prediction, and explains the theoretical motivation of a relation between ego network measures and tie strength. Section 2.3 empirically validates the predictive power of the proposed ego network perspective and assesses ego network measures in combination with other predictors. In Section 2.4, a follow-up study demonstrates that knowledge about tie strength is beneficial in social advertisements. We conclude by highlighting implications and outlining limitations in Section 2.5.

An anonymized version of our research data is available via Stolz and Schlereth (2020).

2.2 Theoretical Background

2.2.1 *Tie Strength in Marketing*

In line with bounded rationality theory (Simon, 1955), there are limitations to the number of connections a person can maintain, which are imposed by the costs of cognitive resources

and the amount of time a person has (Dunbar, 1993). Therefore, naturally, only a limited number of friends can be very close, as reflected by the notion of “tie strength”.

Tie strength in the initial formulation of Granovetter (1973, p. 1361) is defined as a “combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize a tie”. However, Granovetter (1973) explicitly leaves a definition of operational measures that reflect tie strength to future research. Consequently, Marsden and Campbell (1984) observe that next to the “most common approach,” i.e., assessing perceived closeness, “numerous other measures of strength have also been used or proposed” (p. 483-484). Further research highlights additional aspects, such as reciprocity and evolution of tie strength (Friedkin, 1990).

In marketing, researchers frequently noted the relevance of tie strength on consumer decisions: Offline peer influence studies show that strong ties are more likely to be activated for referrals (Reingen & Kernan, 1986) and are more influential (Brown & Reingen, 1987). But also in online environments, strong ties are found to impact peer influence in app adoptions (Aral & Walker, 2014), opinion seeking, and -passing (Chu & Kim, 2011). Additionally, when social advertisements present friend names as visual cues (e.g., “Lisa and 3 other friends also like Company A”), response rates to strong ties are higher (Bakshy et al., 2012).

2.2.2 Tie Strength Assessment

Two general forms of tie strength measurement can be contrasted in the literature: Perception-based assessments and revealed preference assessments. *Perception-based assessments* are typically applied in offline studies. They ask for the individual subjective assessments of real-world tie strength using surveys, i.e., whether a person is an “acquaintance” or someone the respondent feels “close to” and perceives as important (e.g., Brown & Reingen, 1987; Chu & Kim, 2011). On the other hand, *revealed preference assessments* are typically applied in online studies. They infer tie strength via readily available user data, such as mutual friends, common social groups, or interactions (e.g., Aral & Walker, 2014; Bakshy et al., 2012; R. Bapna et al., 2017; Rishika & Ramaprasad, 2019). All these revealed preference measures capture different facets of the online relationship but may miss real-world perceptions.

As applied in our research, tie strength prediction models bridge this gap by obtaining perception-based assessments of tie strength for a sample of individuals by directly asking

them for their closest friend. This information then serves as input for exploring the degree to which the revealed preference traits in online networks reflect real-world perceived assessments. Estimating the strength of a relationship becomes possible for individuals, where knowledge about real-world tie strength is missing.

Prior studies that predict tie strength in OSNs leverage three types of *revealed preferences* as predictors: Firstly, researchers frequently use the *similarity* of user attributes. *Similarity* is a valuable predictor due to homophily effects prevalent in social networks: Individuals tend to befriend others who are similar to them (McPherson, Smith-Lovin, & Cook, 2001). Various types of commonalities are used to reflect similarity, such as demographic attributes (Arnaboldi, Guazzini, & Passarella, 2013), shared job titles, firm affiliation (Xiang, Neville, & Rogati, 2010), political views, and educational degree (Gilbert & Karahalios, 2009).

Secondly, *interactions* among two peers are also common revealed preference measures of tie strength (e.g., Bakshy et al., 2012). In an OSN setting, related predictors can be the amount of exchanged messages, profile posting, tagging activity, logs of likes, and directed comments (Arnaboldi et al., 2013; Xiang et al., 2010). According to Jones et al. (2013) and Kahanda and Neville (2009), these interaction-based predictors have even higher discriminatory power than similarity-based predictors.

Thirdly, *network structures* found in social networks are related to tie strength. First and foremost, already in his initial formulation, Granovetter (1973) hypothesizes common friends to be indicative of higher tie strength. Hence, one of the most common network predictors is the number of common contacts (Gilbert & Karahalios, 2009). Various marketing and peer influence studies use a closely related measure, called “structural embeddedness” (Aral & Walker, 2014; R. Bapna et al., 2017; Rishika & Ramaprasad, 2019). This measure reflects the number of common contacts relative to the network size. Moreover, tie strength prediction papers use network measures to build on knowledge about the full network structure or specific network constellations (Kahanda & Neville, 2009; Rotabi, Kamath, Kleinberg, & Sharma, 2017). Lastly, Backstrom and Kleinberg (2014) highlight that bridging positions in ego networks indicate romantic partnerships and formulate a new structural measure, dispersion, to capture bridging configurations.

Table 1 lists a set of prior tie strength prediction papers, ordered chronologically. They all have in common that they use *similarity, interaction, or network structure* data as predictors, use supervised learning, are conducted on OSNs, and use some form of tie strength as the

response variable. In most of these prediction papers, a perceived assessment is captured, such as surveyed via rating scales (Gilbert & Karahalios, 2009), binary responses (Jones et al., 2013), or inferred via “top friends” tags set by the user (Kahanda & Neville, 2009).

Table 1. Tie strength prediction models

Paper	(Respondents) Ties	Predictors				
		Similarity	Interaction	Common contacts	Full network measures	Ego network bridging positions
Our paper	(41) 18,541	✓	✓	✓	-	✓
Rotabi et al. (2017)	(-) Undisclosed	-	-	✓	✓	-
Backstrom and Kleinberg (2014)	(-) ~ 379m	-	✓	✓	-	✓
Jones et al. (2013)	(789) 1,587	✓	✓	-	-	-
Arnaboldi et al. (2013)	(30) 7,103	✓	✓	-	-	-
Gilbert and Karahalios (2009)	(35) 2,184	✓	✓	✓	-	-
Kahanda and Neville (2009)	(-) 8,766	✓	✓	✓	✓	-

Moreover, as Table 1 shows, other tie strength prediction studies use full network data, which can be difficult to acquire. In contrast, we suggest working with ego networks, which have the advantage that they require only information that is usually visible to one user. In addition to using common contacts, we propose to include centrality measures of bridging positions to complement the insights derived from network structure data. With the exception of Backstrom and Kleinberg (2014), who aim their research on identifying a specific type of strong tie, namely romantic relationships, our paper is the first to uncover the predictive power of ego bridging positions to identify real-world tie strength in an OSN.

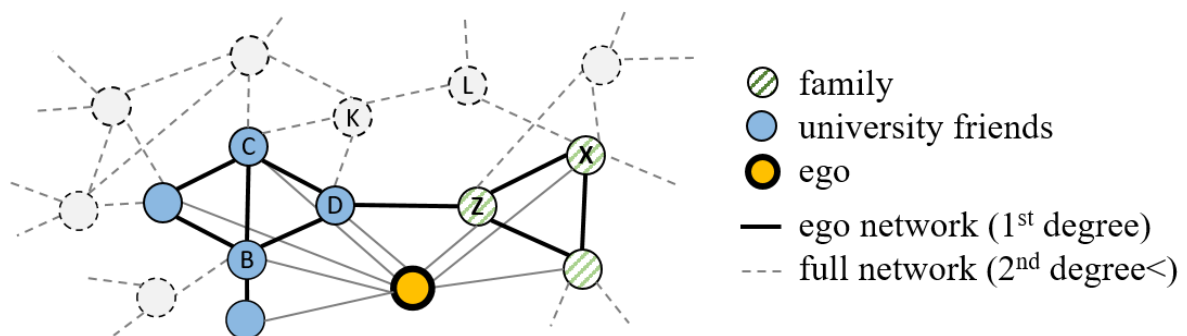
2.2.3 Ego Network Measures and Tie Strength

Social network analysis (SNA) offers a means to quantify the resulting social structures via centrality measures, highlighting different positional properties of a node. Two common measures are the *degree centrality* (hereinafter: *degree*) and *betweenness centrality* (hereinafter: *betweenness*) (e.g., Schlereth et al., 2013). Degree measures the number of

connections that an actor has (Freeman, 1978). Betweenness measures the count of shortest paths that the actor lies on, divided by the number of possible shortest paths in the network. These measures are typically considered as purely global characteristics of the full network (Ebbes et al., 2016). Some studies also use a third measure, namely closeness centrality, measured as the average distance from a focal actor to all other actors within a network to capture communication efficiency. However, closeness centrality requires that all parts of the network are connected, which is not always the case in the ego networks, when excluding the ego. Therefore, we focus only on the former measures in the subsequent discussion.

Figure 1 illustrates the rationale, why degree and betweenness may predict tie strength well in an ego network setting. It shows a synthetic network surrounding a focal ego user. Each node that is connected to ego represents the ego network connections (called alters). Connections between nodes in the ego network are highlighted with solid lines. Dotted lines in Figure 1 represent second-degree (or higher) connections (i.e., friends of friends). Let us assume that only node D is a strong tie, i.e., one person that ego perceives as closest.

Figure 1. Exemplary social network



As observable from Figure 1, the use of *degree* in an ego network provides a different interpretation than its use in a full network. In a full network, degree quantifies the absolute number of contacts, i.e., how popular and well-connected the user generally is. In contrast, degree in an ego network quantifies the common contacts among the friends of the ego (i.e., for node D, the three nodes B, C, and Z). Hence, when neglecting the ego node, degree equates to the frequently used tie strength predictor number of common contacts (see Table 1). This predictor is rooted in the strong triadic closure theorem: strong ties spend large amounts of time together, such that the alter has a high probability of being introduced to many other alters of the ego (Granovetter, 1973, p. 1362). Individuals with many common contacts (i.e., degree in an ego perspective) are assumed to be stronger ties, as the overlap could result from the time spent together. A probabilistic consideration in Granovetter (1973)

highlights this notion: If two persons spend 60% of their time together, then a third person with whom one of those persons spends 40% of her or his time will be together 24% of the time. Hence, individuals with a high ego perspective degree are assumed to be stronger ties, as the common contacts can result from time spent together.

To explain why *betweenness* is a good predictor of tie strength in an ego network setting, we draw on the idea that social connections organize themselves around social circles. Social circles are entities “around which joint activities are organized”, such as workplaces, families, or classes (Feld, 1981). Communities representing multiple social circles are common in a range of major OSN, such as Twitter, Facebook, and Google+ (Leskovec & McAuley, 2012). In Figure 1, we distinguish two exemplary social circles of the ego: family and university friends. The members of a social circle are better connected within the social circle. However, the tie between university friend D and the family circle member Z is special. How come that a university friend had the chance to get acquainted with a family member? The existence of the tie indicates that friend D may have a special standing among the university friends in our example. The intuition is as follows: a bridging position to different social circles may result from any social event which involved multiple social circles of ego (e.g., a birthday party or a visit coincidentally involving other social circles). Following the view of social capital that relationships *require investments* (Lin, 1999), connecting people across distant social circles (in our example, a university class, and a family) requires to create a particular situation for them to connect. In contrast, people within proximate social circles (e.g., two classes at the same university) have more chances to already be acquainted with each other. Thus, a position of an alter that is “more bridging” in the ego network perspective can reflect a higher investment into the relationship.

Backstrom and Kleinberg (2014) propose a complement to the measure *betweenness*, which they call *dispersion*. Dispersion quantifies the distance between common neighbors of ego and the focal alter. The distance of a pair of common neighbors equates one if they are “not directly linked and also have no common neighbors” (other than the ego node and focal alter, cf., Backstrom & Kleinberg, 2014). E.g., for node D in Figure 1, the connection between B to Z and C to Z would count as missing. Only the ego node is the common neighbor. In other words, dispersion focuses on the common contacts of alter and ego (i.e., for node D: just B, C, and Z). In contrast, *betweenness* looks beyond common contacts and considers all nodes within the connected component. While variations of computing dispersion exist, we stick to the default normalized version to compare it to *betweenness*. In Appendix A1, we provide the

respective equations for the computation of all measures and illustrate in Appendix A2 that they provide different computational results.

2.2.4 *Ego- vs. Full Network Perspective*

At first, it may seem counter-intuitive that ego networks predict tie strength well because they discard information that is available in a full network perspective. We argue that removing second-degree connections eradicates distorting information that is redundant in the identification of strong ties. What additional value would the inclusion of all second-degree connections (e.g., nodes K and L) offer to identify D as a strong tie?

Figure 1 depicts why abstracting information from the full network view benefits the tie strength prediction task. The ego network approach focuses only on the alters of ego and the interconnections among them (i.e., the bold lines in Figure 1): Removing second-degree connections may provide a clearer identification of common contacts and bridging positions. In Figure 1, looking at bridging positions between C and X, next to D and Z another bridging position is present in the surrounding full network: Another shortest path leads via K and L. However, this alternative bridge is not capable of constituting a strong tie of ego as it is not part of the first-degree network. Moreover, the ties between ego and the alters of ego are redundant in identifying bridging positions. Including the ego node skews the measures since ego itself will act as a bridge between the alters. We empirically test these considerations in Section 2.3.2 of the subsequent empirical study.

2.3 Empirical Study

2.3.1 *Data Collection*

We conducted an online survey that captured respondents' tie strength perceptions and usage characteristics (i.e., revealed preferences) on Facebook. Facebook is used largely for offline relationship maintenance behavior (Ellison, Steinfield, & Lampe, 2007) and is therefore well suited for our analysis. We recruited respondents via online postings and among university students. A lottery for shopping vouchers served as an incentive for participation.

Through the survey, similar to Jones et al. (2013), we established ground-truth on who is each respondent's closest friend (i.e., the dependent variable in our model) by directly asking respondents for a list of names. The question read: "Who of the contacts in your Facebook profile would you consider as very close to you? / Please indicate the (Facebook-)names of

very close Facebook contacts in the text field below.” Using this open field question for friends’ names, we performed a name matching to usernames found in the network data, which we subsequently extract. We could not match only 12 of the 270 close friend names. We deliberately chose the openly phrased question to leave room for personal interpretation. Our measure simplifies the measurement of tie strength in Marsden and Campbell (1984), who proposes a trichotomous measurement of perceived closeness, i.e., whether an alter is an acquaintance, a good or a very close friend. We deviated from the trichotomous measure, or rating scales (Brown & Reingen, 1987) because those would require each respondent to assess all of their contacts, which in our case averaged 452.2, and amounted up to 1,251 contacts – a task that would be impractical to complete. Moreover, this rare-event binary assessment follows the view that managerial interest is strongest for a small share of users with the highest impact on peer influence (Trusov, Bodapati, & Bucklin, 2010).

To account for *interactions*, further survey items asked for the names of all people messaged on Facebook within the last 5 days and posts on friend’s profiles during the last 6 months. Viswanath, Mislove, Cha, and Gummadi (2009) note that “54% of the interactions between the infrequent interacting user pairs can be directly attributed to Facebook’s birthday reminder feature”, which may reduce the predictive power of this metric. Moreover, we also extracted all friend’s language preferences and gender to account for *similarity* between users. While the granularity of these features is not high, researchers can observe and collect this information for every node in the network.

To extract user network structures, we requested permission from respondents to extract first-degree ego networks via the Facebook research application Netvizz (Bernhard, 2013), i.e., all first-degree connections, in an ego network graph called alters, and the interlinkage among them. Netvizz provides researchers with a user interface to access Facebook’s public APIs. The total number of ties between ego and alters sums to 18,541. Technical challenges and privacy concerns are hypothesized causes for a drop-out rate of 51% when granting Netvizz access to Facebook profile data.

A total of 41 respondents completed all fields in the survey and named between 1 and 27 individuals as their closest friend, averaging 6.2 (SD 4.9). The sample is not representative of the overall population, but with a mean respondent age of 24.5 years (SD 4.5), the sample reflects the young user demographic of social media. Due to the multiplicative effect, the total sample size collected with this method compared to other survey-based studies is still large, as shown in Table 1. Few Facebook profiles appear in the extracted networks that do not reflect

private persons, but companies or organizations, which we removed to avoid a distortion of the network measures.

To derive the centrality measures as predictors, we compute centrality scores for each ego network in isolation. We report the respective equations for the computation of all measures in Appendix A1. Ego network centralities are extracted via the R package *igraph* (Csardi & Nepusz, 2006) and Python package *NetworkX* (NetworkX, 2019). To ensure comparability across the individual networks, we normalized all centrality-based predictors in relation to the number of nodes in the ego network.

2.3.2 *Validation of the Ego Network Perspective*

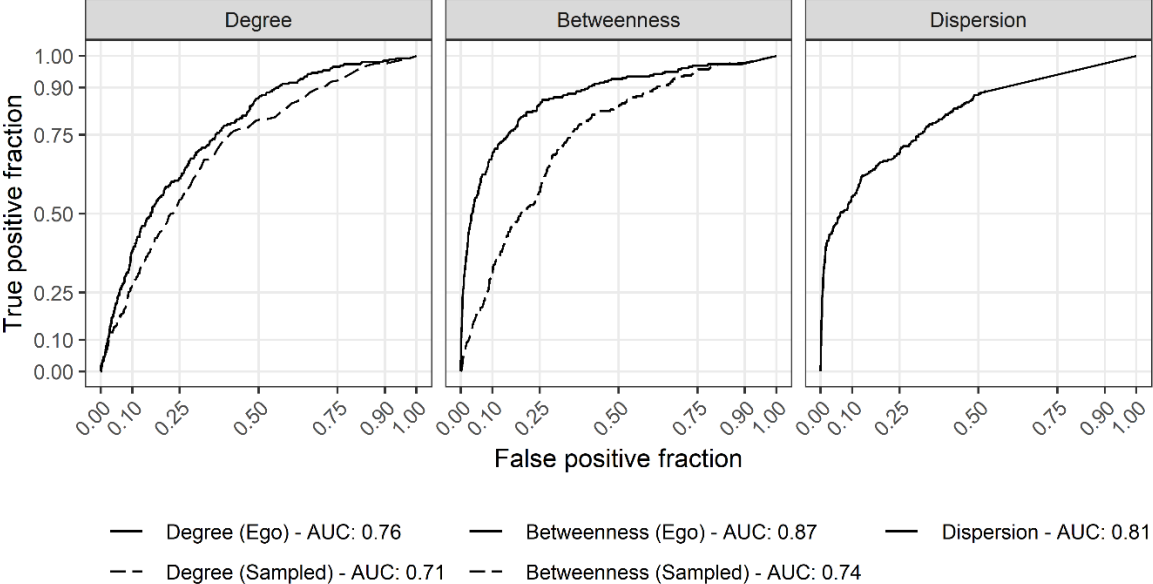
As an initial step in our analysis, we validate our choice of an ego network perspective by comparing it to a sampled network perspective. We test whether merging all ego networks as an approximation towards aggregate "community structures" leads to improvements in the predictive power. The resulting sampled network contains 17,764 nodes and 300,802 edges, with an overlap of 9% of nodes between ego networks.

For degree, betweenness, and dispersion, we compare the predictive power between ego networks and the merged sampled network using ROC curves. Moreover, we report the correlation matrix in Appendix A3. It highlights that for the same network measures the ego and sampled perspective diverge, providing initial support that their interpretation depends on the chosen perspective. ROC curves assess the ability of a continuous measure to rank (i.e., predict) a binary outcome. Classifiers in ROC curves can be any continuous variables alone (e.g., degree) or estimated probabilities between 0% and 100% produced by a prediction model of multiple predictors. To generate a ROC curve, a continuous classifier is ranked in descending order and each observation is successively assessed. For the observation with the highest rank, did the classifier predict the observed value? Starting at the bottom left, if the prediction was correct, the ROC curve extends vertically upward (true positives are mapped on the y-axis) – if not, the curve extends horizontally rightward (false positives are mapped on the x-axis). Then the observation with the second-highest ranking position is assessed, and so forth (see Fawcett, 2004, pp. 6-10). The associated AUC (or "ROC AUC") measure aggregates the ROC curve's performance. It represents the probability that "a randomly chosen positive instance is ranked higher than a randomly chosen negative instance" (Fawcett, 2004). An AUC of .5 reflects random assignment (i.e., ROC curve shows a diagonal line), and 1.0 reflects perfect assignment (i.e., a vertical line adjacent to the "true positives" axis). To

evaluate the discriminatory power of each network measure, we compare the model-free ROC curves in Figure 2.

Note that dispersion (with a model-free AUC of 81%) is a purely ego-network based measure so that it does not exist for sampled networks. For the other centrality measures, we observe the superior performance of the ego-based perspective. Betweenness achieves a model-free AUC of 87% in the ego perspective versus 74% in the sampled network perspective. Similarly, degree in the ego perspective achieves a model-free AUC of 76% versus 70% in the sampled network perspective. This finding implies that a sampled network does not improve predictions given our sampling strategy and selected centrality measures. Hence, our following analysis focuses on centrality measures obtained from the ego network perspective.

Figure 2. Comparison of model-free ROC curves ego vs. sampled



Note: Dispersion only defined in ego networks

As a next step, we assess the differences in the data among our variables of interest. In other words, how are closest friends different from the rest of the network? Table 2 summarizes the collected predictors and contrasts the two sub-groups.

Closest friends have throughout higher ego network centrality measures than non-closest friends. For the remaining binary predictors, the mean can be interpreted as a percentage: 13.3% of the “closest friends” had recently submitted a post to the Facebook wall of the respondent. Among the remainder of observations, this occurred only .8% of the cases. Only for language match, the differences between the subgroups are less striking, as 54.1% of all

closest friends had the same language preference, only slightly above the remainder of the network (48.3%). This suggests that strong ties form across language boundaries, making the predictor less valuable in identifying strong ties among the vast list of connections.

Table 2. Descriptive statistics of strong tie predictors (ego networks)

Group	Predictor Variable	Closest Friends		Not Closest Friends		Type
		Mean	SD	Mean	SD	
Network (Ego) Common Contacts	Degree	.126	.109	.055	.066	Continuous
Network (Ego) Bridging Positions	Betweenness	.050	.077	.003	.013	Continuous
Network (Ego) Bridging Positions	Dispersion	4.418	6.354	.467	1.113	Continuous
Interactions	Profile Post	.133	-	.008	-	Binary
Interactions	Messaged	.372	-	.014	-	Binary
Similarity	Matching Gender	.717	-	.544	-	Binary
Similarity	Matching Language	.542	-	.483	-	Binary

2.3.3 Prediction Model Formulation

Next, we examine how these network predictors complement each other. We assess the collective ability of the predictors to differentiate the *perceived tie strength of a respondent* to a given first-degree *alter a* following the function:

$$ClosestFriend_a = \alpha_0 + \beta_1 Degree_a + \beta_2 Betweenness_a + \beta_3 Dispersion_a + \gamma_1 ProfilePost_a + \gamma_2 Messaged_a + \delta_1 MatchingGender_a + \delta_2 MatchingLanguage_a + \epsilon_a,$$

where “ClosestFriend” is the alter's nomination as closest friend resulting from the survey, i.e., the real-world perceived tie strength. As predictors, we firstly include the *network structure* measures Degree, Betweenness, and Dispersion, denoted with β -coefficients. Secondly, we include the *interaction* measures ProfilePost and Messaged, denoted with γ -coefficients. Thirdly, we include the *similarity* measures MatchingGender and MatchingLanguage, denoted with δ -coefficients.

We perform a supervised learning classification using the linear estimation model logistic regression. To assess non-linear effects (James, Witten, Hastie, & Tibshirani, 2013Chapter 8.1.3), we also assess the performance of tree-based classification via random forests and gradient boosting, as used in prior applications (Jones et al., 2013; Kahanda & Neville, 2009), but cannot improve the prediction accuracy. As logistic regression is an established method and comes with interpretable parameter estimates, we use it throughout the analysis. However, a limitation to the model output is that imbalanced datasets, as in our case, tend to underestimate probabilities (e.g., King & Zeng, 2001). We concentrate our reporting on the predictive results derived through rankings (i.e., ROC curves) and adapt threshold values downwards to account for a systematic underestimation to address this issue.

Following best practice in predictive modeling, we assess the prediction model using hold-out samples (Shmueli, 2010), generated via k-fold cross-validation. To ensure that a sufficient number of positive observations are present in the hold-out sample, we set $k=5$. This means that a random $4/5^{\text{th}}$ of data serves for training purposes (i.e., creating a model), while $1/5^{\text{th}}$ of the data serves for hold-out assessment. We repeat this process another four times so that all data has been considered as hold-out once.

To get an initial impression about the associations to the dependent variable, we report the output of a logistic regression that is estimated with all available data in Table 3. In the model output, degree (5.68, $p < .001$), dispersion (.21, $p < .001$) and betweenness (13.30, $p < .001$) all exhibit significant coefficients, which confirms that all network predictors appear relevant. Moreover, the estimated coefficients are all positive, in line with the predicted behavior. Both interaction-based predictors are also positive and significant. Among similarity-based predictors, only matching gender is positive and significant ($p < .001$), but not matching language ($p > .05$).

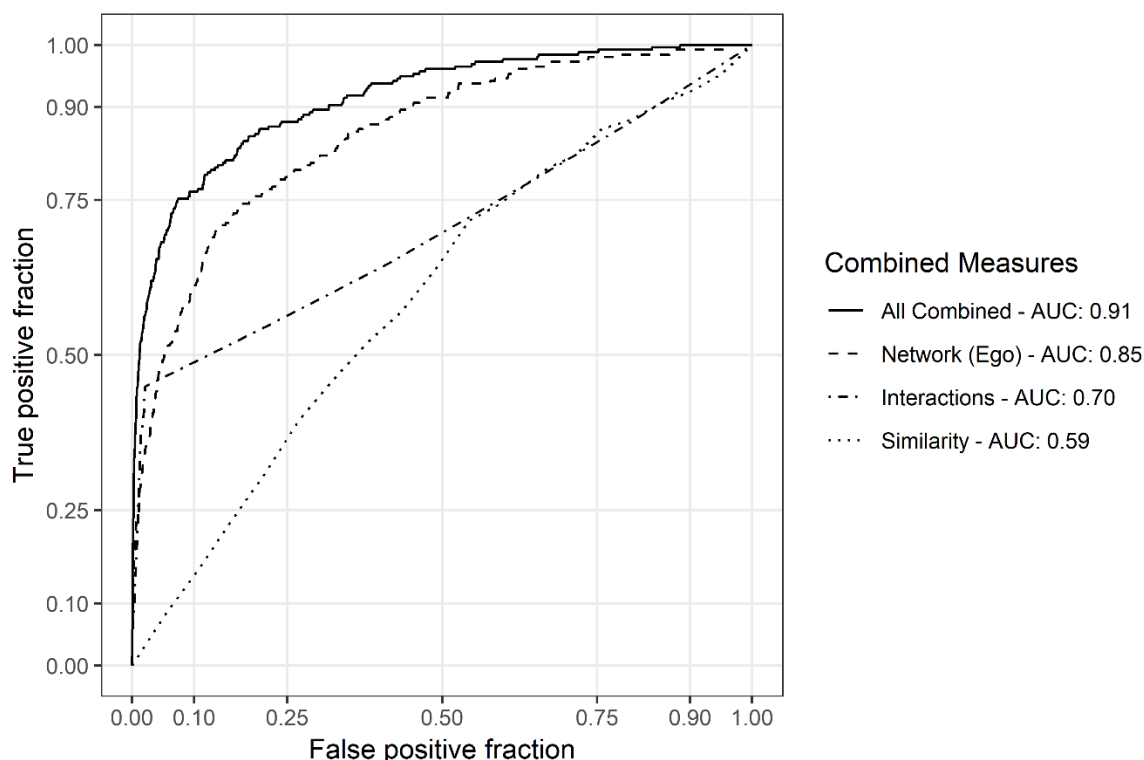
Table 3. Logistic regression ego network model (based on all available data)

<i>Dependent Variable: Closest Friend Nomination (Yes / No)</i>			
<i>Measure</i>	<i>Estimate</i>	<i>Std. Error</i>	<i>Significance</i>
(Intercept)	-6.08	.18	***
Degree	5.68	.69	***
Betweenness	13.30	2.24	***
Dispersion	.21	.03	***
Profile Post	2.12	.27	***
Messaged	3.34	.17	***
Matching Gender	.55	.16	***
Matching Language	.15	.15	
Observations		18,541	
Log-Likelihood		-879.99	
Nagelkerke R ²		.37	

Note: Significance codes: '' .05 '**' .01 '***' .001*

As our methodology's objective is to highlight the predictive abilities of our models, we continue by assessing hold-out samples via cross-validation. Figure 3 shows the corresponding ROC curves and AUC values. The ROC curve exhibits a steep increase for the interaction-based predictor group, which indicates that it is similarly successful in identifying the most likely (order ranked) close friends compared to the network-based group, but less so when identifying the remainder. The non-curvilinear shape of the interaction-based group is attributable to the binary nature of the measures. The network-based predictor group exhibits strong discriminatory power in both models. In isolation, the network centrality group model achieves an AUC value of 85%. The “combined” model with all predictors outperforms all other models with an AUC of 91%. This relatively low incremental uplift shows that the ego network measures largely drive the model fit.

Figure 3. ROC curves for combined models (logistic regression, 5-fold cross-validation)



2.3.4 Practical Demonstration

Our dataset contains a very imbalanced response class so that the naïve classification of predicting the dominant response for all observations results in an accuracy of about 98.6% (i.e., 18,283 / 18,541). Therefore, to demonstrate our model’s practical value, we instead focus on its ability to identify the few rare positives, reflected in the precision. The precision score indicates the share of actual strong ties among all predicted strong ties, as classified by the model. Table 4 provides the corresponding confusion matrix obtained from the k-fold hold-out samples of the “all combined” model.

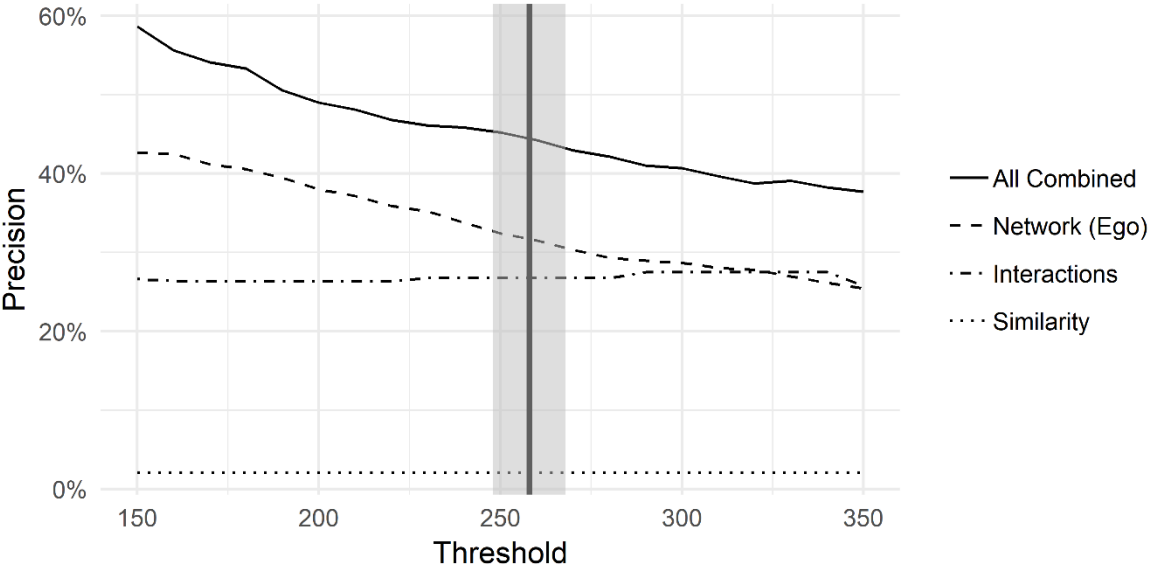
Table 4. Confusion matrix “all combined” model

<i>All Combined (Threshold = 258)</i>		
<i>Actual</i>		
<i>Prediction</i>	<i>No Strong Tie</i>	<i>Strong Tie</i>
<i>No Strong Tie</i>	18,140	143
<i>Strong Tie</i>	143	115

Note: Precision: 45% = 115 / (115 + 143)

Since researchers or practitioners do not know the exact number of strong ties of novel data (represented by our hold-out samples), they can only base their estimates on the number of strong ties in the observed data sample (in our case 258). We perform a sensitivity analysis in Figure 4 for corresponding threshold values of 150 to 350 respondents to reflect this uncertainty. The threshold values refer to the point in the continuous probability distribution generated by the prediction model at which the classification switches between “strong tie” and “no strong tie”.

Figure 4. Precision scores sensitivity analysis



As observable from the confusion matrix in Table 4, using the “all combined” model, we classify 115 of the rare 258 strong ties correctly at the “true cut-off” value 258, which corresponds to a precision of 45% (i.e., 115 / 258). The sensitivity analysis on the precision scores, in Figure 4, shows that the network-based model (precision of 32% at 258) outperforms the interaction-based model (precision of 27% at 258). However, predictions on the sparse similarity predictors alone only achieve low results (precision of 2% at 258). In general, if we include fewer predictions in the hold-out dataset (with higher probabilities), the precision of the models tends to be better. While network-based predictions are more precise for a lower threshold, the predictions of the interaction-based model are more stable. The low granularity of the interaction and similarity data may explain this result because it leads to multiple observations having the same probability score.

In a final robustness test, we examine how the size of the ego networks affects precision. We split the 41 ego networks into three similar-sized groups: small (14), medium (14), and large (13). Large networks and medium-sized networks achieve the best precision (49% and 45% respectively), in contrast to the small networks (37%). Small networks coincide with lower OSN usage intensity, and thus fewer valid predictors. The 14 smallest networks host only 2,663 of the 18,541 ties and are represented less in the training data.

2.4 Display Advertisement Experiment: Impact of Tie Strength on Clicking Behavior²

The key assumption for this research is that knowledge about tie strength is beneficial for social media marketing. Beyond the motivating examples in Section 2.2.1, we aim to demonstrate the benefit of such knowledge in a display advertisement experiment. In this setting, we assess the impact of framing social advertisement messages with names of real-world close friends, good friends, and acquaintances on the tendency to click on the message.

In a survey, we asked participants to indicate the names of their closest friends. As a benchmark to reflect sub-groupings with decreasing tie strength intensity, we asked for the names of good friends and some of their acquaintances (Arnaboldi et al., 2013). After some intermediate questions about social media usage to distract the participants, we displayed six dynamically manipulated advertisement messages and asked respondents for their propensity to click on this message on a 7-point Likert scale.

We created the display advertisements using Adobe Photoshop to mimic the exact layout and feel of a Facebook advertisement message, which is illustrated in Appendix A4. We randomly varied the following conditions: The displayed ad picture (three variations of images featuring a female and male model), the number of likes (15 vs. 1,532), but most importantly, whether one of the close friends, good friends, or acquaintances also liked the page. A total of 109 participants, with comparable demographic characteristics as in our first study, completed the survey.

We estimate a linear regression model displayed in Table 5. Displaying a close friend's name among the likes, we observe a significant ($p < .05$) increase in the propensity to click. In comparison, good friends and no mentions yield a positive but insignificant effect on the propensity to click. In terms of effect size, the mention of a close friend's name was about as

²This section grounds on research data that was collected as part of the Bachelor thesis (Hoelzer, 2019)

strongly associated with the propensity to click compared to over 2,000 additional likes. In conclusion, the results suggest that for the example of social display advertisement, particularly knowledge about the few real-world closest friends is helpful.

Table 5. Linear regression for propensity to click

<i>Dependent Variable: Propensity to Click on Advertisement</i>			
<i>Measure</i>	<i>Estimate</i>	<i>Std. Error</i>	<i>Significance</i>
(Intercept)	2.381	(.275)	**
Close Friend	.380	(.118)	**
Good Friend	.148	(.136)	
No Mention	.161	(.146)	
(Baseline: Acquaintance)			
High Like Count (+ 1,517)	.219	(.087)	*
Image Variation 2	-.146	(.109)	
Image Variation 3	-.270	(.105)	**
(Baseline: Image Variation 1)			
Gender (Male = 1)	-.145	(.090)	
Age	-.002	(.009)	
Observations	654 (= 6 · 109)		
R ²	.043		

Note: Significance codes: '' .05 '**' .01 '***' .001*

2.5 Discussion

2.5.1 Summary

As our analysis shows, firstly, ego network centrality measures obtained from OSNs indicate tie strength. In particular, the predictive accuracy of the two bridging indicators betweenness and dispersion is high. Both are outperforming the established network measure of common contacts, i.e., structural embeddedness. We advance the argument of Backstrom and Kleinberg (2014) about bridging positions in ego networks by pointing to social capital theory to explain how these bridging positions could evolve. We show that this theory on bridging positions extends beyond romantic relationship partners to predicting survey-based tie strength assessments among friends. In contrast to the theory that “no strong tie is a bridge” (Granovetter, 1973, p. 1364), our results suggest that in the constrained perspective of

ego networks, the opposite is true: Individuals that bridge between clusters (i.e., social circles) are the strongest ties.

Secondly, we demonstrate how the set of network-based predictors can complement other predictors. Like Kahanda and Neville (2009) and Jones et al. (2013), we find that our selection of interaction-based predictors has higher discriminatory power than similarity measures. Still, the combination of interaction-, similarity-, and network-based measures yields the highest overall scoring. While the varying number of friend name mentions and varying sizes of ego networks impose substantial variation to the classification problem, the model's predictive power, captured in the AUC value of .91, is remarkable.

Thirdly, we illustrate the practical value of the model in an assessment of precision scores. By identifying only the few “closest” friends, we reflect the assumption that managerial interest is focused on the few most important influencers of a person (Trusov et al., 2010). For example, in our setting, a referral campaign targeted to the closest friends of an individual can be expected to correctly identify around 45% of all closest connections of the seeding person. Moreover, we note an improvement in the precision between purely network-based and the “all combined” model. This suggests that the interplay between multiple predictor classes can generate better results, so that we recommend assessing multiple types of predictors, when possible.

Fourthly, we show that combining the ego networks to a single network sample does not improve tie strength prediction, as often used in studies to approximate full network structures (Ebbes et al., 2016). The relevance of degree and betweenness diminish when merging ego networks, as other nodes may distract from identifying positions between social circles. This implies that the information contained in bridging positions in ego network structures represents further strong revealed preference measures for tie strength, beyond common contacts. Even when researchers have complete network information, for an assessment of the largest influencers of a given person (e.g., the person’s closest friends), it may be beneficial to iteratively extract ego networks and compute the proposed centrality measures for each alter. Moreover, the findings support our proposition that ego networks alone can serve as input in predicting tie strength and, therefore, explain various phenomena of influence in marketing.

2.5.2 Theoretical and Managerial Implications

Our study highlights an alternative tie strength predictor, the bridging positions in ego networks. Our results suggest that researchers and practitioners can gain insight into tie

strength perceptions and, consequently, peer effects by adding ego network measures to their set of “revealed preference” measures. This is particularly valuable in online settings, where the actual perceived tie strength is difficult to obtain. We conjecture that the inclusion of bridging positions of ego networks is beneficial to various applications of tie strength in marketing research, where other network measures, like common contacts, are frequently applied (e.g., Aral & Walker, 2014; R. Bapna et al., 2017).

Various arguments underpin the relevance of this novel measure. Firstly, there exists a discrepancy between perceived tie strength attributes, such as trust, and “revealed preference” tie strength measures, mutual friends, and interactions (R. Bapna et al., 2017). We show that the perceived tie strength of “closest friends”, that may be more strongly associated with trust, can be better approximated via ego bridging measures than the frequently used network measure of common contacts. Secondly, the common tie strength measure interaction has the undesirable properties of being sparse, temporal (Xiang et al., 2010), and incomplete (Wiese et al., 2015), as every person uses a range of on- and offline communication channels. Also, messages related to life events, such as birthday congratulations, distort interaction measures (Viswanath et al., 2009) and may be subject to social visibility biases (Shmargad & Watts, 2016). Thirdly, settings occur when user attributes for a similarity-based prediction are sparse or not captured (Rotabi et al., 2017). In such a setting, knowledge about network measurements that approximate tie strength is especially valuable. Fourthly, in comparison to sociographic network analysis, ego networks are easier to collect. Often information on who the ego is connected to and how these alters are interconnected can be obtained from the person representing the ego node alone (Everett & Borgatti, 2005). Lastly, the difference in size makes processing ego networks computationally inexpensive and parallelizable.

Moreover, our findings contribute to data and privacy protection regulations, as they demonstrate how sensitive information can be inferred from the granular network data present in OSN. A lot of the debate on data privacy in OSN revolves around personality attributes: Facebook “Likes” predict personality traits with high accuracy (Youyou, Kosinski, & Stillwell, 2015), but also allow to infer sexual orientation, political views, intelligence, and ethnicity of a user (Kosinski et al., 2013). Our study casts light on the value of another type of data abundant in OSN: Social network structures. We find that not only romantic relationship partners can be identified (Backstrom & Kleinberg, 2014), but perceived tie strengths can be predicted, even when a user does not disclose these attributes knowingly, for example, by tagging favorite friends or messaging.

Our research is not only relevant to big platforms that own network data. Friend lists are a core feature of OSN and common in social media (Kane, Alavi, Labianca, & Borgatti, 2014). To facilitate integration with other websites and apps, accessing these lists is often possible via APIs. For smaller networks that do not provide APIs, website crawling can be used, if permissions are granted (see examples on Leskovec & Sosič, 2016). The existence of social circles in other OSN settings (Leskovec & McAuley, 2012) implies that adaptations of this approach can be extended to various other OSN platforms. Moreover, various digital services incorporate social networking features into their service (e.g., LastFM, Chess.com). The relevance of the proposed approach is not confined to OSN but can potentially be extended to predictions of tie strength in other on- and offline datasets, such as e-mail data, messaging services and phone records, which reflect real world social networks. Patent filings of prior tie strength prediction approaches demonstrate the economic importance of this research field (Gilbert & Karahalios, 2015).

2.5.3 Limitations and Future Research

One limitation is that we approximate the full network by merging the individual networks of the egos. As the full network structure is unknown to us, we cannot say whether ego networks generally predict tie strength better than full networks. Moreover, we utilize a name matching approach that allows us to assess large amounts of network data. This setup has inherent limitations as we make the simplifying assumption of capturing a single binary metric to assess tie strength. However, tie strength is a multidimensional and continuous construct for which our chosen approach cannot fully account (Marsden & Campbell, 1984). This approximation (and the lack of more granular data on the nodes) might have diminished the predictive power for the analysis.

We strongly encourage future research to build on our findings by exploring the usage of bridging positions in ego networks as an additional measure when evaluating peer influence phenomena in OSN. Further, future studies might want to validate our findings via alternative sampling approaches for large networks (Ebbes et al., 2016). However, a comparison to full real-world networks (e.g., Facebook) is computationally barely feasible. Moreover, future work is needed to explain how ego network structures in OSN emerge over time and under which circumstances bridging positions evolve.

3 Social Capital Accumulation and Premium Memberships in Work-Related Networks³

3.1 Introduction

Social capital (SC) is linked to all kinds of social phenomena (see Portes, 1998; Weiler & Hinz, 2019b for reviews). While SC has received varying interpretations, it essentially reflects the notion that individuals can access tangible and intangible resources through their connections (Adler & Kwon, 2002; Portes, 1998). Among the most widely studied relationships is the one between SC and work-related outcomes. They have been a focal point in early works of social network theory (Granovetter, 1973), with a variety of research finding that SC is positively linked to hiring outcomes (e.g., Gee, Jones, & Burke, 2017), salaries, and career paths (e.g., Seibert, Kraimer, & Liden, 2001).

Today, work-related social media networks (SMNs), like LinkedIn (LI), promise to support the accumulation of SC in an online setting. They intend to support users in managing their professional network and in finding new jobs or other job-related opportunities. Work-related SMNs typically operate under a so-called freemium business model (R. Bapna & Umyarov, 2015; Voigt & Hinz, 2016), i.e., they offer access to the platform free of charge, but require a fee-based premium membership to unlock advanced networking features. Despite their popularity for the job-search process, only few empirical papers assess the value they actually provide (Forret, 2018; Garg & Telang, 2018). Specifically, the literature lacks an understanding of how the advanced networking features available under premium membership increase SC.

Prior research in information systems (IS) points to largely different mechanisms that could explain how premium membership could support SC accumulation: First, premium

³ This chapter is based on the work “How Conversions in Freemium Businesses Impact User Activity” presented at the 42nd INFORMS Marketing Science Conference, Duke University, Fuqua, America (Stolz, Schlereth, & Lanz, 2020). The work is forthcoming at MIS Quarterly under the title “Social Capital Accumulation Through Social Media Networks: Evidence from a Randomized Field Experiment and Individual-Level Panel Data” with co-authors Dr. Michael Weiler, Prof. Dr. Andreas Lanz, Prof. Dr. Christian Schlereth, and Prof. Dr. Oliver Hinz (2022).

As referenced, we built our analysis around section 5 of the dissertation by Dr. Michael Weiler in which he presents his joint work with Prof. Dr. Oliver Hinz (2019a). We investigate research question 1 by reviewing their experiment. Next to replicating their analysis with new data, we substantially revised the manuscript, added a new research question 2, and provide a new framework and analyses. In particular section 3.3 and Appendices A6 to A11 are relating to their experiment. To reflect the independent contribution of my dissertation, I reference analyses, plots, and direct quotes that appear in Weiler and Hinz (2019a) in the text.

badges are common in work-related SMN. Prestigious items confer status, which is closely associated with SC according to social resource theory (Lin, 1999). Empirical findings in virtual communities highlight that ownership of such prestigious items increases SC (Hinz, Spann, & Hann, 2015). Second, premium users can identify their profile visitors. In dating SMNs, where profile browsing visibility likewise depends on the membership type, identifiable profile visits provide “weak signals” of interest to a focal user, which significantly drives matching success (R. Bapna, Ramaprasad, Shmueli, & Umyarov, 2016). These two mechanisms depend on the action of the surrounding network, i.e., they are passive. Third, network scholars (e.g., Borgatti & Halgin, 2011) argue that the individual’s agency is pivotal in SC accumulation. As findings in a music freemium service show, premium converters disproportionately increase their activity alongside SC (R. Bapna, Ramaprasad, & Umyarov, 2018). This explanation, in contrast, emphasizes the activity of the focal users themselves. In the work-related SMN, advanced networking features enhance both sides, active and passive, but it is unclear, which of the two is responsible for the SC accumulation.

With this research, which extends and reviews the findings of Weiler and Hinz (2019a), we seek to examine the impact of having access to the advanced networking features provided by work-related SMNs on SC gains over time (RQ1)—and to identify which specific types of features (active vs. passive) drive these gains (RQ2). As pointed out in Sundararajan, Provost, Oestreicher-Singer, and Aral (2013), despite their inherent relevance for the IS discipline, these questions have received little scholarly attention so far. This is remarkable, given that work-related SMNs enable easier SC accumulation in comparison to offline contexts. Among other things, SMNs eradicate spatial and temporal boundaries in contact formations, make contact lists explicitly visible (Kane et al. 2014), provide algorithmic contact suggestions (Liben-Nowell & Kleinberg, 2007), and stimulate contact formation by pointing to similarities between users (Sun & Taylor, 2020). Hence, SMNs have “potentially altered the processes by which social networks evolve” (Sundararajan et al., 2013, p. 895) and given rise to different network formations, depending on their features (Kane et al., 2014).

To date, the empirical findings in the context of work-related SMNs and SC are mostly derived from cross-sectional data—only a few exceptions use longitudinal data, such as Utz and Breuer (2016). These findings largely agree that one’s digital presence on work-related SMNs is advisable (Nikitkov & Sainty, 2014), especially due to the platform’s professional informational benefits (Utz, 2016; Utz & Breuer, 2016). Moreover, SC maintained within SMNs can play an important role in deriving job benefits (e.g., Aten, DiRenzo, & Shatnawi,

2017; Garg & Telang, 2018). These studies implicitly assume that users exogenously receive the nurtured SC, which is responsible for the corresponding job-related outcomes. The scarcity of studies that address online SC accumulation in a causal fashion can be explained through the difficulty to alter SC via experimental stimuli. Prior studies in IS have found that stimuli that alter the user's status (Hinz et al., 2015) and reduce social boundaries (S. Bapna & Funk, 2021) can result in SC increases in specific settings. With our study we seek to build a broader understanding about how individuals accumulate digitalized SC in general, how this is affected by their individual agency (Ahuja, Soda, & Zaheer, 2012; Stuart & Sorenson, 2007), and what is the role of networking features in particular (R. Bapna et al., 2016; Kane et al., 2014).

To investigate the causal evidence of SC accumulation, we review the field experiment of Weiler and Hinz (2019a), in a major European work-related SMNs (to which we refer to “WorkSMN” hereinafter) and replicate it with individual-level panel data. They issued a free, 12-month premium membership to a special segment of users in the SMN, i.e., freelancers, because they rely on their social network to succeed economically (Van den Born & Van Witteloostuijn, 2013; Wu, 2013). Potential freelancers were randomly assigned to the treatment group, which granted them access to an array of advanced networking features through a premium membership, whereas the control group did not receive such a membership. In other words, the premium conversion serves as treatment variable (as in R. Bapna & Umyarov, 2015). As discussed in Weiler and Hinz (2019a, p. 188), by randomly assigning individuals into a treatment and a control group, they could “address the endogeneity problems that often plague SC research, such as omitted variable bias [...], measurement error, and simultaneity bias”. Hence, their experimental procedure ensures that the estimated effect of premium conversion is unbiased.

To replicate these findings and provide complementary insights into which type of features drive SC accumulation (RQ2), we acquired a second dataset consisting of individual-level panel data that allows us to analyze individual behavior over time. This panel data covers 52,392 freelancers who were targeted by discount mailing campaigns. This dataset contains users' digital footprint data in two directions: on the active side, it captures the outgoing activity of users; on the passive side, it captures incoming activity of the surrounding network. It is complementary to the field experiment, as it provides insights into the types of features that support SC accumulation and allows us to investigate the dynamic evolution of SC and activities.

Our first study reviews the field experiment of Weiler and Hinz (2019a) and the complier average causal effect (CACE) of the experiment. Their findings suggest that “freelancers do not automatically change their digitalized networking engagement just because they have access to” advanced networking features—instead, those premium features can “only prove their full value if the freelancers are also motivated to proactively and purposely utilize the given resources as part of strategic networking behavior” (Weiler & Hinz, 2019a, p. 189). Specifically, Weiler and Hinz (2019a) find that scoring one additional point on the strategic networking behavior scale increases freelancers’ SC under premium membership by 4.609%⁴. Our individual-level panel data backs up this finding: Every doubling in the number of contact invites sent before the discount mailing resulted in SC increases of approximately 4.148 additional contacts among freelancers who converted to premium. Passive features (e.g., prestigious premium badges), which make users more salient, are also positively linked to SC accumulation. Yet, their impact is substantially lower compared to active features (e.g., personal messages to non-contacts). Thus, we concur that “the possession of an efficacious ‘networking weapon’ is not enough by itself; it must also be accompanied by the intent to ‘shoot it’” (Weiler & Hinz, 2019a, p. 189) and provide further evidence for their “*theory of purposeful feature utilization*”, which reflects that users need a motivation to proactively utilize advanced networking features.

Our work contributes to the IS literature in multiple ways. We conceptualize the dynamics of the evolution of SC accumulation and the role of advanced networking features in work-related SMNs. Using two complementary data angles, we not only tease out causal effects and test for robustness, but we also disentangle the value that passive features contribute to the SC accumulation, relative to active features. Our research is relevant for work-related SMNs as they gain a better understanding of which advanced networking features foster SC accumulation among their users. It is in the best interest of SMNs to encourage networking to gain a competitive advantage from network effects (Shapiro & Varian, 1999). Our results show that giveaway trials of premium memberships, which are a common marketing measure among freemium networks (Koch & Benlian, 2017), will do little to encourage users to accumulate SC if they have a low agency for networking.

Moreover, for the product managers of work-related SMN, our study points to the role of active features in SC accumulation, which is an important consideration in managerial

⁴ Weiler and Hinz (2019a) report 4.593%, which we slightly refined in the current review by re-estimating the model with different covariates.

decisions concerning pricing, versioning, and feature development (Shapiro & Varian, 1999). Finally, platform users benefit from our insights: A premium membership alone is not sufficient to increase SC. Only premium users willing to actively approach others (e.g., through personal messages) will be able to derive additional SC.

The remainder of the paper is structured as follows: Section 3.2 reviews the related research on SC and networking in an online setting. Section 3.3 reviews the causal evidence of SC accumulation from the field experiment (RQ1) of Weiler and Hinz (2019a). Section 3.4 complements the experimental insights into which types of features drive SC accumulation (RQ2): firstly, by providing evidence for the robustness of these findings; secondly, by developing further insights from the individual-level panel data. In the final Section 3.5, we discuss our results and outline potential future work.

3.2 Theoretical Background

3.2.1 *Social Capital*

Since the concept's mainstream scientific inception almost 40 years ago, a plethora of different SC definitions have permeated the disciplines of social science (i.e., sociology, economics, and political science—see Adler and Kwon (2002) for an overview). These myriad definitions generally align with two camps: individual-level and collective-level SC (Adler & Kwon, 2002; Borgatti, Jones, & Everett, 1998). In broad terms, representatives of the individual-level perspective explicitly focus on the individual and his/her relationships with fellow humans. This perspective perceives SC as a private good of one individual (i.e., a node in a network). Thus, researchers commonly use network analytical measures, like the node degree, to operationalize it (Borgatti et al., 1998). Maintaining relationships with others provides individuals with exclusive access to all kinds of tangible and intangible resources, which in turn will ultimately assist them in attaining their specific goals. In the words of Nahapiet and Ghoshal (1998, p. 243), SC is “[...] the sum of the actual and potential resources embedded within, available through, and derived from the network of relationships possessed by an individual or social unit. Social capital thus comprises both the network and the assets that may be mobilized through that network”. Thus, “[...] social capital is a ‘metaphor about advantage’” (Takac, Hinz, & Spann, 2011, p. 189), which helps to explain why some individuals are evidently more successful than others.

Scholars who look at the collective-level perspective of the concept (e.g., Putnam, 2000) focus on larger social structures, such as groups and communities. They look at the members’

ties, which exist among each other within these collectives and assess whether these groups eventually achieve cohesion (Adler & Kwon, 2002). This perspective emphasizes that every individual in the collective will benefit from the created SC, regardless of whether he/she put effort into the creation process (Borgatti et al., 1998). Thus, according to this perspective, “[...] social capital is a collective good and [...] it is non-exclusive in consumption [...]” (Rostila, 2011, p. 311).

In this study, we follow Nahapiet and Ghoshal’s (1998) understanding of individual-level SC. We do so because it is the individual user who eventually benefits from his/her resources on the SMN platform and thus reaps assets, such as information about potential working opportunities.

3.2.2 *Networking*

Networking refers to the process of cultivating and establishing SC (Forret & Dougherty, 2001). The term “networking” is generally understood to mean a “[...] strategic (i.e., rationally motivated) behavioral effort that involves the dyadic exchange of interpersonal resources, which are directed toward building and maintaining network relationships with specific network contacts and motivated by whether they have access to specific interpersonal resources” (Porter & Woo, 2015, p. 1481). For instance, individuals who engage in such behavioral efforts of networking are rewarded with a higher salary (e.g., Wolff & Moser, 2009) and more job offers (e.g., Van Hove, Van Hooft, & Lievens, 2009).

Although individuals are aware of the importance of networking, they often do not, or only reluctantly, participate in such behavior, whether due to difficulties in socializing or a belief that they are insincere by establishing ties solely for instrumental purposes (Casciaro, Gino, & Kouchaki, 2014; Kuwabara, Hildebrand, & Zou, 2018). Others, by contrast, enjoy creating instrumental ties and actively strive to establish them (Bensaou, Galunic, & Jonczyk-Sédès, 2014). To put it differently, the motivation to proactively adjust one’s career trajectory largely seems to depend on an individual’s beliefs and attitudes toward such instrumentally oriented behaviors (Kuwabara et al., 2018). Next to attitudes of networking behavior, experimental evidence indicates that specific external interventions can promote and encourage networking. For instance, S. Bapna and Funk (2021) highlight that female IT conference participants who receive a nonreciprocal list of contact recommendations significantly extended their digitalized SC on LI. This observation raises the question of whether online stimuli can likewise positively affect SC.

The process of networking has greatly changed due to the introduction of SMNs. The online setting eliminates frictions that exist in an offline setting: SMNs provide explicit contact lists (Kane et al., 2014) and support networking through a range of features (Karahanna, Xu, Xu, & Zhang, 2018). Individuals not only receive recommendations on potential contacts directly from SMNs, but they can also leverage the IS-inherent search functions to identify potential contacts (Kane et al., 2014). Therefore, SMNs mirror a huge repository of diverse social contacts, which the user can tap into and access other individuals who may otherwise be (spatially) out of reach. SMN users can also browse through their personal network members' profiles, which allows them to inform themselves about their contacts' friends, whom they can in turn easily befriend by simply sending a contact request.

Moreover, SMNs commonly provide users with contact suggestions that are algorithmically derived based on network structures (Li, Fang, Bai, & Sheng, 2017; Liben-Nowell & Kleinberg, 2007) and commonalities between users (Sun & Taylor, 2020). Facilitating network formation among users is a primary strategic interest to SMN platform providers, as a better-connected network reflects a lock-in of users in the form of network effects (Shapiro & Varian, 1999). In other words, users are unwilling to undertake the effort to recreate their network on a competing SMN platform. As contact lists evolve over several years of usage, SMNs form an ensemble of actively maintained and inactive (dormant) ties, both of which can be valuable in terms of SC (McCarthy & Levin, 2019).

This rich set of features enables users of SMNs to not only maintain their offline contacts, but also reach out to users who they do not know well offline or even at all. Manago, Taylor, and Greenfield (2012, p. 6) found that on Facebook, networks grow “primarily through relatively more distant kinds of relationships”. In work-related SMNs, such weak connections are particularly commonplace (Garg & Telang, 2018). Freelancers—the target group of our study—especially rely heavily on SC to acquire project-based labor arrangements and can achieve job security through a diversified set of contacts (Van den Born & Van Witteloostuijn, 2013; Wu, 2013). As in the offline world, looking at each other and initiating a conversation are typically the first steps to forming a connection. The equivalents to such offline behaviors in an SMN context are profile visiting and messaging (R. Bapna et al., 2016).

In summary, while the intrinsic agency for networking differs across users, SMNs strongly support SC accumulation through a wide range of features, eliminating frictions that exist in offline contexts (Kane et al., 2014; Karahanna et al., 2018). We expect in work-

related SMNs, and for freelancers in particular, that this desire for SC accumulation through more distant connections is strong, such that advanced networking features will be particularly helpful for accumulating SC.

3.3 Accumulating Social Capital: Causal Evidence (RQ1)

3.3.1 *Experimental Set-Up*

Weiler and Hinz's experiment (2019a) aimed to causally test whether freelancers with access to advanced networking features accumulate more SC than those without (RQ1). It built on the idea that such features in a work-related SMN can bolster freelancers' networking opportunities, meaning their ability to accumulate SC. As Weiler and Hinz (2019a, p. 194) describe, this general notion aligns with the affordances concept, in "which SMN features are seen as the foundation for influencing the users' action possibilities." (see also Bucher & Helmond, 2017; Leonardi, 2013). In other words, users approach the same SMN features with different motivations: "[...] [U]sers will appropriate certain features of a technology only when they perceive that those features offer them affordances for action [...]" (Leonardi, 2013, p. 752).

Weiler and Hinz (2019a) set up an experiment in one of the largest European work-related SMNs ("WorkSMN"), which features approximately 17 million users. Like its American counterpart, LI, WorkSMN enables its users to accumulate and sustain SC in the form of professional contacts and to find new jobs or projects—either free of charge or by signing up for a premium membership. Examples of such advanced features of the premium membership are functionalities that enable users to see which other users visited their profile, enhanced search capabilities, or the possibility to send messages to non-contacts (see Appendix A5). Essentially, users' networking capabilities are either more constrained or enabled based on their type of membership.

In the experiment of Weiler and Hinz (2019a) and our individual-level panel data we focus on freelancers, a population that is barely represented in current research (Lo Presti, Pluviano, & Briscoe, 2018), because they are quite difficult to access (Kuhn, 2016). Nonetheless, freelancers epitomize a promising research population in our context due to their heavy reliance on networking and the resulting SC to gain projects and succeed in their careers (e.g., Van den Born & Van Witteloostuijn, 2013). Qualitative studies highlight that freelance musicians value networking via SMNs (Haynes & Marshall, 2018) and that today's freelancers see an active SMN profile as imperative to their success (Gandini, 2016). So far,

few studies have used quantitative data to examine how freelancers engage in a digital environment and those, which have, paid more attention to how freelancers behave in online labor markets (e.g., Leung, 2014; Shevchuk & Strebkov, 2018) or enterprise SMNs (Wu, 2013) rather than in work-related SMNs. Hence, there is a need for deeper insights into this research population and how they increase and utilize their SC in work-related SMNs.

Weiler and Hinz (2019a) recruited freelancers as participants for their experiment through a pre-study. This pre-study was solely intended for recruitment purposes. They used the premium features of WorkSMN to search for users who self-reported themselves as freelancers in their digital user profile and asked them to fill out a questionnaire about “The perceived usefulness of work-related SMNs in generating job offers”. The questionnaire concluded with two questions that determined whether respondents “qualify for the experiment: (1) whether they are willing to use a donated premium membership for six months and (2) whether they are willing to support further studies from the researchers” (Weiler & Hinz, 2019a, p. 195). If the participants responded affirmatively to the two questions, they were forwarded to another questionnaire. This questionnaire then asked for the respondent’s personal contact details, name, and WorkSMN profile ID.

In total, it took them “roughly ten months (January to October 2017), and approximately 6,350 manually sent contact requests to collect a sample of 243 eligible freelancer responses” (Weiler & Hinz, 2019a, p. 195). Furthermore, they state that “among the remaining 217 freelancers, we used random number seeds to select 75 participants who received a voucher code to upgrade their basic account to a premium membership.(...) The platform service provider graciously donated the vouchers.” (Weiler & Hinz, 2019a, p. 196). For simplicity, we refer to those who received the vouchers as treatment group, and those 142 who did not receive vouchers (non-gifted) as the control group. Further, Weiler and Hinz (2019a, p. 196) describe that

“we did not want the members of the treatment [and control] group to perceive that they were participating in an experiment. Thus, we explicitly communicated to [the treatment group] that they won a free, 12-month premium membership as a thank-you gift for participating in our pre-study. Our message [...] included the premium voucher code, the redemption instructions, and a notice about the code’s automatic expiration outside of a special period to discourage them from postponing their redemption. Treated freelancers who did not redeem their voucher after two weeks received a personal reminder note via their WorkSMN profile.”

In summary, among the treated group, a total of 65 freelancers converted to premium.

3.3.2 Data

Before the experiment, Weiler and Hinz (2019a) manually harvested digital footprint data from each freelancer's profile. The data included "gender", "freelancers' tenure on WorkSMN", profile details, such as freelancers' self-reported number of "haves", "wants" (i.e., tags that can be set on the profiles), and "number of subscribed groups". While the "have" section allows the freelancers to promote their skills, products, and services, the "wants" section enables them to specify what they are looking for, e.g., finding new project partners. They additionally collected the "number of direct contacts" shortly before the experiment and after the experiment had run for about six months. Across many disciplines, including IS (e.g., Hinz et al., 2015), this metric is a well-established operationalization of SC. In short, individuals who are well-connected necessarily maintain access to a wider set of valuable resources (Borgatti et al., 1998). Lastly, as their dependent variable they computed the average percentage change in the number of direct contacts over six months, reflecting the increase in SC among participants (" Δ SocialCapitalAccumulation").

While Weiler and Hinz (2019a) collected the number of contacts (SC) from the freelancers' profiles for the second time, they excluded two participants from their analysis: Technical issues occurred in the data collection for one of them and the other deleted the SMN account. After their exclusion, they had a final sample size of 215 freelancers. We show a flow diagram to summarize the procedure of Weiler and Hinz (2019a) in Appendix A6.

Table 6 summarizes the covariate balance check to test for differences between the treatment and control groups. Due to unequal sample sizes, we report the Welch's t-test. Treatment ($n = 74$) and control ($n = 141$) group do not exhibit statistically significant differences for any of the pre-treatment variables ($p > 0.05$). We conclude that the randomization was successful and both groups are comparable.

Six months later, the participants received a follow-up questionnaire. Weiler and Hinz (2019a) utilize two mechanisms to build trust and further motivate participants (for a discussion of such mechanisms see Laurie & Lynn, 2009): First, they sent out summary slides about the pre-study and, second, also sent out a voucher (of 5€) for an online retailer accompanied by a message that the continued participation would be highly appreciated.

Table 6. Covariate balance check before manipulation—adapted from Weiler and Hinz (2019a, p. 198)

<i>Covariate</i>	<i>Treatment (assigned)</i>	<i>Mean (SD)</i>	<i>Difference</i>	<i>t-value</i>	<i>p-value</i>
Gender (Female = 1; Male = 2)	0 (n = 141)	1.567 (0.497)	-0.068	-0.965	0.336
	1 (n = 74)	1.635 (0.485)			
Social Capital (Number of direct contacts)	0	153.348 (153.882)	-16.693	-0.668	0.506
	1	170.041 (183.990)			
Number of groups	0	6.482 (7.858)	0.360	0.363	0.717
	1	6.122 (6.372)			
SMN Tenure (in years)	0	7.028 (3.969)	-0.310	-0.557	0.578
	1	7.338 (3.815)			
Number of ‘haves’	0	11.879 (10.430)	-1.229	-0.855	0.394
	1	13.108 (9.776)			
Number of ‘wants’	0	4.333 (5.618)	0.211	0.313	0.754
	1	4.122 (4.151)			

The follow-up questionnaires each contained a six-digit token (e.g., “Xv5P32”) to match the participants’ “digital footprint” data with their self-reported survey data. Weiler and Hinz (2019a) collected several covariates: socio-demographic variables like “age” (M = 44.94 years; SD = 10.98) and “level of education” (high: 83.08%; low: 0%; middle: 16.92%); freelancers’ personality trait “openness” (M = 5.49; SD = 1.13, Cronbach’s $\alpha = 0.65$) using an established seven-point Likert scale to capture the Big Five Inventory (BFI) (Hahn, Gottschling, & Spinath, 2012); freelancers’ “SMN usage intensity” (several times a month: 34.62%; less than once a month: 6.15%; once a month: 17.69%; once a week: 11.54%; several times a week: 14.62%; once a day: 10.77%; and several times a day: 4.62%), and the “number of weekly hours” working as a freelancer (M = 38.33; SD = 16.29) (Forret & Dougherty, 2001). Further, they surveyed the freelancers about the “field” they work in (e.g., Shevchuk & Strebkov, 2018) using a pre-chosen list of response categories (e.g., graphic design, with 15.38%, or journalists/PR, with 15.38%). Moreover, they asked whether the freelancers had “other earnings” apart from their freelance work (Van den Born & Van Witteloostuijn, 2013), which was not the case for the majority (66.92%). They also asked freelancers to assess their “strategic networking behavior” by adapting the four-item scale by Utz (2016). One of these items, for example, read: “I send contact requests to a great number of people, in order to get a large network”, to which participants could indicate their agreement on a five-point Likert scale ranging from “strongly disagree” (1) to “totally agree” (5). Weiler and Hinz (2019a) also evaluated the Cronbach’s alpha measure ($\alpha = 0.66$), which suggests consistency among the four items (M = 2.35; SD = 0.91).

3.3.3 *Challenges and Estimation*

The first challenge they encountered is sample attrition, i.e., that not every freelancer reacted to the follow-up questionnaire. Weiler and Hinz (2019a, p. 201) report that

“In order to alleviate this issue, we implemented different measures: (a) We used two different channels (i.e., e-mail and the messaging function of the platform itself) to invite and motivate our freelancers to participate in our follow-up questionnaire; (b) in total, we sent our participants three reminders. Freelancers who still refused to take part in the follow-up questionnaire after two reminders were offered a small delayed cash incentive (5 Euro), as a means of increasing the response rate [...]. While several freelancers responded to the third reminder, only five participants actually requested their promised cash incentive.”

Eventually, 65.12% (i.e., 140 of our 215 participants) took part in their follow-up questionnaire six months after the experiment. Specifically, 130 of these respondents reported that they are still working as a freelancer, while ten stated they quit pursuing freelance work and that they transitioned into traditional forms of employment. Thus, Weiler and Hinz (2019a) first focused on those who were still active as freelancers.

To rule out a violation of the random assignment assumption, they repeated the covariate balance check of Table 6 for the remaining 130 respondents. They found no statistically significant differences between the two groups ($p > 0.05$) (see Appendix A9). The results suggest that the observed attrition is unrelated to any pre-treatment covariate, so that the randomization remained valid.

A second challenge they encountered is two-sided noncompliance, which occurred because the freelancers did not always comply with the intended behavior: Treated freelancers did not redeem their vouchers and freelancers in the control group decided to buy premium membership on their own. At the same time, a few freelancers assigned to the control group decided to buy a premium membership on their own. Two-sided noncompliance is a common threat to experiments that utilize some sort of encouragement design (Gerber & Green, 2012). For instance, this issue often appears in clinical trials, where some patients are encouraged to take a medical treatment while others are not. Eventually, the decision about whether to redeem the randomly assigned voucher or not (i.e., whether to comply) may be driven by observed or unobserved individual characteristics, which is troublesome because it may result in the perilous problem of self-selection (Angrist & Pischke, 2008).

Thankfully, some approaches enable researchers to handle two-sided noncompliance, which “[...] allow the recovery of a causally interpretable estimate of the treatment effect, even though they alter the interpretation and the generalizability of the experimental results” (Lonati, Quiroga, Zehnder, & Antonakis, 2018, p. 26). These are the intention-to-treat (ITT) analysis, the as-treated (AT) analysis, the per protocol (PP) analysis, and the complier average causal effect (CACE, also referred to as local average treatment effect or LATE) (Imbens & Rubin, 2015; Sagarin et al., 2014). Among these approaches, CACE is the only one that recovers an unbiased causal estimate of treatment received on the outcome (Imbens & Rubin, 2015). As such, we focus on estimating the CACE in the upcoming analysis, as it captures the average treatment effect for a subset of participants—namely the so-called “*compliers*”—using an instrumental variable (IV) approach (see Appendix A8). IVs are an established and widely used approach for deriving causal estimates (Angrist & Pischke, 2008). We can define compliers as freelancers when a) they actually redeem a premium membership voucher after being assigned to receive it (treatment group), and b) they remain untreated after being assigned to the control group (i.e., they do not voluntarily buy the premium membership) (Gerber & Green, 2012; Imbens & Rubin, 2015).

As stressed by several scholars (e.g., Bollen, 2012), randomly generated variables inherently possess the ideal conditions to pass as a valid instrument. Thanks to their encouragement design, we already have such a promising IV that we can leverage: namely, the randomly allocated assignment to the premium voucher (Treatment assigned). Using this established setup, Weiler and Hinz (2019a) exactly follow recent articles from the IS domain, which also utilized the CACE approach with the random assignment of the treatment as an IV (C. Sun et al., 2019; Sun, Shi, Viswanathan, & Zheleva, 2019). Specifically, their choice of this IV gave them the perfect conditions to tackle the non-compliance issue, as they are subsequently able to uncover the causal effect for those freelancers who complied with their initial treatment assignment (Angrist & Pischke, 2008). As a result of utilizing this IV, the identified CACE is independent of unobserved confounding factors that give rise to self-selection. To put this into perspective: By only looking at the compliers, and thus remedying self-selected movement between treatment conditions, they ensure that the treatment and control groups remain comparable. Thus, the merit of employing the outlined approach is that the CACE estimate is “[...] undiluted by non-compliance and unaffected by selection bias” (Angrist, 2006, p. 35). Weiler and Hinz (2019a, p. 206) point to Bollen (2012) and Hinz, Hill, and Kim (2016) when recapitulating:

“In general, a valid instrument should meet two requirements: (1) It is highly correlated with the independent endogenous variable—in our case ‘actual redemption of premium membership voucher (Treatment received)’ (‘instrument relevance’)—and (2) it is uncorrelated with the unobservable error term ‘ u ’ (‘instrument exogeneity’) [...]. While it is possible to evaluate ‘instrument relevance’ with statistical means such as the first-stage F-statistic [...], scholars cannot consult statistical methods to test ‘instrument exogeneity’. Instead, they have to provide theoretical reasons and facts to substantiate why a specific variable fulfills this assumption. [...]”

They eventually summarize that given the randomized allocation of the treatment the IV derived from their experiment must meet the exogeneity requirement. We moreover repeat their assessment of the IV in the upcoming section, which shows that the IV meets all those criteria.

3.3.4 Results

3.3.4.1 Model-Free Evidence

In the following we review the findings of Weiler and Hinz (2019a) to answer RQ1—i.e., what is the impact of having access to the advanced networking features on SC gains. Looking at their plain results, without considering the IV, freelancers receiving the treatment accumulated, on average, 11.41% (SD = 16.72) more WorkSMN contacts during the experiment. In comparison, their counterparts in the control group gained an average of 9.41% (SD = 15.37) more WorkSMN contacts. In other words, the model-free analysis provides first empirical evidence that treated freelancers, with their access to advanced networking features, did not accumulate significantly more SC ($t(104.79) = -0.690, p = 0.492$).

3.3.4.2 CACE Estimation and IV

Moreover, Weiler and Hinz (2019a) include the covariates and derive a CACE estimate via the IV (Gerber & Green, 2012; Imbens & Rubin, 2015).⁵ We summarize their two-stage least squares (2SLS) IV model in the adapted Equations (1) and (2). Equation (1) reflects the first stage equation of the model in which the “Treatment received” (i.e., premium membership) is explained by the IV “Treatment assigned” (i.e., free premium voucher receipt) and the aforementioned covariates.

⁵ We display the correlation among these variables in Appendix A7

$$(1) \text{ TreatmentReceived}_i = \pi_0 + \pi_1 \cdot \text{TreatmentAssigned}_i + \pi_2 \cdot \text{Age}_i + \pi_3 \cdot \text{Gender}_i + \pi_4 \cdot \text{EducationLevel}_i + \pi_5 \cdot \text{SMNtenure}_i + \pi_6 \cdot \text{StrategicNetworking}_i + \pi_7 \cdot \text{Openness}_i + \pi_8 \cdot \text{UsageFrequency}_i + \pi_9 \cdot \text{AdditionalRevenueFreelancer}_i + \pi_{10} \cdot \text{WeeklyHours}_i + \pi_{11} \cdot \text{FreelanceArea}_i + \varepsilon_i$$

In Equation (2) we represent their second stage estimate, in which the endogenous self-selection variable “Treatment received” is replaced by the predicted values derived from the first-stage regression and the same covariates.

$$(2) \Delta \text{SocialCapitalAccumulation}_i = \beta_0 + \beta_1 \cdot \text{TreatmentReceived}_{(\text{predicted})}_i + \beta_2 \cdot \text{Age}_i + \beta_3 \cdot \text{Gender}_i + \beta_4 \cdot \text{EducationLevel}_i + \beta_5 \cdot \text{SMNtenure}_i + \beta_6 \cdot \text{StrategicNetworking}_i + \beta_7 \cdot \text{Openness}_i + \beta_8 \cdot \text{UsageFrequency}_i + \beta_9 \cdot \text{AdditionalRevenueFreelancer}_i + \beta_{10} \cdot \text{WeeklyHours}_i + \beta_{11} \cdot \text{FreelanceArea}_i + \varepsilon_i$$

Weiler and Hinz (2019a, p. 207) further theorize that individuals differ in their aspirations to participate in networking behaviors:

“We have reason to believe that there could be a mechanism at play that obfuscates the relationship between our treatment and dependent variable (“ Δ Social Capital Accumulation”). Specifically, it isn’t enough that the premium membership subscription offers more visibility on the work-related SMN as well as access to advanced and additional features; the freelancers also have to proactively use these new networking opportunities in order to gain SC benefits. In other words, they need to have the corresponding attitudes, values, and beliefs necessary to motivate their pursuit of instrumental ties.”

Hence, they follow the calls of various scholars in adding a variation to their model specification (Ahuja et al., 2012; Bensaou et al., 2014; Kuwabara et al., 2018; Porter & Woo, 2015). Weiler and Hinz (2019a) use “strategic networking behavior” (Utz, 2016) to reflect whether freelancers are truly motivated to befriend other users and consider their SC as a strategic instrument to support their career. In their regression model they enter an interaction term of both variables “treatment received” and “strategic networking behavior”. This interaction term reflects whether an individual who had a higher motivation or agency for networking, was able to benefit more from the advanced networking features. Yet, as the interaction term also contains the endogenous variable “treatment received” a further IV variable needs to be derived to identify the CACE for it (Ebbes et al., 2016). Weiler and Hinz (2019a) follow the examples of others (e.g., Hoisl & Mariani, 2017) and derive this second instrument from the interaction term between the first instrument and strategic networking behavior. We report a variant of the second-stage estimates (see footnote 4) of Weiler and

Hinz (2019a) in Table 7 and report their first-stage model in Appendix A10. Whereas model E1 shows the CACE estimate of the main effect “Treatment received”, model E2 also includes the interaction term “Treatment received * Strategic networking behavior”, which we focus on in the following.

Table 7. CACE estimation of the effect of having access to an IS with advanced networking features on SC accumulation—adapted from Weiler and Hinz (2019a, p. 209)

	<i>Model E1.</i> <i>2nd stage estimates:</i> <i>“Δ Social capital</i> <i>accumulation”</i>	<i>Model E2.</i> <i>2nd stage estimates: “Δ Social capital</i> <i>accumulation” (including interaction</i> <i>term)</i>
Treatment received	0.300 (2.472)	-10.867 (6.224)
Strategic networking behavior	3.306 (1.247) **	1.020 (1.698)
Treatment received * Strategic networking behavior		4.609 (2.338) *
Male (Ref.: female)	-4.104 (2.397)	-3.874 (2.380)
High level of education (Ref.: middle)	-4.723 (3.053)	-3.804 (3.063)
Age (in years)	-0.219 (0.112)	-0.233 (0.112) *
Openness	-1.150 (1.022)	-1.432 (1.023)
SMN tenure	-1.195 (0.287) ***	-1.243 (0.285) ***
Work hours per week	0.031 (0.067)	0.038 (0.066)
Other income (Ref.: no other income)	0.807 (2.242)	0.317 (2.237)
Usage frequency (once a month)(Ref.: less often than once a month)	-1.496 (4.771)	-2.427 (4.756)
Usage frequency (several times a month)	1.012(4.639)	1.044 (4.601)
Usage frequency (once a week)	0.591 (5.029)	0.550 (4.988)
Usage frequency (several times a week)	-1.082 (5.139)	-0.623 (5.102)
Usage frequency (once a day)	2.498 (5.185)	2.075 (5.146)
Usage frequency (several times a day)	9.438 (6.201)	9.035 (6.153)
F	3.452	3.435
R ²	0.351	0.362
RMSE	10.338	10.252
Underidentification test (Anderson canon. corr. LM statistic)	108.070 ***	104.394 ***
Weak identification test (Cragg-Donald Wald rk F-statistic)	473.076	193.656
N	130	130

*Note: 2SLS instrumental variable regression providing the complier average causal effect (CACE). Standard errors in parentheses; unstandardized regression estimates. Constant not reported. The paper of Stock and Yogo (2005) lists the critical values for the Cragg-Donald Wald F statistic. The Stock-Yogo weak ID test critical value for a 10% maximal IV size is 16.38 for model E1 and 7.03 for model E2, respectively. In our model specification, we partialled out the variable freelance field from all the other variables in our estimation using the Frisch-Waugh-Lovell theorem. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$*

As validation metrics in Table 7, we also included the “weak identification test”, showing that the receipt of the voucher had a sufficiently strong effect on the decision to become a premium member: The Cragg-Donald Wald F-statistic measure (193.656), exceeds the critical thresholds proposed in Stock and Yogo (2005). Also the under-identification test derived via the Anderson canonical correlation LM statistic rejects the null hypothesis that the model is

overall under-identified, at $p < 0.001$. In sum, we are confident that the IVs used in Weiler and Hinz (2019a) are sufficiently strong and valid.

3.3.4.3 Interaction between Strategic Networking Behavior and Treatment Received

As described in Weiler and Hinz (2019a, p. 210),

“The result of the second-stage regression of Model [E2] (Column 2) provides a first indication that the treatment (receiving access to advanced and additional networking features via the premium membership) has no direct significant causal effect on establishing more social capital. Instead, we found a significant positive interaction effect between our treatment and strategic networking behavior [$B = 4.609$; $p < 0.05$]. [...] This finding implies that complying freelancers who score higher on the strategic networking behavior scale could significantly increase their amount of social capital. To put it another way, our finding indicates that if a freelancer exhibits a strong motivation for strategic networking behavior and has access to an IS with suitable networking features, he/she will substantially increase the amount of his/her social capital relative to that of his/her counterparts in the control group. Specifically, scoring higher on the strategic networking behavior scale by one point will increase the amount of social capital of a freelancer by [4.609%] relative to that of his/her non-treated equivalents.”

In Table 8, adapted from Weiler and Hinz (2019a), we show a two-by-two matrix to highlight the magnitude of the interaction effect. Weiler and Hinz (2019a, pp. 210-211) reflect on this:

“We included those freelancers with a value smaller than the median value (median = 2.25) of our variable strategic networking behavior into the “low” group and classified those freelancers with a score equal or greater than the median into the “high” group. We find that, on average, treated freelancers in the latter group accumulated 13.24% more social capital during the course of the experiment, while their counterparts in the former group achieved, on average, only 8.93% more social capital. Likewise, we see that freelancers accumulate more social capital on average if they have access to an IS that equips them with suitable networking features.”

Table 8. Descriptive mean comparisons of strategic networking behavior (median split) and treatment received groupings—adapted from Weiler and Hinz (2019a, p. 211)

		<i>Treatment received</i>	
		No (0)	Yes (1)
<i>Strategic networking behavior</i>	Low	8.33%	8.93%
	High	10.39%	13.24%

The interplay of the treatment and strategic networking behavior implies that the advanced networking features that freelancers’ need to utilize actively are of particular importance for gaining SC. In Appendix A11, we test whether the agency variable “strategic networking behavior” potentially suffers from post-treatment bias because it was elicited in a follow-up questionnaire. We provide theoretical arguments why this is not the case and empirically rule out group differences between the treatment and control group. Ultimately, we conclude that it is quite unlikely that the motivation to engage in strategic networking behavior was susceptible to change due to our issued treatment.

In summary, the results of the field experiment suggest a causal link between having access to advanced networking features and the accumulation of SC, but this is conditional on the users’ agency for networking. In the next section, we explore which types of features drive this accumulation.

3.4 Feature Utilization: Supportive Evidence (RQ2)

The field experiment suggests that the mere increase in status portrayed by a premium membership badge, or an increase in visibility, is not sufficient to attract SC in a work-related SMN by itself. Otherwise, we would have observed a significant main effect for all types of treated users. Instead, we observed that users with a high agency for networking were able to significantly increase their SC. This begs the question: How did these users achieve this positive outcome? That is, we want to investigate which types of advanced networking features drive SC accumulation (RQ2). This question is of paramount interest to platform providers, who seek to understand the products that they are offering and the magnitude of the value they add. Likewise, platform users want to know exactly what a premium membership entails.

To investigate RQ2, we follow the needs-affordances-features framework, which states that affordances and features are closely related: features enable affordances, i.e., “action possibilities” (Karahanna et al., 2018). Hence, we shift our focus to the actual change in how

users make use of these affordances following a premium conversion, as they are directly linked to the advanced networking features.

3.4.1 Framework

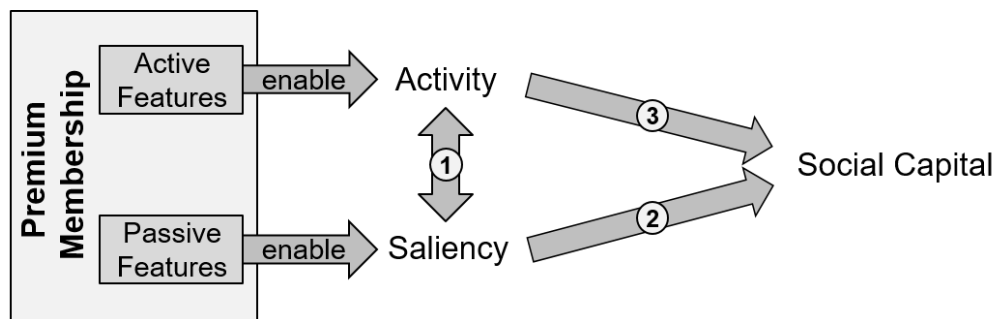
We distinguish two types of advanced networking features: Active features operate via *activity*, i.e., targeted actions of the focal user, whereas passive features operate via *saliency*, i.e., incoming attention that the focal user receives. In Appendix A5, we list these features in detail. The arguably most important active features are the following: Premium users have access to an exclusive advanced search filter. While basic users can only search by name, premium users can search through WorkSMN’s user directory by industry, firm, and role, or even a combination of criteria. For example, by these means, premium users can identify profiles of targets in the buying center of a specific company—without knowing their names—and get in contact with them. This advanced search filter is especially important to freelancers because they rely heavily on acquiring project-based labor arrangements (Wu, 2013). Besides, only premium users can send personal messages to such non-contacts and thereby reach out to them. One can think of profile visiting and messaging as steps on a funnel of SC accumulation: users first identify potential targets and then actively “make a move” (R. Bapna et al., 2016) on targets outside of their network.

Passive features take effect via the premium user’s saliency. They include better and more highlighted positioning in search results and a prestigious premium badge that accompanies the user’s profile. Among others, this badge also highlights the profile in contact suggestions, in groups, and contact lists. As passive features unfold their effectiveness irrespective of the user’s activity (e.g., a badge is always highlighting premium users), all premium users are likely to experience a lasting increase in SC accumulation due to these passive features. Moreover, premium users can access user statistics on who visited their profile. The ability to follow up on such “weak signals” created by profile visits has been noted as a pivotal feature in achieving matches in dating SMN (R. Bapna et al., 2016). Likewise, we assume that this feature may improve SC accumulation on work-related platforms: For example, incoming profile visitors may give freelancers hints about interested clients or colleagues.

Given that advanced networking features primarily support targeted actions, we operationalize profile visits and messages sent as *activity*, and profile visitors and messages received as *saliency*. As illustrated in Figure 5, activity relates to active features while saliency relates to passive features. The seeding literature on SMNs suggests that saliency is a

result of activity (X. Chen, van der Lans, & Phan, 2017; Lanz, Goldenberg, Shapira, & Stahl, 2019); however, it could also be the other way around, i.e., that a user’s saliency triggers activity. Correspondingly, we expect a reciprocate *association between activity and saliency* (see arrow 1). The seeding literature also finds that activity is effective in accumulating SC (Hinz et al., 2011; Lanz et al., 2019). Hence, for the *association between activity and SC accumulation* (arrow 2), we expect that profile visiting and personal messaging drive SC accumulation. Finally, regarding the *association between saliency and SC accumulation* (see arrow 3), research shows that prestigious items, like a premium badge, can result in more contacts on SMNs (Hinz et al., 2015).

Figure 5. Hypothesized framework



3.4.2 Data

WorkSMN kindly shared data with us on three discount mailing campaigns conducted in May 2015, May 2016, and November 2016. During these periods, a selected set of 52,392 freelancers received a 50% discount coupon on their first 12-month subscription. Of them, 626 freelancers converted to premium. Even though it is impossible to warrant unbiased causal estimates from such observational data, it allows us to further investigate the process of SC accumulation under premium membership, as well as account for the dynamic interplay of activity and saliency with SC. The restricted discount mailing campaign allows us to carry out our analysis in a uniform environment; in most cases involving SMNs, the discounts and their communication tend to be highly personalized and vary over time. This approach also boosted the number of conversions, which might otherwise be too low due to the rare-event nature of premium membership conversions (R. Bapna et al., 2018; Oestreicher-Singer & Zalmanson, 2013).

The dataset contains digital footprint data about the users, which we summarize in Table 9. Tracking user activity on the platform is multifaceted; thus, WorkSMN only stores a selected set of variables that it considers the most relevant over a longer period. These include

the monthly individual-level panel data on users' number of profile visits and messages sent, as well as their number of contacts four months before and six months after the mailing campaigns. While some scholars have studied similar activity information (e.g., R. Bapna et al., 2018; Oestreicher-Singer & Zalmanson, 2013), only a few have succeeded in also capturing the incoming activities (for an exception in the context of dating platforms, see R. Bapna et al., 2016). In our case, we have access to the monthly numbers of profile visits and messages received from all focal users. In addition to the individual-level panel data, WorkSMN granted us access to a granular set of snapshot data, which they collected in advance of each mailing campaign to select the targeted users. This snapshot data contains the number of contact invites sent before the mailing campaign—along with variables that largely overlap with the ones used in the field experiment. To overcome the high skewness of the digital footprint variables, and to implement the notion that excessive networking behavior is rewarded with diminishing returns to scale, we took natural logarithms of “+1”-transformed values for our independent variables.

As reflected in Table 9, non-converters appear similar to converters with regard to SMN tenure, age, gender, education, and career level. Nevertheless, converting freelancers sent out more contact invites, had more visits, higher seniority, and were more job-seeking. We found strong behavioral changes (Δ) for premium converters across all captured dimensions of SC, activity, and saliency. Also, a correlation analysis (see Appendix A13) revealed a general association between activity and saliency measures.

Table 9. Covariate summary

Covariate	Premium Conversion	Mean (SD)	Difference	t-value	p-value	Measure
SC Accumulation (Δ)	0 (n = 51,766) 1 (n = 626)	11.719 (25.205) 30.254 (66.854)	-18.535	6.933	<0.001	Monthly
Profile Visits (Δ Log)	0 1	0.003 (0.998) 0.809 (1.242)	-0.806	16.212	<0.001	Monthly
Messages Sent (Δ Log)	0 1	0.038 (0.659) 0.501 (0.854)	-0.463	13.529	<0.001	Monthly
Profile Visitors (Δ Log)	0 1	-0.079 (0.751) 0.285 (0.830)	-0.364	10.925	<0.001	Monthly
Messages Received (Δ Log)	0 1	0.062 (0.700) 0.387 (0.826)	-0.325	9.808	<0.001	Monthly
Contact Invites Sent Before (Log)	0 1	2.842 (1.471) 3.173 (1.427)	-0.331	5.766	<0.001	Before campaign
Website Visits Before (Log)	0 1	1.746 (1.201) 2.075 (1.223)	-0.329	6.679	<0.001	Before campaign
SMN Tenure (in years)	0 1	8.052 (2.439) 7.939 (2.686)	0.113	1.051	0.294	Before campaign
Age	0 1	43.040 (9.684) 42.376 (9.726)	0.664	1.697	0.090	Before campaign
Gender (Female = 1; Male = 2)	0 1	74.971% 75.680%	-0.709%	0.130	0.719	Before campaign
Education Title (Yes = 1, No = 0)	0 1	50.994% 53.920%	-2.926%	2.000	0.157	Before campaign
Level: Entry & Experienced	0 1	58.024% 55.040%	2.984%			
Level: Management	0 1	13.335% 13.280%	0.055%	2.938	0.230	Before campaign
Level: Senior	0 1	28.641% 31.680%	-3.039%			
Job seeking: Active	0 1	36.840% 48.320%	-11.480%	34.436	<0.001	Before campaign

Note: Test statistic reflect Welch's t-test for continuous variables and Pearson's Chi-squared test for discrete variables

3.4.3 Challenges and Estimation

One challenge of observational data is that self-selection is inherent. Unlike in the field experiment, in our observational data freelancers had to pay for the premium membership by themselves. Thus, if a variable is associated with the self-selection effect and the dependent variable at the same time, it will skew the estimated effect size (i.e., a confounder). We acknowledge that there are countless hypothetical confounding factors. In our setting, the latent job situation of a freelancer is the most obvious candidate. A freelancer who lost an important client is more likely to convert to premium and increase networking activity. We include variables to capture this confounder in our analyses. Later on, we provide an extensive set of robustness tests using propensity score matching methods, enabled by the rich dataset.

Hence, even though we cannot categorically rule out all potential confounding factors, the observational data provides us with a valuable complementary perspective: It allows us to assess the previous findings' convergent validity while also resolving some of the previously mentioned limitations of the experiment. The overall idea is that if the results of the two approaches converge, we can gain more confidence in the findings. One example is that it was not possible to rule out the post-treatment bias in the experiment completely. With the individual-level panel data, we can be sure it does not exist, because we utilize in our analysis the number of contact requests sent before the mailing campaign.

Another challenge is the assumed directionality of activity and saliency on the SC accumulation (i.e., arrows 2 and 3 in Figure 5). Studies that seek to quantify content contribution provide evidence for an audience effect in some contexts (e.g., blogging networks in Shriver, Nair, & Hofstetter, 2013): A bigger audience can incentivize activity, which is why researchers commonly refer to this reverse association as an audience effect. In contrast, experimental findings suggest that this effect is rather nuanced (Toubia & Stephen, 2013). Because our study primarily focuses on directed networking actions, we expect this effect to be insubstantial in our context. Nevertheless, we also assessed whether reverse associations, such as an audience effect, occur in our setting.

We continue as follows: First, we replicate the analysis of Weiler and Hinz (2019a) on RQ1 with the observational data. Thereby, we acknowledge that we cannot warrant unbiased causal estimates due to the possibility that not all determinants of the self-selection are necessarily controlled for. Nonetheless, we can now objectively measure strategic networking behavior by operationalizing the number of contact invites sent before the campaign. Second, we turn our attention to the outgoing and incoming activities (i.e., activity and saliency) after the mailing campaigns. We use model-free evidence to investigate which types of advanced networking features drive SC accumulation. Third, we seek empirical support on the directionality of activity and saliency on SC accumulation. For this, we follow the examples of Dewan and Ramaprasad (2014) and H. Chen, De, and Hu (2015) and used a panel vector autoregression model. Finally, we conduct a difference-in-differences analysis to assess the influence of behavioral changes over time.

3.4.4 Results

3.4.4.1 Accumulating Social Capital: Robustness Check (RQ1)

Consistent with the experiment, we examine SC before the discount mailing and six months after. Since the campaigns target a large and heterogeneous user base of freelancers, the number of contacts ($M = 210.12$, $SD = 336.87$) is sometimes low, even up to no contacts at all. For this reason, we measured changes in SC accumulation as the *absolute* number of direct contacts.⁶ To objectively measure strategic networking behavior, we use the individual number of contact invites sent before the mailing campaign ($M = 47.56$, $SD = 194.69$). Replacing the self-reported agency variable “strategic networking behavior” of the 2SLS-IV regression presented in Equation 2 with “contact invites sent before”, we derive the OLS regression estimates presented in Table 10. To provide conservative estimates, we report heteroscedasticity-robust standard errors.

In line with our experimental findings, we observe a positive association between pre-campaign behavior and SC accumulation. Freelancers who exhibit a 1% higher number of contact invites sent before converting increased their SC accumulation by 0.059 contacts ($= \log(1.01) \cdot 5.985$; $p < 0.01$) after their premium conversion. In other words, as a premium user, every doubling of the number of contact invites sent results in an increase in SC accumulation of approximately 4.148 ($= \log(2) \cdot 5.985$) contacts. Table 10 reveals a significant main effect of increased SC accumulation ($p < 0.001$) for “premium conversion”, which was not significant in the field experiment. We assume that this difference is due to the non-experimental setting where freelancers pay for premium membership (R. Bapna et al., 2018).

Upon the inclusion of the interaction term, the main effect of the conversion is no longer significant, which suggests that freelancers with an agency for networking are accumulating SC when they have access to advanced networking features. These results provide strong support for the field experiment’s main conclusion.

⁶ We show in Appendix A12 that our findings are robust for the percentage change, when following R. Bapna and Umyarov (2015) and removing the freelancers with the lowest 15-percentile in SC prior to the mailings.

Table 10. OLS regression of the effect of having access to an IS with advanced networking features on SC accumulation

	<i>Model P1: „ASC accumulation”</i>	<i>Model P2: „ASC accumulation” (including interaction with contact invites sent before)</i>
Premium conversion	16.027 (2.604) ***	-2.918 (4.489)
Contact invites sent before (Log)	4.351 (0.162) ***	4.284 (0.163) ***
Premium conversion *		5.985 (1.993) **
Contact invites sent before (Log)		
Male (Ref.: female)	1.009 (0.219) ***	0.991 (0.220) ***
Level of education (Ref.: No acad. title)	1.006 (0.224) ***	1.010 (0.224) ***
Age	-0.068 (0.015) ***	-0.068 (0.015) ***
SMN tenure	-0.134 (0.055) *	-0.134 (0.055) *
Career segment: Manager (Ref.: Entry)	1.795 (0.351) ***	1.809 (0.351) ***
Career segment: Senior Executive	0.231 (0.282)	0.222 (0.280)
Job seeking: active	0.820 (0.238) ***	0.808 (0.237) ***
Usage frequency as Logins before (Log)	2.683 (0.142) ***	2.684 (0.142) ***
F	560.364	523.712
R ²	0.114	0.115
RMSE	24.643	24.626
N	52,392	52,392

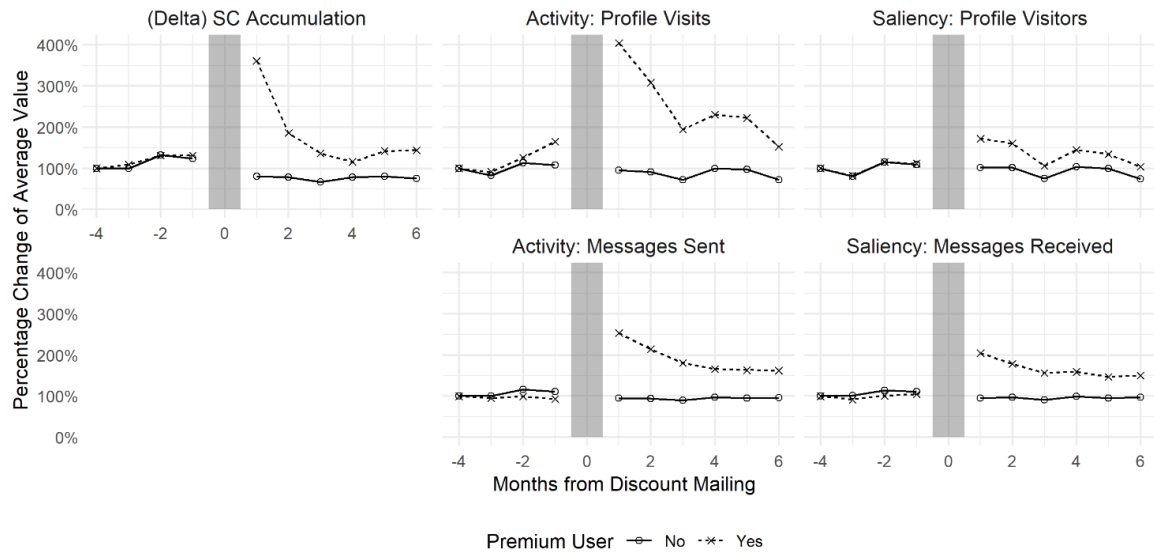
*Note: Constant and fixed effects of the three mailing campaigns are not reported. Robust heteroscedasticity-consistent standard errors in parentheses. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$*

3.4.4.2 Model-Free Evidence

We now visually explore which specific types of features (active or passive) drive these gains (RQ2). Figure 6 visualizes the percentage changes in SC of those who became premium users vs. non-premium users, and contrasts them in terms of the other digital footprint variables. For ease of interpretation, we index all variables to four months before the mailing campaign. We grey out the month of the mailing campaign because the monthly data does not allow us to decompose the numbers into the ones before and after upgrading to premium.

The indexed SC of premium subscribers is similar to the non-subscribers before the mailing campaign. After the campaign, we observe a peak for premium users that declines quickly yet stays on a significantly higher index level ($p < 0.001$) compared to non-premium users. Concerning activity (i.e., profile visits and messages sent), its trajectory coincides with SC accumulation in magnitude and duration. The peak is particularly pronounced for profile visits (404% versus 253% for messages sent). A possible reason is that, in contrast to message sending, profile visiting is a lower-involvement activity. When looking at the graphs on the right-hand side, saliency does not peak at the same magnitude but followed a similar pattern (172% for profile visitors and 205% for messages received).

Figure 6. Monthly co-evolution of average SC accumulation, activity, and saliency before and after the premium conversion



The peak in the first month of premium conversion and the quick decline afterward suggest that SC accumulation has a stronger relationship with activity than saliency. R. Bapna et al. (2018) observe a similar activity peak in the context of music freemium services. They attribute this change to users’ desire to extract value from the service that they now pay for. Regarding saliency as an explanation of SC gains, we would expect a more durable increase in SC after the premium conversion; after all, the features that help a user stand out do not change in the course of the premium membership. Yet, because saliency also peaks in the first month, the graphs suggest that activity may be the underlying driver of the saliency increase. Next, we turn our attention to a difference-in-differences model that allows us to control for user characteristics. However, as discussed above, before we could do this, we need to first to rule out a reverse association between activity and SC accumulation as well as saliency and SC accumulation.

3.4.4.3 Panel Vector Autoregression

We follow the examples of studies in IS (H. Chen et al., 2015; Dewan & Ramaprasad, 2014) and use panel vector autoregression (PVAR) to model the dynamic interdependencies between our main variables via Generalized Method of Moments (GMM). GMM uses transformed observations as instruments for the lagged dependent variables (Hansen, 1982). As detailed in Appendix A14, we first downsample non-converters due to the computational intensity of PVAR calculations via GMM. We test the log-transformed individual values for stationarity (Harris & Tzavalis, 1999) and found an optimal lag length of three months

(Andrews & Lu, 2001). The resulting estimates of a system PVAR model, which controls for time fixed effects, are presented in Table 11 below.

Table 11. Panel vector autoregression model with three lags

	<i>Profile visits</i>	<i>Profile visitors</i>	<i>Messages sent</i>	<i>Messages rec.</i>	Δ <i>SC acc.</i>
<i>Profile visits</i> _{<i>t-1</i>}	0.209 (0.015)***	0.059 (0.009)***	0.041 (0.011)***	0.046 (0.010)***	0.048 (0.012)***
<i>Profile visitors</i> _{<i>t-1</i>}	0.126 (0.019)***	0.207 (0.013)***	0.057 (0.013)***	0.052 (0.014)***	0.089 (0.016)***
<i>Messages sent</i> _{<i>t-1</i>}	-0.015 (0.02)	0.015 (0.013)	0.117 (0.016)***	0.039 (0.016)*	-0.005 (0.019)
<i>Messages rec.</i> _{<i>t-1</i>}	0.021 (0.019)	-0.019 (0.012)	0.045 (0.015)**	0.111 (0.014)***	0.016 (0.018)
Δ <i>SC acc.</i> _{<i>t-1</i>}	0.020 (0.015)	0.000 (0.010)	0.009 (0.011)	0.016 (0.011)	0.029 (0.015)
<i>Profile visits</i> _{<i>t-2</i>}	0.043 (0.014)**	-0.01 (0.008)	0.003 (0.009)	0.008 (0.009)	-0.023 (0.012)*
<i>Profile visitors</i> _{<i>t-2</i>}	0.073 (0.018)***	0.091 (0.013)***	0.016 (0.013)	0.016 (0.013)	0.063 (0.016)***
<i>Messages sent</i> _{<i>t-2</i>}	-0.021 (0.019)*	0.011 (0.012)	0.037 (0.015)*	-0.022 (0.014)	-0.018 (0.018)
<i>Messages rec.</i> _{<i>t-2</i>}	-0.024 (0.018)***	-0.019 (0.011)	-0.004 (0.014)	0.052 (0.013)***	-0.015 (0.017)
Δ <i>SC acc.</i> _{<i>t-2</i>}	0.023 (0.013)	0.013 (0.009)	0.013 (0.010)	0.010 (0.010)	0.035 (0.015)*
<i>Profile visits</i> _{<i>t-3</i>}	0.029 (0.012)*	0.011 (0.008)	0.000 (0.008)	0.011 (0.009)	-0.009 (0.01)
<i>Profile visitors</i> _{<i>t-3</i>}	0.080 (0.016)***	0.135 (0.011)***	0.012 (0.011)	0.000 (0.011)	0.012 (0.013)
<i>Messages sent</i> _{<i>t-3</i>}	-0.022 (0.018)	-0.015 (0.011)	0.016 (0.014)	-0.025 (0.014)	-0.017 (0.016)
<i>Messages rec.</i> _{<i>t-3</i>}	-0.022 (0.017)*	-0.019 (0.011)	0.001 (0.013)	0.030 (0.013)*	-0.005 (0.017)
Δ <i>SC acc.</i> _{<i>t-3</i>}	-0.001 (0.012)	-0.014 (0.008)	-0.005 (0.010)	0.008 (0.009)	0.022 (0.013)
<i>Premium con.</i> _{<i>t</i>}	0.476 (0.034)***	0.229 (0.023)***	0.287 (0.026)***	0.189 (0.026)***	0.264 (0.027)***

Note: Constant and seasonality estimates not reported; Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Both saliency and activity are positively related with SC accumulation. In Table 11, all lagged observations of activity and saliency are associated with each other and SC accumulation. Converting to premium is positively associated with all of the digital footprint variables (all $p < 0.001$). However, the reverse association—namely that a higher number of contacts might lead to an increase in saliency and a higher activity level (e.g., Shriver, Nair and Hofstetter 2013)—is not supported. SC accumulation had no explanatory value for consecutive activity nor saliency; thus, we can rule out the presence of a reverse effect. Moreover, the results show that activity and saliency are interrelated, but they do not allow us to conclude that one precedes the other.

3.4.4.4 Active Versus Passive Features

To explore which specific features drive SC accumulation, we further develop the aggregate difference-in-differences based modeling approach from the robustness check of RQ1 (Table 10). We extend the control variables introduced in model P1 as follows: For each

freelancer, we model the change in the activity variables (i.e., profile visits and messages sent) and saliency variables (i.e., profile visitors and messages received) by differencing the log-transformed value before with the average six months after the campaign. Moreover, we add interaction terms between “premium conversion” and the increases in activity and saliency, resulting in equation (3):

$$(3) \Delta SocialCapitalAccumulation_i = \beta_0 + \beta_1 \cdot PremiumConversion_i + \sum \beta_C \cdot Controls_i + \sum \beta_A \cdot \Delta Activity_i + \sum \beta_S \cdot \Delta Saliency_i + \sum \beta_{IA} \cdot PremiumConversion_i * \Delta Activity_i + \sum \beta_{IS} \cdot PremiumConversion_i \cdot \Delta Saliency_i + \varepsilon_i$$

Next to the direct effect estimates of activity β_A and saliency β_S , the interaction estimates β_{IA} (interaction activity) and β_{IS} (interaction saliency) reflect the ability of active and passive advanced networking features to accumulate SC. If access to advanced networking features plays no role in SC accumulation, we would expect these interaction terms to turn out insignificant.

When testing this model for concerns of multicollinearity, we find that all variance inflation factors are in a moderate range (smaller 3), suggesting that we can distinguish activity from saliency. Because the PVAR results do not indicate whether saliency follows activity or vice versa, we perform the subsequent analysis in a stepwise manner: We report the results with either activity or saliency alone, and then we include their interaction term with premium conversion.

As shown in Table 12, model P3 supports that changes in activity—i.e., profile visits ($B = 3.777$; $p < 0.001$) and messages sent ($B = 3.830$; $p < 0.001$)—are associated with SC accumulation. In model P4, we find similar effects for the saliency variables—i.e., profile visitors ($B = 3.706$; $p < 0.001$) and messages received ($B = 3.898$; $p < 0.001$)—albeit with lower R^2 .

When including the interaction terms of activity, the significant main effect of the premium conversion disappears in model P5 ($B = -0.003$; $p > 0.05$). At the same time, both interaction terms—namely between premium conversion and profile visits ($B = 5.669$; $p < 0.05$) as well as messages sent ($B = 13.500$; $p < 0.05$)—prove to be positive and significant. On the other hand, the interaction terms of the saliency variables (profile visitors and messages received) with premium conversion in model P6 and P7 yield no significant parameters ($p > 0.05$) and only partially explain the premium conversion main effect in model P6 ($B = 6.671$; $p > 0.01$).

Table 12. OLS regression of the effect of changes in activity, saliency, and their interactions with premium conversion on SC accumulation

	Δ SC accumulation explained by...				
	Model P3: Δ Activity	Model P4: Δ Saliency	Model P5: Δ Activity and interactions	Model P6: Δ Saliency and interactions	Model P7: Δ Activity, Δ Saliency, and interactions
Premium Conversion	11.133 (2.455)***	13.496 (2.519)***	-0.003 (2.738)	6.671 (2.212)**	-0.074 (2.767)
Δ Profile visits (Log)	3.777 (0.168)***		3.666 (0.160)***		3.255 (0.150)***
Δ Messages sent (Log)	3.830 (0.277)***		3.553(0.257)***		2.108 (0.362)***
Δ Profile visitors (Log)		3.706 (0.241)***		3.584 (0.211)***	2.161 (0.186)***
Δ Messages received (Log)		3.898 (0.322)***		3.693 (0.296)***	1.458 (0.391)***
Premium conversion *			5.669 (2.740)*		4.239 (2.060)*
Δ Profile visits (Log)					
Premium conversion *			13.500 (5.412)*		12.695 (5.886)*
Δ Message sent (Log)					
Premium conversion *				8.317 (6.925)	3.814 (6.012)
Δ Profile visitors (Log)					
Premium conversion *				11.817 (8.040)	0.956 (9.058)
Δ Messages received (Log)					
F	674.598***	606.619***	610.314***	545.072***	503.201***
R ²	0.153	0.140	0.157	0.143	0.161
GVIF < 3	Yes	Yes	Yes	Yes	Yes
RMSE	24.095	24.283	24.034	24.238	23.977
N			52,392		

*Note: We include all control variables as in model P1 (Table 10). Robust heteroscedasticity-consistent standard errors in parentheses. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$*

Overall, increases in activity and saliency are associated with increases in SC. That means that irrespective of membership type, users can enhance their SC by changing their activity. Moreover, looking at the estimated interaction coefficients suggests that, in the setting of work-related SMN, activity does not uniformly drive SC for basic and premium users. Instead, we observe a significant interaction between “premium conversion” and changes in those activities that are supported by active features. The interaction effects of saliency with premium conversion are not significant, indicating that passive features are less effective than active features for premium users.

These patterns complement our conclusions from the field experiment of Weiler and Hinz (2019a), which revealed that only freelancers who convert to premium and exhibit a strategic networking behavior accumulate more SC. The individual-level panel data further suggests that among the advanced networking features, particularly active features support the accumulation of SC, thereby enabling users to actively extend their network. The passive features, such as prestigious premium badges and the ability to follow up on profile visitors, provide limited value in terms of accumulation.

3.4.4.5 Propensity Score Matching

A limitation of the previous analysis is that self-selection among premium converters may exist. Table 9 shows that premium converters and non-converters differ substantially, so that they may not be directly comparable. We follow previous IS research (e.g., R. Bapna et al., 2018; Hinz et al., 2015; Oestreicher-Singer & Zalmanson, 2013) and address this potential limitation by conducting propensity score matching (PSM). We can test for the robustness of the results by selecting only those non-converting users who closely resemble premium converters.

The objective of this approach can be thought of as “finding an artificial twin” that closely resembles a premium converter (Rosenbaum, 2020). We first have to form propensity scores based on pre-treatment variables. Following model P1, we include all basic variables such as age, gender, education, and self-indicated job-seeking activity. However, the strength of our observational data is that it contains a wide range of variables beyond these basic ones. As a large scale comparison of experimentation and matching shows, granular behavioral matching can yield results indistinguishable from experimental results (Eckles & Bakshy, 2020). We build upon this idea and try to account for a potential self-selection process: In addition to the basic variables of model P1, we further include pre-treatment behavioral variables, namely in terms of activity (i.e., profile visits and messages sent) and saliency (i.e., profile visits received and messages received).

Behavioral variables are especially promising because they can uncover latent processes, such as the users’ current job situation (Ebbes & Netzer, 2018). For example, we find in Figure 6 that profile visiting already increases before a premium conversion, which supports the assumption. In line with this phenomenon, profile visits before treatment have a significant ($p < 0.01$) positive effect on the decision to convert to premium, as estimated via the underlying propensity model (reported in Appendix A15).

The propensity scores are the basis for matching, where the most common form is 1:1 matching. Since 626 freelancers converted to premium, the corresponding “artificial twins” also amount to 626 among the abundant control observations (i.e., 1.209% from $n = 51,766$). To reduce sampling variability, one can further apply an elbow criterion (Rosenbaum, 2020), which allows us to adapt the matching ratio to 1:4 and increase the selection to 1,841 most identical observations (i.e., 3.556% from $n = 51,766$). Based on these two matching criteria

and the two sets of variables, Table 13 shows the modeling results of four PSM specifications. For direct comparison, we also include the unmatched sample (model P7 from Table 12).

Table 13. Propensity score matching overview

<i>Matching variables</i>	<i>Propensity Score Matching</i>				<i>Unmatched (comparison)</i>
	Controls	Controls, Activity before, Saliency before	Controls	Controls, Activity before, Saliency before	(unmatched)
<i>Matching ratio</i>	1:1	1:1	1:4	1:4	-
<i>Matching caliper</i>	-	-	0.002	0.002	-
<i>Treatment obs. (dropped)</i>	626 (0)	626 (0)	623 (3)	617 (9)	626
<i>(Matched) controls</i>	(626)	(626)	(1,841)	(1,841)	51,766
<i>Interaction Coefficients of Model P7</i>					
Premium conversion *	5.468	3.286	4.053	4.292	4.239
Δ Profile visits (Log)	(2.632)*	(2.424)	(2.343) ⁺	(2.321) ⁺	(2.060)*
Premium conversion *	13.073	12.822	14.886	13.345	12.695
Δ Message sent (Log)	(5.994)*	(5.732)*	(6.180)*	(5.822)*	(5.886)*
Premium conversion *	1.189	-0.982	3.541	3.792	3.814
Δ Profile visitors (Log)	(6.097)	(9.150)	(5.951)	(5.747)	(6.012)
Premium conversion *	2.286	2.429	0.752	0.292	0.956
Δ Messages received (Log)	(9.221)	(5.732)	(9.368)	(9.111)	(9.058)

Notes: Robust heteroscedasticity-consistent standard errors in parentheses. Significance levels: ⁺ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Matching caliper: to be interpreted in standard deviations of treatment propensities, where 0.002 reflects an inflection point to dropping treatment observations.

Table 13 shows that for all four PSM specifications, the interaction of message sending and premium conversion remains positive and significant. As with the unmatched sample, no significant interaction with saliency occurs. The interaction of profile visiting and premium conversion is significant for the most common form of PSM—and at a higher level ($p < 0.1$) when increasing the ratio to 1:4 (potentially caused by the smaller dataset). These results further underpin that the advanced networking features, especially messaging, are valuable for SC accumulation.

3.5 Discussion

Our review of the field experiment conducted in Weiler and Hinz (2019a) and the additional findings of an observational dataset, allow us to advance the discussion on the actions that lead to accumulation of SC. As various scholars point out, SC is not an exogenously given trait (Ahuja et al., 2012; Bensaou et al., 2014; Borgatti & Halgin, 2011; Stuart & Sorenson, 2007), with which our empirical findings agree. Our research particularly explains how advanced networking features help freelancers to accumulate more SC, and uncovers the role of their motivation to network in this process.

The field experiment of Weiler and Hinz (2019a) provides answers to RQ1, by giving an estimate of the CACE. The CACE estimates reflects an unbiased causal estimate for the sub-

group of compliers (Gerber & Green, 2012; Imbens & Rubin, 2015), despite the two-sided noncompliance issue that occurred in the experiment. For those compliers, their findings show that the motivation to network is pivotal for the efficacy of the premium membership: Only users who were motivated beforehand, also increased their SC.

In this study, we tested for the robustness of these findings: We find converging results of a significant interaction effect between premium conversion and SC accumulation for the objectively measured number of contact invites sent utilizing a second, individual-level panel dataset. We further utilized the observational dataset for our second study (RQ2) and found that observed changes in user behavior concerning profile visiting and message sending is positively linked with SC accumulation. Along these lines, a freelancer who is joining a work-related SMN cannot rely on passive features alone to accumulate SC. Prestigious premium badges, as they are very frequently used in freemium settings (e.g., R. Bapna et al., 2018), may act as an encouragement to other users to also obtain premium and also impact saliency. Yet, for SC accumulation, these badges play a minor role.

3.5.1 Theoretical Implications

First, scholars have hypothesized that individual features of an SMN define the way individuals network and build SC online (Kane et al., 2014; Karahanna et al., 2018; Sundararajan et al., 2013). Building on these theories, we contribute a theoretical framework that describes how advanced networking features drive SC accumulation by enabling two pivotal forces: Activity and saliency. Using this framework, we further investigated those mechanisms among self-selected individuals who pay for the premium membership and thus generally exhibit a high baseline agency for networking. As our field experiment demonstrates in work-related SMN, a prestigious appearance in the form of a premium badge is, by itself, insufficient to accumulate significantly more SC. Our individual-level panel data further uncovered the mechanisms of increased SC accumulation among premium users: They increase their SC by sending more messages and outgoing profile visits. The ability to follow up on “weak signals” of visible profile visits, a crucial mechanism in dating SMNs (R. Bapna et al., 2016), appears to have no effect in work-related SMNs: increasing profile visits are not significantly associated with more SC accumulation if the focal user has access to a premium membership. Overall, our model-free evidence and PVAR model also indicate that premium converters generally become more active (confirming findings of R. Bapna et al., 2018), which is generally positively associated with SC.

Second, our joint findings—i.e., the experimental study of Weiler and Hinz (2019a) and our replication—add to the stream of experimental research on SC accumulation in SMN. Specifically, we offer a better understanding of which and how experimental stimuli drive the formation of online SC. The multilayered stimuli, the provision of advanced networking features, does not automatically increase SC among individuals, but is conditional on the heterogeneity in the recipient’s agency for networking. This is interesting to researchers facing the tricky objective to experimentally manipulate SC in SMN. For example, artificial manipulations of connections via fake profiles (Toubia & Stephen, 2013) would not properly reflect the resources of SC. Hence, to assess the effect of SC on further outcomes, such as job search (S. Bapna & Funk, 2021), job security (Wu, 2013), or user behavior (Wasko & Faraj, 2005), researchers should organically manipulate SC. Reflecting on the experiment of Weiler and Hinz (2019a), such an organic manipulation in the form of donating a premium membership voucher is only successful if the users are motivated to network. This insight complements previous findings in IS that seek to identify the mechanisms of SC accumulation experimentally. Such studies show that nonreciprocal online stimuli (i.e., e-mails with contact suggestions), known only to the recipient, increase SC (S. Bapna & Funk, 2021). Meanwhile, in non-work-related contexts, experimental donations of prestigious items that are visible beyond the recipient can increase SC by making a user more salient (Hinz et al., 2015).

Third, our joint findings about individual motivation and agency advance the insights of prior IS studies, which assessed measures to enable network change in the wake of additional digital affordances (R. Bapna et al., 2018; Leonardi, 2013; Wu, 2013). The experiment of Weiler and Hinz (2019a) agrees with these studies that the provision of online networking capabilities “was, by itself, incapable of bringing [...] any network change” to them (Leonardi, 2013, p. 772) and that gifting access to premium features does not automatically change user activity (R. Bapna et al., 2018). Beyond this finding, we provide further support that the premium membership’s advanced networking features will only unfold their effectiveness if freelancers are motivated to utilize those features and enhance their network (Weiler & Hinz, 2019a). The users’ agency (i.e., their networking behavior) is what particularly determines the benefit of having access to advanced networking features. This finding adds to the research on individual agency for networking in SMN (e.g., Tröster, Parker, Van Knippenberg, & Sahlmüller, 2019; Vissa, 2012) and confirms theorizing that individual networking agency is the key driver of SC accumulation (Borgatti & Halgin, 2011). Thus, the affordances (i.e., additional action possibilities for networking) of work-related

SMNs do not automatically prompt an action. They only invite the freelancers to form valuable ties for instrumental purposes. Through our replication findings, we provide further support for this “*theory of purposeful feature utilization*” (Weiler & Hinz, 2019a).

In sum, our joint findings contradict a deterministic view of SC accumulation in SMN: Digitalized networks do not emerge exogenously. Researchers must be cautious when interpreting the effect of an individual’s social network position on a certain outcome if they do not take the endogenous nature of the networks into account. For example, Lee (2010) strikingly demonstrated that the positive relationship between individuals’ network positions and their performance vanished, once he controlled for their past-performance. Our findings agree, as they show how the networking affordances of a SMN are only utilized if the users are sufficiently motivated and have an agency for SC accumulation (Ahuja et al., 2012; Stuart & Sorenson, 2007).

3.5.2 *Practical Implications*

Our findings have several practical implications for (work-related) SMNs that want to optimize the features offered by their platforms and strengthen the network of their platform, on the one hand, and for users who want to make the most out of the premium membership, on the other hand. From a platform provider’s perspective, the chief focus should be encouraging users to network and accumulate SC. Users who have established a large network on one SMN are unlikely to repeat the effort on another SMN that fulfills the same purpose. Giveaways of trial periods of premium membership are a common marketing instrument in freemium businesses (Koch & Benlian, 2017), and they can be used to foster network effects, among others, in work-related SMNs. However, such a free giveaway only leads to the formation of new contacts if premium users are motivated to network. Hence, free giveaways are less suitable for “waking up” passive networkers. Instead, SMNs should target those users with a high agency for networking. Otherwise, the platform risks missing out on revenues because some users would otherwise have decided to pay for their premium membership.

The strong association between premium users’ messaging and SC accumulation highlights the role of a particular premium feature: The improved ability to reach out via personal messages. This finding provides two strategic directions for platform operators: Because of the value of personal messages to their users, SMNs should think about differentiating their offer to the actual user needs through the manifold possibilities of

nonlinear pricing (e.g., Schlereth & Skiera, 2012). Instead of offering this important feature in an unlimited way, WorkSMN could differentiate its premium memberships according to the number of messages that one can send to non-contacts (e.g., ten messages for 5€ per month; 30 messages for 10€ and unlimited for 99€). A different strategic direction could be that SMNs incentivize users to actively network: Because SMNs generate a high value from dense networks with active users, the platform could remunerate the most active users through free access to advanced networking features.

Moreover, our findings imply that SMNs should provide templates once a user intends to send a personal message to a non-contact. An algorithm could generate such a template, including a target user, easing the contact initiation based on similarities between profiles and common contacts. Our study also highlights that the opportunities provided by advanced networking features are complex and may be challenging to comprehend for users. As a solution, we suggest providing video tutorials that display and summarize the advanced networking features, alongside with special use cases. Platforms can further improve clarity by embedding instructions and encouragements in their design (Oestreicher-Singer & Zalmanson, 2013).

From a user's perspective, our insights are beneficial for understanding the value of converting to a premium membership. Work-related SMNs, just like other freemium networks, commonly advertise premium memberships with claims along the lines of "Make new connections" and "Your profile is gaining traction".⁷ However, the underlying mechanisms of advanced networking features are usually unclear. Hence, attempts to estimate the contribution of individual premium features in freemium networks have garnered interest (R. Bapna et al., 2016). Findings from a field experiment (Weiler & Hinz, 2019a) and individual-level panel data agree that passive features supply little benefit to SC accumulation itself. Premium users, in work-related SMNs, should not expect that the visual highlighting via prestigious premium badges suffices for successful SC accumulation. Likewise, the "weak signals" provided through profile visiting are not a strong mechanism in work-related SMNs. Instead, the agency for networking is the very basis for successful SC accumulation. Only those users who actively use targeted actions—i.e., profile visiting and personal messaging—that are enabled by active premium features can accomplish increased SC accumulation through their premium membership.

⁷ <https://premium.linkedin.com/>, Retrieved on 07/13/2020

3.6 Limitations and Future Research

A number of limitations could have influenced the identified results. First, Weiler and Hinz (2019a) conducted their follow-up questionnaire about six months after our manipulation. It is difficult to assess whether this period was too short to explicitly see the benefits of the freelancers' networking endeavors, as these might take more time to emerge fully. At least, the model-free visualization of the individual-level panel data suggests that the advanced networking features unfold most of their effectiveness at the beginning of a premium membership (Figure 6).

Second, the generalizability of our findings might be limited because the studies only involved freelancers from one country. Therefore, we cannot say whether the results also hold for other study populations or countries. Furthermore, we solely looked at one specific work-related SMN; hence, our findings may not be generalizable to other platforms. To achieve external validity, future studies should transfer a similar experiment to other platforms (e.g., LI), ensuring that participants differ in their characteristics (e.g., the unemployed or graduates) and/or their geographical location.

Third, our conclusion on RQ2 is based solely on observational data: We generated supportive evidence—without being able to ensure unbiased causal estimates, as in RQ1. To further understand the role of advanced networking features in SC accumulation, future research should try to conduct online randomized field experiments—ideally by directly manipulating platform features (R. Bapna et al., 2016; Gee, 2019)

Finally, all our work focused on the structural dimension of SC, which we operationalized in line with prior research as the number of direct contacts, i.e., node degree (e.g., Hinz et al., 2015; Lanz et al., 2019). While this measure is easy to harvest manually by visiting the member's profile, it is also a very crude measure of the concept: It does not differentiate between strong and weak ties (i.e., meaningful connections and “meaningless” contacts). In addition, by only looking at SC in terms of the plain acceptance of a sent digital contact request, our work did not consider that contacts might possess unequal interest in terms of their potential instrumentality. Thus, future research should explore how and why individuals decide to cultivate some of those accumulated instrumental ties, while allowing others to remain a digital artifact.

4 Online Lead Generation: Learnings from an Emerging Industry⁸

4.1 Introduction

As digital devices become increasingly popular, the amount of digital touchpoints within the customer journey is gradually growing. Whereas consumers frequently research and compare durable and high-priced products and services online, this behavior is less popular in consumable categories (Nielsen, 2016). Therefore, the more expensive, complex, and individualized a product or service is, the more likely it is that consumers will actively investigate these products online and seek guidance in the purchase process. Whereas previously, the purchase of a new kitchen started with a conversation with a sales agent in a brick-and-mortar kitchen store, these conversations are replaced by touchpoints and interactions in the digital space. Novel conversational commerce interfaces, like chatbots, even increase this trend further. This is a challenge for product and service providers, who typically do not have the competencies, knowledge, or reach to engage users online across all channels.

Online lead generation companies (OLGC) fill this gap. As intermediaries, they connect the customers with the service or product providers. OLGCs generate their revenue by capturing leads through various value propositions and eventually selling them to third parties. This way, a completely digital industry has emerged. Major players, like Verivox for insurances and telecommunication, Audibene (i.e., hear.com) for hearing aids, and Enpal for solar modules are estimated to make several millions of Euros in revenue. This nascent business model has transferred large parts of traditional sales conversations to the digital domain. Through forms, chatbots, messenger services, and phone consultation, these OLGCs fulfill large parts of the sales conversation digitally.

The business model of OLGCs is, to a great extent, a blind spot in the literature. With this article, we follow the overarching objective to illuminate this blind spot by investigating the OLGC industry through a qualitative study. We develop a framework of the lead generation process and highlight novelties of the digital customer journey. Along the funnel, we derive

⁸ This project is based on the interviews conducted in Wisskirchen (2020) “Online Lead Generation Companies: A Segmentation”. The project is co-authored with Kilian Wisskirchen, Prof. Dr. Christian Schlereth, and Alexander Hoffmann and was published in the St. Gallen Marketing Review under the title “The Emerging Business of Online Lead Generation” (2021).

best practices and recommendations for the industry, which are also relevant to managers seeking to expand their company's digital channels.

4.2 Theoretical Background

4.2.1 *Online Lead Generation Companies*

The term lead generation reflects the “generation of information about prospective customers through targeting, prospecting, information dissemination, and persuasion.” (Bannerjee Bhardway, 2019) OLGCs specialize in this task. One can think of the core business of OLGCs, metaphorically, as the activity of fishing: OLGCs capture online leads (the fish) through varying value propositions (the hinge and the bait). A common value proposition is a *personalized* overview of quotes (e.g., for a credit loan), which would take a long time to aggregate manually. Next to the value of personalization, the example of credit loans highlights how *transparency* is a further important value proposition of these businesses by offering price comparisons and customer references. Besides, some OLGCs act as a *reliable partner* in complex acquisitions and orchestrate steps of the value chain (e.g., kitchens or solar modules). Lastly, the online setting offers *discretion*, which, for example, may be important in the case of plastic surgery mediation or other health services and is not possible in an offline setting. In summary, the value propositions for consumers can differ substantially, depending on the vertical (i.e., particular offering).

This way, for cooperating third party providers, OLGCs provide an appealing source of customers and revenue. For routing the leads to those third parties, OLGCs are commonly paid per lead or conversion and act as an intermediary platform. However, some OLGCs have also started to orchestrate related activities and services (e.g., installation), occasionally even offering their own products. This way, these companies can remain part of the customer journey but forego the lean processes and cost-efficiency of a pure platform provider.

Few research papers explicitly assess the process of online lead generation and, to our best knowledge, none describe the business model of OLGCs in detail. While aspects of the lead generation context have been explored in offline settings (e.g., Smith, Gopalakrishna, & Chatterjee, 2006), our research reflects specifically on the digital setting of OLGCs. It adds insights into how these firms thrive in the online domain.

4.2.2 *Related Frameworks*

By acting as intermediaries, OLGs operate at the marketing-sales intercept (Smith et al., 2006) and inherently capture a major part of the sales process. Sales funnel frameworks commonly describe the sequence of stages in the sales process. With each stage, fewer customers convert to the next stage, generating a “funnel” shape. The most popular model to segment the sequence of stages that a consumer traverses psychologically is the Attention-Interest-Desire-Action (AIDA) framework (Strong, 1925).

However, also from a practical perspective, the sales literature offers models to classify individual stages in the process. For example, Johnston and Marshall (2016) describe the sales process stages as prospecting customers, opening a relationship, qualifying prospects, presenting a sales message, and closing the sale. Similarly, from a firm perspective, Smith et al. (2006) model the lead generation activity along the sequence of milestones: Phone call, sales visit appointment, and sales. This paper deviates from this classical “offline” perspective and proposes a model for the online lead generation case.

4.2.3 *Best Practices and Trends from the Literature*

Prior studies point to best practices in the online sales process. For example, whereas traditionally lead generation draws from offline marketing channels, like radio or direct marketing (Smith et al., 2006), online marketing is especially important in the digital customer journey (Kannan & Li, 2017). Moreover, to capture user information, online forms provide an automated interface through which interested users can generate requests (Banerjee & Bhardwaj, 2019). In comparison to in-person sales conversations, online forms are less costly for the companies to implement. An emerging trend is to use chatbots that dynamically respond to users within the scope of a pre-defined topic range. Compared to static input forms, chatbots evoke trust, which results in higher recommendation acceptance (Hildebrand & Bergner, 2020). Whereas chatbots are noted to more naturally follow turn-based conversation patterns, forms allow to include progress bars, which motivate users to progress.

Another important aspect is the timing of follow-up calls, Smith et al. (2006) find in offline settings that delays in responding to customer requests lead to fewer sales. Similarly, studies of online businesses indicate that the conversion probability is 60 times higher if companies respond within an hour (Oldroyd, Mcelheran, & Elkington, 2011) and that the probability of reaching an online lead within five minutes versus 30 minutes diminishes

sharply (Elkington, 2013). Hence immediate follow-ups appear optimal to convert leads to appointments and consequentially to sales. Regarding additional call metrics, a total number of seven attempts, six to eight days between the first and the last attempt, and a time gap of one to two days between the attempts are noted best practices (Elkington, 2013; Xant.ai, 2019).

4.3 Research Context and Methods

We conducted 19 semi-structured interviews with founders and managers from different OLGs and diverse verticals founded between 1991 and 2018. We capture a wide array of characteristics and OLG business models: One differentiating factor was the industry. We conducted at least one interview in the industries real estate services, mobility services, financial services, funeral services, energy services, house services, luxury products, travel services, and health services. Among the 19 interviewees were five founders. The rest of the interviewees had high-level marketing management roles.

Most of the analyzed firms had more than 200 employees (13 companies), generated more than 35 Mio. EUR in revenue per year (10 companies) and were founded before 2017 (15 companies). Thus, most of these companies have a proven business model with a functioning lead management system. Even though we do not cover the whole online lead management market, the selection of interview partners is exhaustive and they cover representatives of the most important OLG verticals.

The interviews (with an average duration of 32 minutes) were conducted in German and recorded via Zoom. Afterward, the recordings were transcribed, coded, and evaluated to summarize qualitative coding analysis (Mayring, 2010). The interview guide consisted of four high-level questions touching the four dimensions of lead generation, monetization, marketing, and industry specialties (i.e., How do you design the lead management process? How do you monetize your leads? What marketing channels do you use? What are the specifics of your industry for OLGs?). Most of the questions of our semi-structured interviews were open, with no pre-defined answer suggestions. This allowed us to deepen some questions, depending on how the conversation developed.

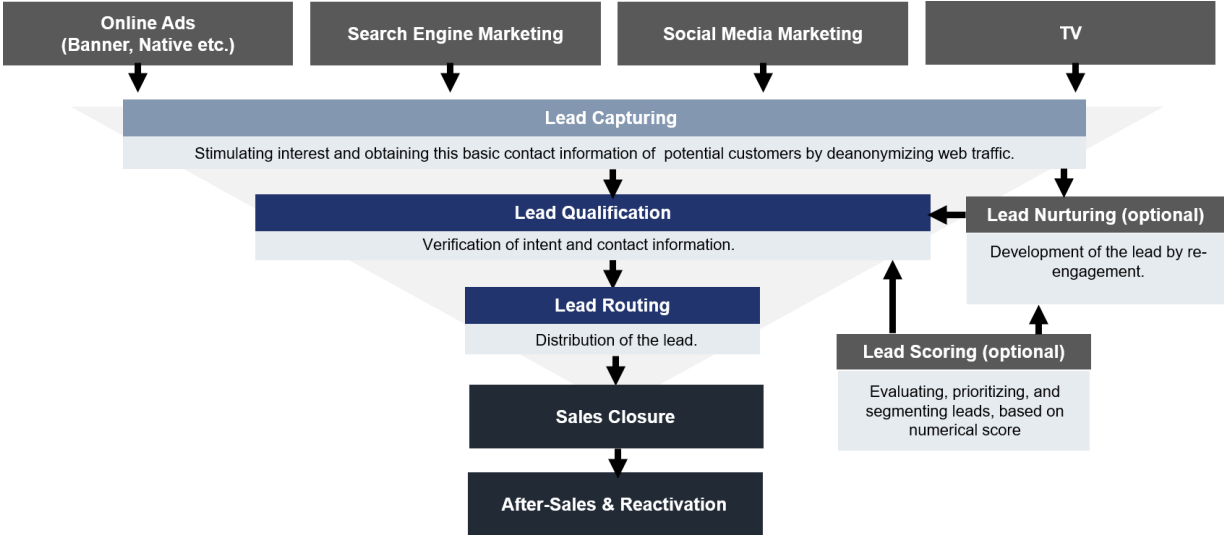
4.4 Online Lead Generation Funnel

Building on prior literature and the insights of our interviews, we offer a high-level overview of the lead generation process. We distinguish the key stages in the OLG process.

This practical online lead generation framework, shown in Figure 7, serves to structure the best practices gathered in the following section.

The lead generation funnel consists of various marketing and advertisement activities: Whereas in traditional settings, the marketing activities directly transfer to a sales contact (Smith et al., 2006), in the online setting, the users find their way to the OLGC website. Through advertisements users are directly routed to the OLGC’s website, which are the starting point of the lead generation funnel. In contrast to traditional sales processes (Johnston & Marshall, 2016, pp. 44-50), that involve prospecting users (e.g., finding interested customers, often by “knocking on doors”), the users are primarily attracted through marketing efforts and make themselves “the first move”. This is an important advantage, as leads who visit the OLGC websites commonly already have a high interest in a purchase.

Figure 7. Online lead generation funnel (own illustration)



The first stage of *lead capturing* has the two-fold objective of stimulating interest and obtaining potential customers’ basic contact information. The OLGC website commonly provides both functions. That implies that the OLGC websites offer information on the product or service that they are offering. Additionally, the websites create trust by presenting customer references, certification badges, and mentions in popular news outlets. Importantly, a user interface (e.g., a form) is prominently displayed on the websites, enabling users to easily configure their specific setting and, importantly, also input their contact data.

Lead qualification refers to determining if a lead meets the requirements of purchase (i.e., is a “qualified lead”). In the online domain, this often involves verifying and assessing the collected information. Hence, technical procedures can partly verify the buyer’s intent and

contact information. This stage can already involve dynamic conversations. In our interviews, several conversation channels appeared and were often associated with very different conversion rates. For example, chatting with WhatsApp might be perceived as a lower communication hurdle than a personal phone call. However, it has limited scaling possibilities due to missing interfaces and strict government regulations.

Additionally, leads may optionally enter the *lead nurturing* stage. The goal of nurturing is to develop the lead towards the qualification requirements and develop the lead towards a purchase decision. Lead nurturing targets users who already disclosed their contact information but remain in an ambiguous position in the middle of the funnel (Patterson, 2007). Through re-engaging, customer development, and education activities, those undecided leads are further developed towards a conversion. As our interviews show, the intensity of those activities often depends on the vertical. Our interviewees report that personal communication is key in this stage so that OLGC typically use phone- or e-mail conversations.

The stages of *lead qualification* and *lead nurturing* are often supported by *lead scoring* systems. Lead scoring aims to quantify the potential of leads and to segment them based on behaviors and attributes (McGlaughlin, Doyle, & Balegno, 2012). For example, lead scores can reflect whether a lead matches the typical buying persona and how developed the lead is. The initial lead score can change over time because of the lead's subsequent behaviors (e.g., visiting the pricing website). A lead scoring system also helps companies to prioritize leads and allocate their resources. Studies show that predictive lead scoring and prioritizing can significantly increase revenue (Säuberlich, Smith, & Yuhn, 2005).

If leads are qualified, they are assigned to sales agents for the finalization of the sale (*lead routing*). Among OLGCs, the strategies of lead routing vary—while some OLGC forward their leads exclusively to a single third party provider, others forward the same lead to multiple ones. Some OLGCs capture the following closure of the sale themselves, so that the lead routing only implicitly occurs on the OLGC website.

Lastly, reflecting the *sales closure* stage, an offer is drafted, and the sale is eventually closed (i.e., concluded). Moreover, *after-sales* activities aim to sell complementary services and products. More recently, more and more OLGCs realized the potential of *reactivating* successfully routed leads by following-up after some time has elapsed.

4.5 Findings and Recommendations

We derive various practical insights from our interviews, which we present in the following in sequence of their occurrence along the online lead generation funnel.

4.5.1 *Lead Capturing: Attracting Leads Through Information*

An important aspect of sales conversations is which party initiates them. Our interviewees highlight that it is important to motivate users to disclose their personal information by offering desirable information. Two strategies became apparent in the interviews. Firstly, interviewees report providing related information pieces (e.g., rent index by postal code, residual value estimations, or industry reports) on the main websites directly as a “quid pro quo” incentive for the user to disclose contact information themselves—the frequently used industry term of a “lead magnet” captures this intuition nicely. Secondly, in a similar vein, six interviewees pointed to the practice of informing leads via additional websites. One interviewee reports that they “look at what kind of topics the users are interested in” and offer this content on “satellite websites”. These satellite websites are typically branded differently and, at first glance, appear to provide an objective assessment (e.g., imitating a “newspaper” article). However, small logos and the disclaimer section reveal the affiliation of the website to the OLG. As our interviewees report, users browsing those websites often have a substantially higher conversion rate.

Recommendation: Offer related information pieces and satellite websites to direct the attention to your business or service and encourage leads to make the “first move”.

4.5.2 *Lead Capturing: The Power of Multi-Step Forms*

Replacing the first steps of a sales conversation in a brick-and-mortar store, the OLG websites offer means to capture user preferences and personal details. Most commonly, our interviewees emphasize multi-step forms as an essential lead capturing element. As one interviewee reported, similar to in-person sales conversations in a brick-and-mortar store, multi-step forms typically first ask for simple preference questions related to their search (e.g., desired credit loan amount). Personal details, like phone numbers, are typically among the last questions.

Moreover, the emerging technology of chatbots is particularly noted by our interviewees, as a trend to automate conversations, while still keeping a natural conversation flow. Three participants reported having tested chatbots. While some found inferior conversion

performance as compared to multi-step forms, another OLG reports having experienced decreases in cost. A key difference is that click-based interfaces provide a passive means of interaction by the user, whereas initiating a new conversation requires proactivity. In this regard, one interviewee pointed out experimenting with chatbots that include click-based interfaces, so that the technologies of multi-step forms and chatbots converge.

Recommendation: In the first stage of capturing leads, utilize multi-step forms to abstract a sales conversation. The forms should ask customer preferences first and ask about personal details last.

4.5.3 Lead Nurturing and Qualification: Prompt Responses Are Not Always the Best Option

While some interviewees report that they indeed try to follow-up on leads as immediate as possible, as suggested by the literature, they found via experimental testing that a delayed response is favorable. E.g., one interviewee stated that an immediate response led potential leads to be “scared of in the sense that it generated the impression of ‘we are watching you’”. This observation may be a specific of online businesses that enable almost immediate follow-ups, previously impossible in an offline setting. Another interviewee who is working for an internationally operating company, explained that a lead gets ‘cold’ much faster in the American than in the German market. Hence, country-specific characteristics may provide an alternative explanation for this finding. To illustrate this, we contrast the reported average optimal call metrics with those reported in US studies, discussed in Section 4.2.3, in Table 14. Besides, customers get annoyed more quickly, as reflected by the lower optimal number of attempts.

Table 14. Comparison of average metrics stated by the participants and prior best practices

<i>Metric</i>	<i>Lead Management Practices (Our study)</i>	<i>Literature Best Practice*</i>
<i>Response time</i>	88 minutes	5 minutes
<i>Number of attempts</i>	4.8 attempts	7 attempts
<i>Duration to last attempt</i>	2 days	6-8 days
<i>Time between attempts</i>	1 day	1-2 days

*Elkington (2013); Xant.ai (2019)

Recommendation: When initiating in-person follow-up response, assess whether your calls are too prompt or too slow through customer surveys and experimentation. A prompt response can often be favorable, yet in some situations, a slower response is advisable.

4.5.4 *Lead Scoring: Data-Rich Environments Enable Lead Scoring*

Not only does the customer enter their self-disclosed information, but additional information is available through technical browsing data. This enables data-driven lead scoring based on numerous attributes, previously impossible in the offline domain. An interviewee described: “The variables for the lead scoring system are either provided intentionally by the user, e.g., postal code, or unintentionally, their device. (...) An iPhone user is worth much more than an Android user”. Overall, the interviews highlight that OLGCs understand lead scoring as an important trend and seek to implement it in their business model. Among our 19 interviewees, seven already had implemented a lead scoring system, and four reported wanting to implement it.

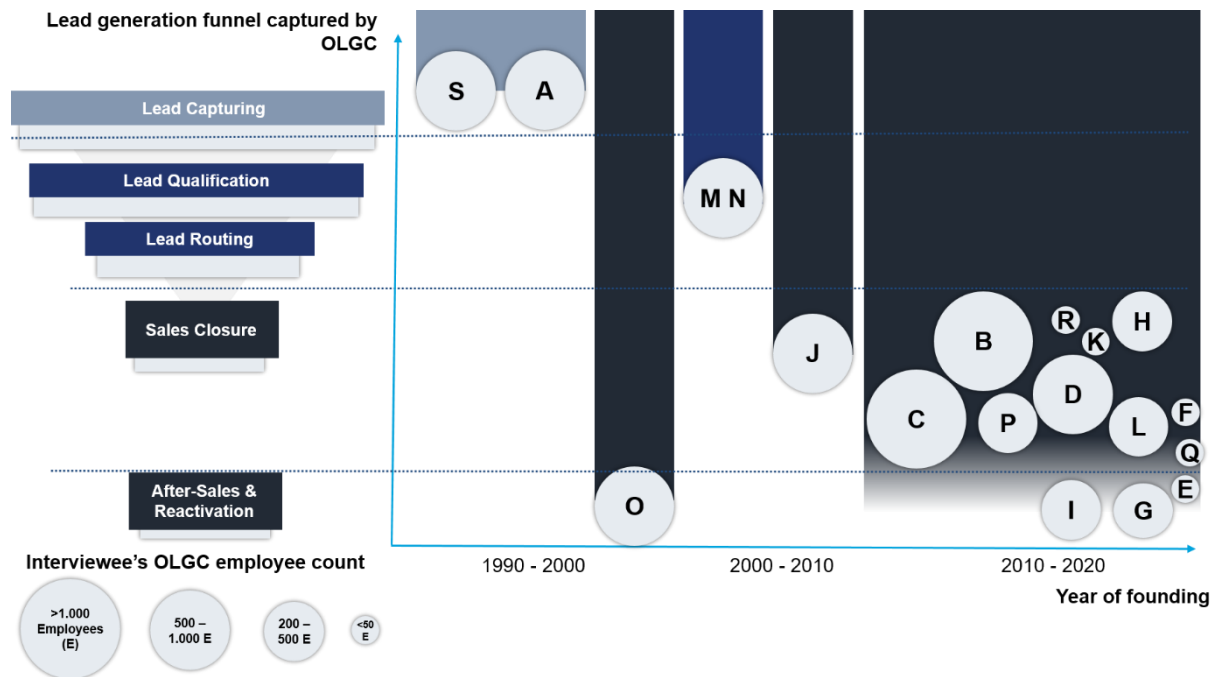
Recommendation: Lead scoring based on the abundantly available data in an online setting helps to prioritize leads.

4.5.5 *Sales Closure and After-Sales: Trend Toward Business Model Expansion*

Lastly, our interviews highlighted that the OLGCs differ in their coverage of online lead generation funnel stages. Typically, OLGCs cooperate with multiple partners that seek to offer their products or services. However, various OLGCs even handle the closure of the sale on their website. Moreover, some OLGCs expressed an ambition to engage in after-sales activity and to reactivate leads. Often OLGCs start with a lean “lead capturing” approach and successively expand their activities from there. As one interviewee remarks, “*we used to have a business model that was purely based on the capturing and forwarding of leads. (...) Today we fulfill those [authors note: related business] activities ourselves*”.

To visually explore the market segmentation, we form three categories along with the previously outlined online lead management funnel. Thereby, we link the anonymized OLGCs of the interviewees alongside the OLGCs’ founding date in Figure 8.

Figure 8. Lead Combination Trend (own illustration)



OLGCs in our research founded after 2010 offer the full online lead management funnel, whereas, before 2010, they only fulfill *lead capturing*, *-qualification*, and *-routing*. As one of our interviewees reflects, “*by selling the product, we can increase our margin and at the same time define the full customer journey*“. Regarding after-sales stages, only a few OLCGs are active. The primary reason is that, especially for OLCGs that capture leads for durable and expensive products, there is only a single customer contact. In summary, increased coverage of lead generation funnel steps, serving additional verticals, and offering products are frequent growth strategies.

Recommendation: Lead capturing is an excellent starting position to grow a business through additional lead generation stages. Ambitions to engage in after-sales activities should reflect the characteristics of the vertical—especially for durable products, after-sales is often not promising.

4.6 Discussion and Further Research

Our study contributes to the marketing and sales literature by providing insights into the emerging business model of OLCGs that largely redefines the sales conversation process. To the best of our knowledge, we are the first to provide an empirical study on this topic. Our study points out important differences in the online lead generation funnel compared to the classical sales process (Johnston & Marshall, 2016, pp. 44-50). Overall, we find that OLCGs

have replaced large parts of the sales conversation through conversational commerce. Especially at the funnel's beginning, conversational touchpoints are automated by the OLGCs through their websites. In contrast, in-person conversations start at a later stage in the sales process, if at all.

We find that traditional conversational commerce techniques, especially online forms, are currently the most widely used tool to converse with leads in the early stages of the online lead generation funnel. This is in line with prior descriptions of online lead generation (Banerjee & Bhardwaj, 2019). However, OLGCs recognize chatbots as an evolving trend that fits to their digital business model very well. Currently, OLGCs are still in the stage of testing and evaluation for which tasks to utilize chatbots and how to reap the most benefits from them. Hence, while chatbots offer appealing advantages (Hildebrand & Bergner, 2020), they are still in an early adoption stage.

When involving sales personnel to converse with users, our research challenges the common recommendation of immediate follow-ups (Smith et al., 2006) in that our interviewees report varying instances of delaying follow-ups to be optimal. In contrast to the companies in the sample of Oldroyd et al. (2011), the OLGCs in our study have rigorous testing in their DNA. We, therefore, hypothesize that these differences may be due to cultural norms. For example, Americans are less inclined to scrutinize processes and are more open-minded towards technological advances based on the Hofstede cultural dimensions theory (e.g., Sorge & Hofstede, 1983). Moreover, the urgency of a product could also determine the optimal response time. Hence, another explanation for the differences in the results could be that optimal response times vary for each vertical. For example, the finance industry is more tolerant about delays compared to health-care customers⁹.

We identify emerging trends within the OLGC industry. One trend is the increased usage of user data to enable lead scoring, especially, when following up on leads must be prioritized. Further, we find a trend towards the expansion of OLGC business models. As OLGCs reflect digital platform businesses, which commonly can quickly build market power, prior studies (Kannan & Li, 2017) particularly emphasize the increasing efficiency and returns (e.g., ability to negotiate better prices). As our research shows, another form of utilizing this increasing market power is to start competing with the third-party partners through capturing further sales generation funnel stages, expanding to additional verticals, orchestrating related

⁹ We thank James Oldroyd for contributing this interesting thought and discussing this idea with us.

services, or even developing their own product offerings. Yet, as our interviewees reflected on their strategic positioning, the disadvantages of an expansion may be that they would upset third-party partners and increase the complexity of their business operations drastically.

Concluding, our study points to future research opportunities regarding the conversational process of online lead generation. Future research could investigate the role of cultural norms and differences in product or service characteristics to develop a framework to identify factors that determine optimal call response times. Additionally, our study points out various conversation possibilities like forms, chatbots, messenger services, and in-person conversations—surprisingly, the simple multi-step form and phone call combination appears the most popular. Hence, future research could investigate the adoption of these alternative conversational channels in practice or compare them empirically in a randomized experiment.

5 General Conclusion

This dissertation sets out to advance the understanding of targeting and empowering users in a digital era. I present three dissertation projects that tackle the topic from different angles: Firstly, “Predicting Tie Strength with Ego Network Structures” looks at the setting of online social networks and demonstrates the feasibility of targeting best friends of a given person via (ego) network measures. Secondly, “Social Capital Accumulation and Premium Memberships in Work-Related Networks” shows that targeting and empowering users with premium membership in a work-related network setting especially benefits those “power networkers” that already have a high networking activity. Thirdly, “Online Lead Generation: Learnings from an Emerging Industry” looks at the intermediary function that online lead generation companies fulfill by capturing users in their online product research process and uncovers best practices of the industry.

On a high level, the first two projects in this dissertation take two fundamentally different quantitative approaches, exemplifying the distinction of explaining vs predicting (Shmueli, 2010). While the first project “Predicting Tie Strength with Ego Network Structures” deals with the question of whether one can predict a “hidden” attribute with observational data (i.e., focuses on the ability to predict hold-out samples using k-fold cross-validation, precision scores, and ROC curves), the second project on “Social Capital Accumulation and Premium Memberships in Work-Related Networks” seeks to explain what the isolated effect of a treatment is (i.e., focuses on the output of inferential statistics, p-values, and effect size estimates). Prediction papers are a rare occurrence in the field of management literature (Shmueli & Koppius, 2011) but gain in popularity with the maturity and propagation of Machine Learning (ML). However, researchers should not conflate both worlds. As explained in Ebbes, Papies, and Van Heerde (2011), removing unobserved confounders (e.g., through using instrumental variables) when trying to explain an effect will decrease the predictive power of a model. Similarly, studies that seek to explain an effect commonly address the issue of underestimated effects due to imbalanced rare-event datasets (King & Zeng, 2001) through over- and undersampling. As we show in “Predicting Tie Strength with Ego Network Structures”, adapting probability threshold values can be a feasible solution to address a systematic underestimation in prediction settings.

Concluding, as the three dissertation projects show, especially in networking services the availability of user-specific data enables businesses to target users with high precision, using

observed characteristics and even “hidden” attributes that can be inferred from their user data. The possibilities of this targeting remains a disputed subject, due to practices of political micro-targeting (Cadwalladr & Graham-Harrison, 2018). However, as the projects also show, digitalization likewise empowers users: Through personalization and the targeted provisioning of premium features, users benefit from this digitalization process and can derive great benefits from the digital platforms. After all, there are countless convenient digital services that arguably most people would not want to give up on (Brynjolfsson et al., 2019).

6 Appendices

A1 Computation of Centrality Measures

Degree Centrality C_D
(Freeman, 1978, p. 220)

$$C_D(p_k) = \sum_{i=1}^n a(p_i, p_k)$$

is calculated for each focal node p_k in a network consisting of n nodes, where $a(p_i, p_k) = 1$ if p_i and p_k are connected. I.e., C_D is the count of all adjacent nodes.

Betweenness Centrality C_B
(Freeman, 1978, p. 223)

$$C_B(p_k) = \sum_{i=1}^n \sum_{j=1}^n b_{ij}(p_k)$$

is calculated for each focal node p_k in a network consisting of n nodes, where $b_{ij}(p_k)$ is the probability that p_k falls on a randomly selected geodesic (i.e., shortest path) linking p_i and p_j .

Dispersion D
(Backstrom & Kleinberg, 2014)

$$D(U, A) = \sum_{S, T \in CN} dist(S, T)$$

where CN refers to the set of common neighbors of ego node U and adjacent alter A . S and T refers to a pair of common neighbors of U and A . Distance measure $dist$ equals 1 if there is no direct link between S and T and no common neighbors exist in the ego network, and 0 otherwise.

A2 Computation of Centrality Measures Computation of Centrality Measures

<i>Structural Embeddedness</i>	<i>Bridging Position</i>	
<i>Degree Centrality</i>	<i>Betweenness Centrality</i>	<i>Dispersion</i>

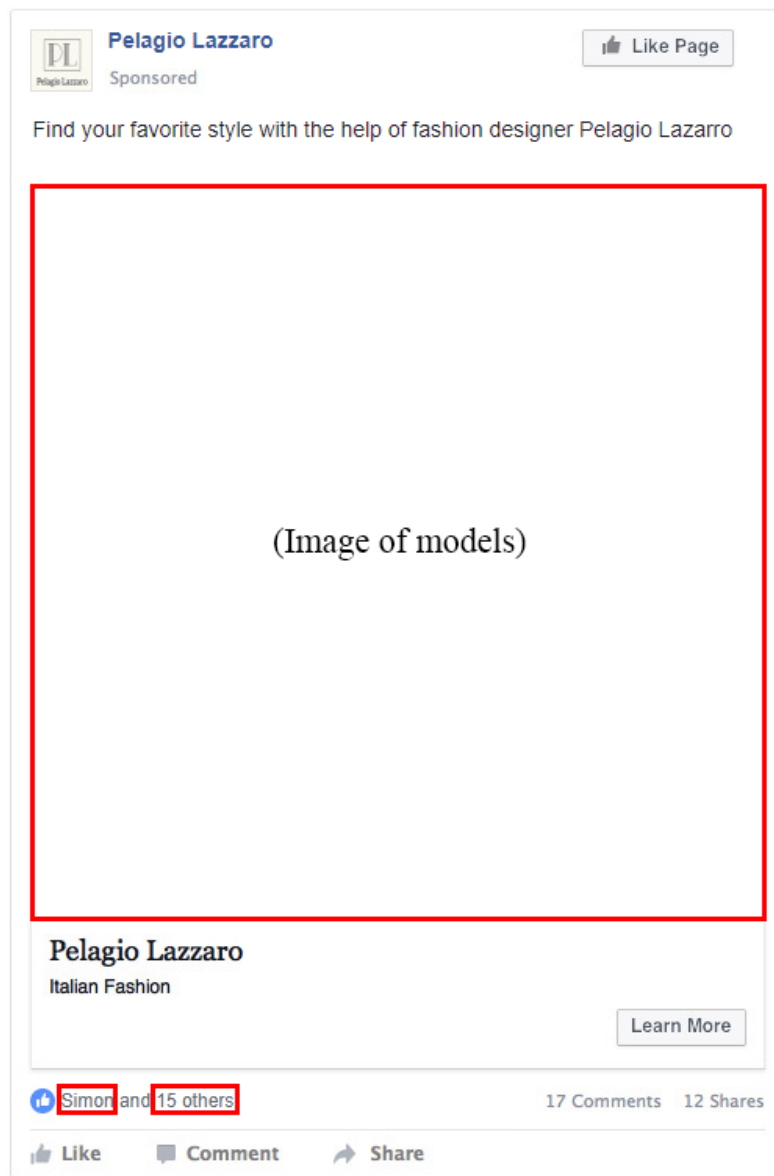
A3 Kendall Correlation Matrix – All Network Measures for Adjacent Alters of Ego

	<i>Degree (Ego)</i>	<i>Betweenness (Ego)</i>	<i>Dispersion (Ego)</i>	<i>Degree (Sampled)</i>	<i>Betweenness (Sampled)</i>
<i>Degree (Ego)¹</i>	1.00				
<i>Betweenness (Ego)¹</i>	.42	1.00			
<i>Dispersion (Ego)¹</i>	.13	.55	1.00		
<i>Degree (Sampled)</i>	.68	.39	.20	1.00	
<i>Betweenness (Sampled)</i>	.38	.49	.39	.55	1.00

Note: We report the Kendall correlation coefficients, to account for non-normality.

¹*All Generalized Variation Inflation Factors (GVIF) are below the threshold of 3, which indicates that no multicollinearity is present when including degree, betweenness, and dispersion*

A4 Impact of Close Friends Survey Items



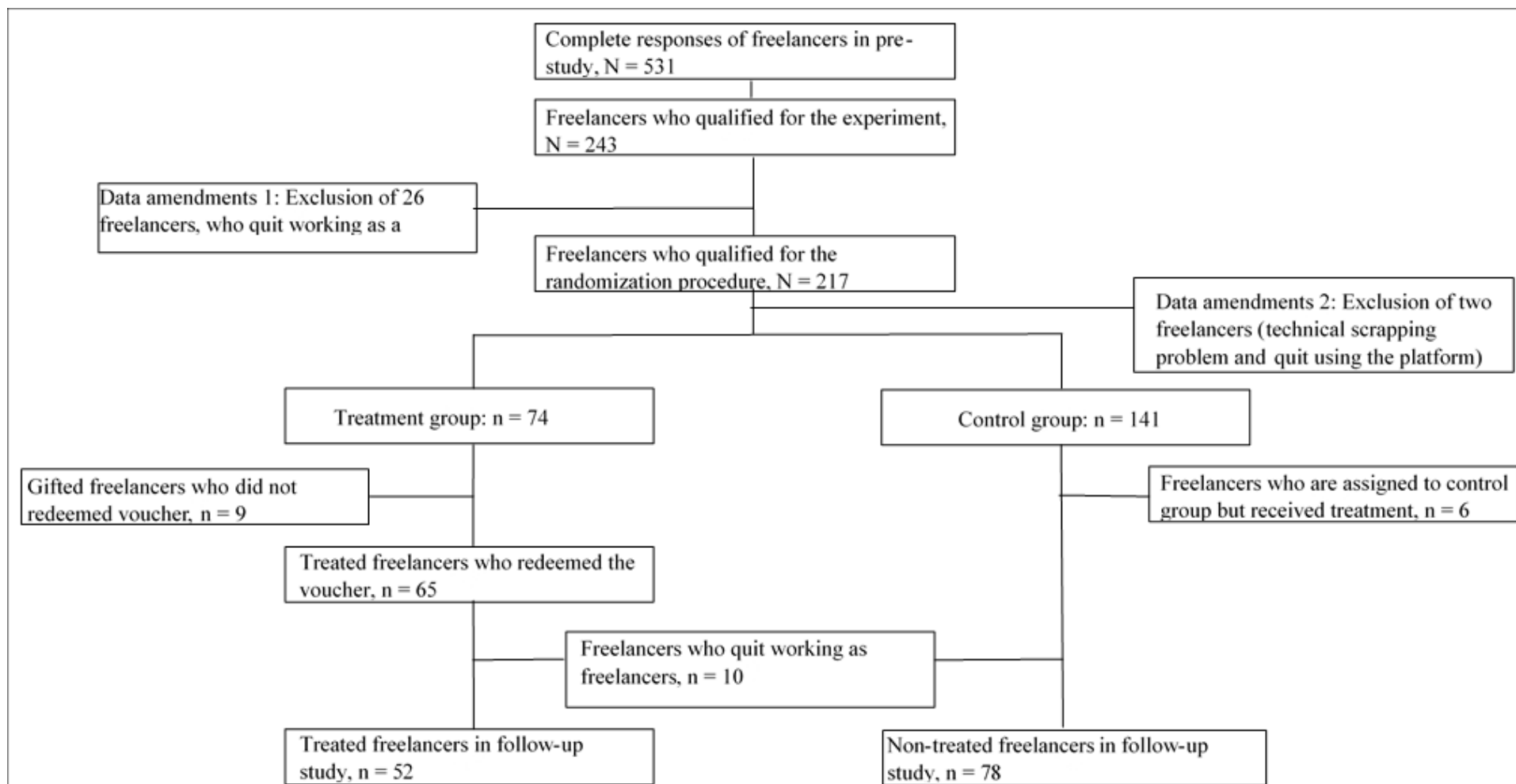
Elements framed with a red box were randomly displayed dynamically (with names based on prior mentions). All images showed a female and male model to ensure gender-neutral targeting.

A5 Overview of Features on WorkSMN by Membership Type

	<i>Basic membership</i>	<i>Premium membership</i>	<i>Active or Passive</i>
<i>More detailed profile visitor's statistics</i>			
<ul style="list-style-type: none"> • View profile visitors, including their full name, position, and company 	-	✓	Passive
<ul style="list-style-type: none"> • Find out how profile visitors came across your profile and the date they visited it 	-	✓	Passive
<ul style="list-style-type: none"> • Link to the visitor's profile page 	-	✓	Passive
<i>More professional appearance through customization of profile page</i>			
<ul style="list-style-type: none"> • No ad banners visible 	-	✓	Passive
<ul style="list-style-type: none"> • Possibility to customize business card with individual cover image 	-	✓	(Both)
<ul style="list-style-type: none"> • Possibility to highlight your top skills 	-	✓	(Both)
<ul style="list-style-type: none"> • Possibility to add and embed more multimedia content to your profile page including videos from YouTube 	-	✓	(Both)
<i>More advanced member search and getting found more easily</i>			
<ul style="list-style-type: none"> • Member search, maximum number of displayed search results 	10	300	Active
<ul style="list-style-type: none"> • More search options, including special search filters (e.g., working field, company, location) 		✓	Active
<ul style="list-style-type: none"> • Possibility to create automatic search alerts 	-	✓	Active
<ul style="list-style-type: none"> • Highlighted entry of your profile in search results 	-	✓	Passive
<i>More efficient communication and networking</i>			
<ul style="list-style-type: none"> • Possibility to send contact requests with a personalized message plus to browse and edit pending contact requests 	-	✓	Active
<ul style="list-style-type: none"> • Possibility to send messages to non-contacts 	-	✓	Active
<ul style="list-style-type: none"> • Updates of network contacts are visible at a glance (e.g., if they changed a job) 	-	✓	Passive

A6 Experiment: Flow Diagram

Figure 9. Flow diagram of the freelancers' recruitment, treatment condition assignment, and attrition—as in Weiler and Hinz (2019a, p. 203)



A7 Experiment: Descriptive Statistics and Bivariate Correlations

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
[1] <i>ΔSC accumulation</i>												
[2] <i>Treatment received</i>	0.027											
[3] <i>Treatment assigned</i>	-0.020	0.920										
[4] <i>Gender</i>	-0.118	0.076	0.158									
[5] <i>Level of education</i>	0.028	-0.008	-0.058	0.022								
[6] <i>Age</i>	-0.212	0.062	0.081	0.236	0.005							
[7] <i>Openness</i>	-0.054	0.136	0.067	-0.193	0.012	-0.278						
[8] <i>SMN tenure</i>	-0.455	0.058	0.072	0.187	0.103	0.294	-0.019					
[9] <i>Strategic networking behavior</i>	0.301	0.038	0.052	-0.014	0.044	-0.150	0.090	-0.162				
[10] <i>Other income</i>	-0.083	0.093	0.071	0.083	0.056	0.134	-0.104	-0.007	-0.073			
[11] <i>Work hours per week</i>	0.014	-0.124	-0.148	0.130	0.094	0.005	0.070	0.041	-0.038	-0.227		
[12] <i>WorkSMN usage intensity</i>	0.275	0.209	0.153	0.077	0.163	0.164	-0.021	0.043	0.236	0.033	0.037	
<i>Mean</i>	10.209	0.400	0.392	1.569	2.831	44.946	5.494	7.184	2.358	0.331	38.331	3.615
<i>SD</i>	15.890	0.492	0.490	0.497	0.376	10.989	1.126	3.847	0.911	0.472	16.296	1.572

Notes: Significance level $p < 0.05$, as indicated by bold numbers

A8 Experiment: Elaboration on the Complier Average Causal Effect (CACE)

In this section, we elaborate on why the CACE is the most appropriate strategy for the research purpose of Weiler and Hinz (2019a) and provide additional (technical) details:

Issues of the Per Protocol and As-Treated Approaches to Handle Noncompliance

The per protocol (PP) and as-treated (AT) approaches are rather inadequate in their way of addressing noncompliance, as they take away the salient and unique benefit that is established through the randomization process, i.e., the introduction of balance of observed and unobserved characteristics across the treatment and control condition. While the AT approach reassigns noncompliant participants for the analysis, the PP approach excludes those deviant ones altogether (Sagarin et al., 2014). To put this into perspective, the AT approach looks at participants as they self-select into the treatment condition and, therefore, pays no attention to the initial group assignment (Sagarin et al., 2014). For instance, if a freelancer assigned to the control group does buy a premium membership subscription on his/her own, this formerly untreated freelancer will now be analyzed as treated. Likewise, a freelancer who is assigned to receive the treatment but does not redeem the corresponding premium membership voucher will be reassigned to the untreated condition. The PP approach, however, keeps the initial group assignment intact, but simply eliminates from the analysis all those participants who did not stick to their assigned treatment condition (Sagarin et al., 2014). Unsurprisingly, “[...] in the presence of noncompliance, there is no compelling justification for these two [...] approaches [i.e., AT and PP]” (Imbens & Rubin, 2015, p. 535; Sagarin et al., 2014).

In comparison with the AT or PP approach, the intention-to-treat (ITT) approach focuses solely on the groups as they were initially created by the randomization process (i.e., the actual assignment). Consequently, this approach is advantageous, as it perpetuates the initial groups and is therefore able to produce causal estimates. Despite this undeniable benefit, however, the ITT neglects the actual receipt of the treatment and therefore no empirical evidence emerges about this effect of interest (Sagarin et al., 2014). Taken together, as highlighted, none of these three outlined approaches can produce unbiased causal estimates regarding the effect of the treatment received (Imbens & Rubin, 2015; Sagarin et al., 2014). Thus, against this backdrop, we chose the CACE approach on account of the fact that, in the face of noncompliance, it is the only strategy that is able to reveal a causal estimate of treatment received on the outcome, namely for the subgroup of the compliers (for a detailed description of the term see the next subsection) (Imbens & Rubin, 2015).

Rationale of the CACE approach

We follow Gerber and Green (2012) and distinguish experimental subgroups. We use the letter z to reflect if freelancers are assigned to the control group ($z = 0$) or the treatment group ($z = 1$). Also, d_i refers to whether the corresponding freelancer is actually treated ($d_i = 1$) or not ($d_i = 0$). As described in Weiler and Hinz (2019a, p. 204), four different groups exist:

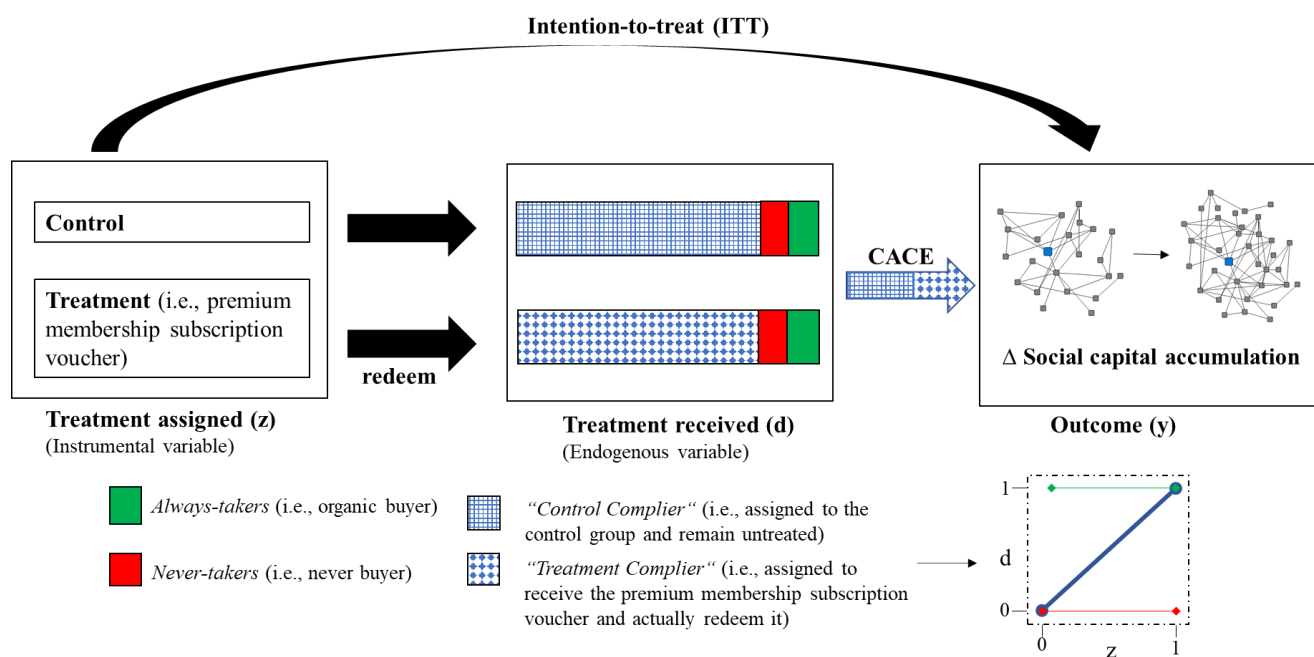
“(1) We can classify freelancers as ‘*compliers*’ when they are a) assigned to receive the premium membership subscription voucher (treatment group) and they actually redeem it ($d_i(1) = 1$), as well as when they are b) assigned to the control group and they remain untreated (i.e., they do not voluntarily buy one premium membership subscription on their own) ($d_i(0) = 0$). [Note: The rationale of what they delineated is visualized by the blue cross-hatched areas in our Figure 10]

(2) We can categorize freelancers as ‘*never-takers*’ when they are a) assigned to the treatment group, but do not actually redeem the received premium membership subscription voucher ($d_i(1) = 0$), as well as when they are b) assigned to the control group and do not voluntarily buy a subscription on their own ($d_i(0) = 0$). [Note: In our Figure 10, never-takers are displayed by the red colored area].

(3) We can group freelancers as ‘*always-takers*’ when they are a) assigned to receive the premium membership subscription voucher and they actually redeem it ($d_i(1) = 1$), as well as when they are b) assigned to the control group, but voluntarily buy a premium membership subscription on their own ($d_i(0) = 1$). Regardless of their treatment assignment, always-takers and never-takers do not change their corresponding treatment condition, respectively. [Note: In our Figure 10, always-takers are displayed by the green colored area].

(4) Finally, we can categorize freelancers as ‘*defiers*’ when they are a) assigned to receive the premium membership subscription voucher, but do not actually redeem it ($d_i(1) = 0$), as well as when they are b) assigned to the control group, but voluntarily buy a premium membership subscription ($d_i(0) = 1$) [Note: Due to the monotonicity assumption, we assume no ‘defiers’ in our study, and therefore, there is no reason to visualize them in our Figure 10.]”

Figure 10. Graphical representation of CACE and ITT—as in T. Sun et al. (2019)



Notes: The semicolon rimmed box in the right corner (in the style of Lousdal (2018)) graphically details how the different groups (i.e., always-takers (green), never-takers (red), and compliers (blue)) respond to their assigned treatment condition (z). For the compliers, the group of interest within the CACE approach, we see that their treatment status mirrors exactly one-to-one their randomly assigned treatment condition. The other two groups, however, remain unresponsive to the corresponding treatment offer (Angrist & Pischke, 2008).

We created Figure 10 to briefly illustrate the technical construction of the CACE estimate. For this purpose, we show how the CACE can be manually estimated using two separate OLS regressions, as it demonstrates nicely the basic intuition of the framework. However, extreme caution must be exercised using this strategy in actual practice, as it can produce incorrect standard errors. Thus, researchers planning to estimate the CACE should use the genuine and expedient 2SLS-IV regression routines available in statistical analysis software tools (Angrist, 2006). Moreover, we also have to keep in mind that the identification of the CACE rests on a series of assumptions, as also addressed in Appendix A10.

Specifically, from a statistical point of view, we can regard the CACE estimate as the ratio of two components (i.e., ITT and share of compliers), whereby each one is the output of a specific OLS regression. To put this into perspective, first, we regress “treatment assigned (z_i)” on “treatment received (d_i)”, which ultimately reveals the share of compliers in the study. Second, we regress “treatment assigned (z_i)” on the outcome of interest “ Δ SocialCapitalAccumulation”, which then estimates the so-called ITT parameter. Finally, we divide both of the obtained coefficients to identify the $CACE = \frac{ITT}{Share\ of\ Compliers}$ (Gerber & Green, 2012).

A9 Experiment: Covariate Balance Check After Attrition

Table 15. Covariate Balance check after attrition

<i>Covariate</i>	<i>Treatment</i> <i>(assigned)</i>	<i>Mean (SD)</i>	<i>Difference</i>	<i>t-value</i>	<i>p-value</i>
<i>Gender (Female = 1; Male = 2)</i>	0	1.506 (0.503)	-0.160	-1.833	0.069
	1	1.667 (0.476)			
<i>Social Capital (Number of direct contacts)</i>	0	156.696 (150.184)	-20.970	-0.643	0.522
	1	177.667 (199.313)			
<i>Number of groups</i>	0	7.418 (9.401)	1.222	0.868	0.387
	1	6.196 (6.627)			
<i>SMN tenure (in years)</i>	0	6.975 (3.876)	-0.535	-0.776	0.440
	1	7.510 (3.818)			
<i>Number of 'haves'</i>	0	13.519 (12.047)	0.754	0.395	0.694
	1	12.765 (9.622)			
<i>Number of 'wants'</i>	0	5.127 (6.587)	1.264	1.347	0.180
	1	3.863 (4.109)			

A10 Experiment: First-Stage Estimate and IV Validation

As Weiler and Hinz (2019a, p. 228) elaborate

“The measures that we have presented [*Note*: Table 7 in the main text] allow us to verify the overall strength of our instrumental variables (IVs). Nonetheless, we can also separately evaluate each of our instruments in Model [E2] by using the Sanderson and Windmeijer (2016) first-stage underidentification and weak identification tests. We assess this statistic over the standard first-stage F-test of excluded instruments because the Sanderson and Windmeijer (2016) metric is required in cases where the model has more than one endogenous regressor, which is obviously the case in Model [E2]. Both conditional tests confirm that each of our IVs is relevant and strong, also if we examine them separately [*Note*: see Table 16].”

Table 16. First-stage estimates—as in Weiler and Hinz (2019a, p. 229)

	<i>Model E1. First-stage regression estimates: Treatment received</i>	<i>Model E2. First-stage regression estimates: Treatment received</i>	<i>Model E2. First-stage regression: Treatment received * Strategic network behavior</i>
Treatment assigned (IV)	0.891 (0.041)***	0.841 (0.107)***	-0.214 (0.235)
Treatment assigned *		0.021 (0.041)	0.990 (0.090)***
Strategic networking behavior (IV)			
Male (Ref.: female)	-0.054 (0.045)	-0.053 (0.045)	-0.133 (0.099)
High level of education (Ref.: middle)	0.044 (0.057)	0.047 (0.057)	0.069 (0.125)
Age (in years)	-0.000 (0.002)	-0.000 (0.002)	-0.000 (0.005)
Openness	0.022 (0.019)	0.020 (0.019)	0.048 (0.042)
SMN tenure (in years)	-0.000 (0.005)	-0.000 (0.005)	0.006 (0.012)
Strategic networking behavior	-0.021 (0.023)	-0.031 (0.030)	-0.027 (0.067)
Work hours per week	0.001 (0.001)	0.001 (0.001)	0.002 (0.003)
Other income (Ref.: no other income)	0.037 (0.042)	0.035 (0.042)	0.123 (0.092)
Usage frequency (once a month) (Ref.: less than once a month)	-0.029 (0.089)	-0.033 (0.089)	-0.033 (0.195)
Usage frequency (several times a month)	-0.005 (0.086)	-0.004 (0.087)	0.031 (0.189)
Usage frequency (once a week)	-0.085 (0.093)	-0.085 (0.093)	-0.206 (0.204)
Usage frequency (several times a week)	0.066 (0.096)	0.067 (0.096)	0.115 (0.210)
Usage frequency (once a day)	0.170 (0.096)	0.168 (0.096)	0.423 (0.210)*
Usage frequency (several times a day)	0.011 (0.115)	0.009 (0.116)	0.034 (0.253)
F-test of excluded instruments	473.08***	234.84***	347.99***
Underidentification test (Sanderson and Windmeijer 2016, first-stage chi-squared test)	640.62***	698.56***	1080.36***
Weak identification test (Sanderson and Windmeijer 2016, multivariate F-test of excluded instruments)	473.08***	510.49***	789.49***

Notes: Corresponding second-stage estimates are shown in Table 7

A11 Experiment: Discussion of Post-Treatment Bias

Given the central role of the agency variable “strategic networking behavior” in our study, we need to evaluate and discuss whether this variable suffers from post-treatment bias. As mentioned in 3.3.2, we measured this key variable only as part of our follow-up questionnaire. Thus, a potential concern is that the treatment might have influenced the strategic networking behavior of freelancers, and the inclusion of this post-treatment variable in our statistical model could have caused biased causal estimates (Montgomery, Nyhan, & Torres, 2018). However, as outlined in the following paragraphs via statistical and theoretical arguments, we are confident that our self-reported variable “strategic networking behavior” is not contingent upon the treatment. Therefore, it seems unlikely that post-treatment bias affected our results.

As a statistical check regarding the presence of this issue, we again ran a Welch’s t-test. Table 17 proves that there are no statistically significant differences between the treatment and control groups regarding our agency variable. In particular, the well-balanced groups

suggest that the treatment did not cause the treated freelancers to exhibit greater strategic networking behavior.

Table 17. Group comparison to assess post-treatment bias of strategic networking behavior

<i>Covariate</i>	<i>Treatment (assigned)</i>	<i>Mean (SD)</i>	<i>Difference</i>	<i>t-value</i>	<i>p-value</i>
<i>Strategic networking behavior</i>	0	2.316 (0.882)	-0.105	-0.629	0.531
	1	2.421 (0.959)			

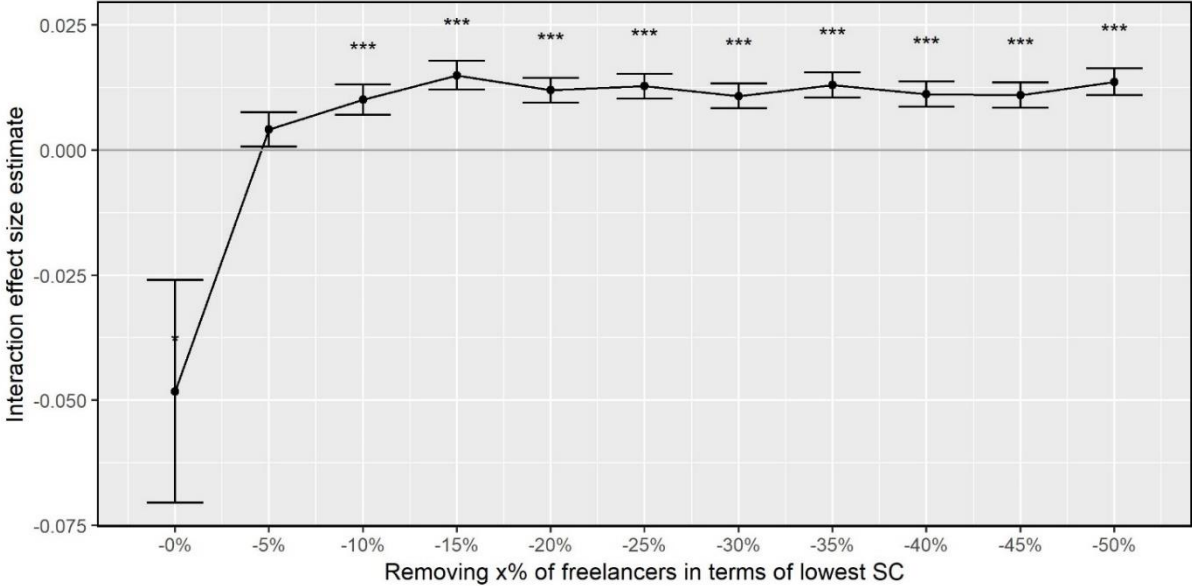
After ruling out potential group differences, theoretical arguments provided further support that the motivation to engage in a strategic networking behavior was not a consequence of the treatment. While we can equate the behavior of strategic networking with a personality trait (see Utz & Breuer, 2016), it is also plausible to assume that such intrinsic characteristics are more or less fixed over the course of time (Gertler, Martinez, Premand, Rawlings, & Vermeersch, 2016). For instance, a number of studies (e.g., Cobb-Clark & Schurer, 2012; Gustavsson, Weinryb, Göransson, Pedersen, & Åsberg, 1997) have demonstrated the stability of personality traits over several years. Against this backdrop, we can reasonably assume that, within the six months’ timeframe in which we conducted our experiment (i.e., between the manipulation and the follow-up questionnaire), it is quite unlikely that the motivation to engage in a strategic networking behavior was susceptible to change due to our issued treatment. We also found no statistically significant differences between the two groups on the other post-treatment covariates that were collected ($p > 0.05$), including openness, age, number of weekly working hours, frequency of WorkSMN usage, level of education, and other sources of income.

A12 Individual-Level Panel Data: Robustness of Percentage Change Computation

We expect that freelancers who pay for the premium membership will generally have a high agency for networking and thus affordances provided by WorkSMN. Hence, in percentage terms, the magnitude in SC accumulation is disproportionately higher for freelancers with just few contacts on WorkSMN. To address this potential distortion (in the form of outliers), we conduct a sensitivity analysis when repeating the OLS regression with percentage values for “ Δ Social capital accumulation” as dependent variable, i.e., “ Δ % Social capital accumulation”. More specifically, we follow the example of R. Bapna and Umyarov (2015) and apply a stepwise exclusion approach by percentile regarding the freelancers’ SC before the discount mailing. For each step, we assess the estimated effect size of the interaction term of “premium conversion * contact invites sent before”. The results in Figure

11 show that only after removing freelancers who are in the lowest 10-percentile, the estimates stabilize around a positive interaction of 0.012 - 0.014.

Figure 11. Interaction effect size estimates on „Δ % Social capital accumulation“



A13 Individual-Level Panel Data: Descriptive Statistics and Bivariate Correlations

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
[1] <i>SC Accumulation_{it}</i>										
[2] <i>Profile Visits (Log)_{it}</i>	0.273									
[3] <i>Messages Sent (Log)_{it}</i>	0.266	0.508								
[4] <i>Profile Visitors (Log)_{it}</i>	0.241	0.404	0.409							
[5] <i>Messages Rec. (Log)_{it}</i>	0.254	0.423	0.691	0.468						
[6] <i>Gender_i</i>	-0.019	-0.089	-0.036	-0.031	-0.042					
[7] <i>Educ_i</i>	0.011	0.042	0.064	0.082	0.067	0.022				
[8] <i>Age_i</i>	-0.046	-0.097	-0.020	-0.157	-0.081	-0.071	-0.021			
[9] <i>SMN tenure_i</i>	-0.033	0.011	0.012	0.039	0.010	-0.021	0.186	0.133		
[10] <i>JobSeeking_i</i>	0.022	0.037	0.039	0.060	0.043	0.029	0.015	-0.085	-0.148	
<i>Mean</i>	1.800	1.005	0.441	2.061	0.537	1.250	0.510	43.032	8.051	0.370
<i>SD</i>	8.086	1.095	0.702	0.930	0.738	0.433	0.500	9.684	2.442	0.483

Note: Significance level $p < 0.05$, as indicated by bold numbers. Observations over freelancer i and time t , as in Figure 6.

A14 Individual-Level Panel Data: Panel Vector Autoregression Model

Vector autoregressive (VAR) modeling allows us to assess the coevolution of multiple variables as interdependent system. In other words, VAR models do not impose a priori assumptions about the directionality of effects among the variables entered. Panel VAR, or PVAR, essentially is a generalization of this approach to panel data sets that contain time-series information on sets of heterogeneous units, i.e., freelancers in our setting. The nascent modeling approach has found recent application in IS literature (e.g., H. Chen et al., 2015; Dewan & Ramaprasad, 2014).

Because regular OLS estimation of PVAR models yields biased coefficients, most applications are estimated via generalized method of moments (GMM). GMM uses transformed observations as instruments for the lagged dependent variables (Hansen, 1982). An advantage of PVAR is that it captures unobserved differences of the microunits.

We assess the variables associated with SC accumulation, activity, and saliency. We take monthly first differences in SC accumulation and take natural logarithms to eradicate the positive trend and skewness. Monthly observations surrounding the discounting campaigns amount to 16 observations. In the first part of our data we report Harris-Tzavalis tests to assess for the stationarity of the data, reported in Table 18. Harris-Tzavalis is particularly suitable for short period, large cross-unit panel datasets, as in our case. All variables in the panel data set are stationary.

Table 18. Harris-Tzavalis test

<i>Variable</i>	<i>Test statistic</i>	<i>p-value</i>
Profile Visits	0.191	$p < 0.001$
Messages Sent	0.136	$p < 0.001$
Profile Visitors	0.139	$p < 0.001$
Messages Received	0.123	$p < 0.001$
SC Accumulation	0.028	$p < 0.001$

Specification

We specify a system GMM model with forward orthogonal deviations (FOD). FOD minimizes data loss, as it takes averaged future observations as instruments. For the sake of computational efficiency, we down-sample the 47,341 non-converters in equal proportion to converters and arrive at 1,430 observations. The presented results have been confirmed with

multiple random seeds. We look at the Andrews and Lu (2001) statistics to assess the optimal lag length and choose 3 lags, reflecting a full quarter—a reasonable duration for interdependencies between the variables to occur. As shown in Table 19, MAIC, MBIC, and MQIC values concur that a 3-lag specification is optimal.

Table 19. MAIC, MBIC, and MQIC (Andrews & Lu, 2001)

<i>Lags</i>	<i>MBIC</i>	<i>MAIC</i>	<i>MQIC</i>
1	-3,603	326	-1,074
2	-3,673	179	-1,198
3	-3,952	-115	-1,492

A15 Individual-Level Panel Data: Propensity Score Matching

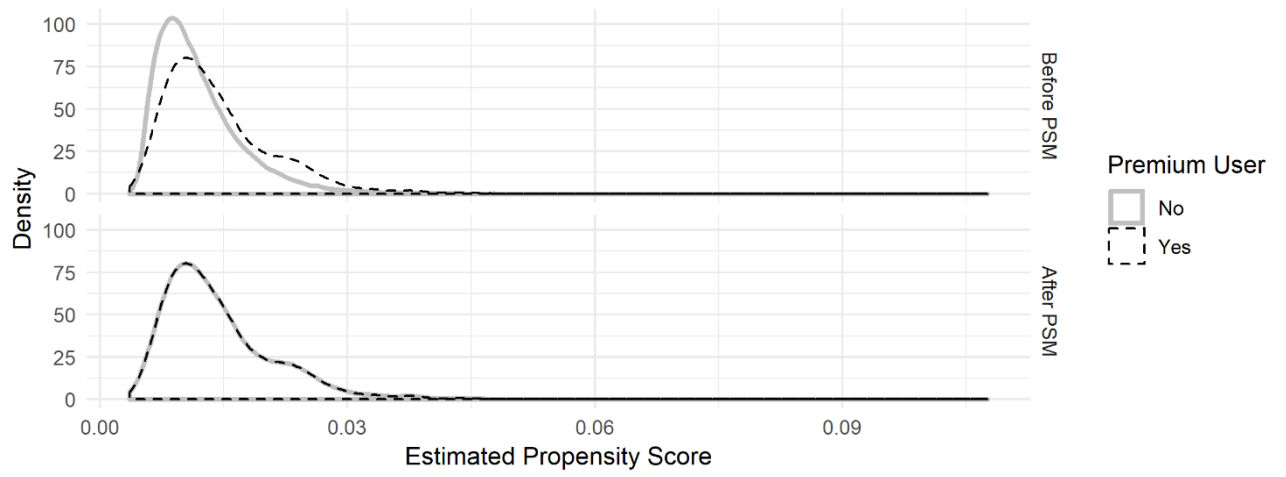
Table 20. Logistic regression of propensity to convert to premium

	<i>Premium conversion</i>
Contact invites sent <i>before</i> (Log)	0.045 (0.033)
Male (Ref.: female)	-0.036 (0.097)
Level of education (Ref.: No acad. title)	0.087 (0.083)
Age	0.001 (0.004)
SMN tenure	-0.004 (0.017)
Career segment: Manager (Ref.: Entry)	0.004 (0.124)
Career segment: Senior Executive	0.168 (0.092)
Job seeking: Active	0.434 (0.082) ***
Logins <i>before</i> (Log)	0.098 (0.042) *
Messages sent <i>before</i> (Log)	0.013 (0.074)
Profile visits <i>before</i> (Log)	0.130 (0.045) **
Messages rec. <i>before</i> (Log)	-0.037 (0.071)
Profile visitors rec. <i>before</i> (Log)	0.101 (0.058)
AIC	6,709.561
Nagelkerke R ²	0.017
n	52,392

Notes: Constant not reported. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Following the propensity score matching (PSM), we evaluate the match via standardized mean differences (SMD). Given the specification with ratio 1, the focal variables SMD reduce from 0.232 to 0.013 for Job seeking, from 0.180 to 0.041 for Logins before (Log), and from 0.275 to 0.005 for Profile Visits before (Log). All matching variables in the matched groups appear below the 0.1 threshold to conclude that matching was successful. The distributions of the propensity scores after matching appear close to identical. We obtain similar results for matching with ratio 4.

Figure 12. Distribution of propensity scores



7 Bibliography

- Adler, P. S., & Kwon, S.-W. (2002). Social capital: Prospects for a new concept. *Academy of Management Review*, 27(1), 17-40.
- Ahuja, G., Soda, G., & Zaheer, A. (2012). The genesis and dynamics of organizational networks. *Organization Science*, 23(2), 434-448.
- Andrews, D. W., & Lu, B. (2001). Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models. *Journal of Econometrics*, 101(1), 123-164.
- Angrist, J. D. (2006). Instrumental variables methods in experimental criminological research: what, why and how. *Journal of Experimental Criminology*, 2(1), 23-44.
- Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*: Princeton university press.
- Aral, S., & Walker, D. (2014). Tie strength, embeddedness, and social influence: A large-scale networked experiment. *Management Science*, 60(6), 1352-1370.
- Arnaboldi, V., Guazzini, A., & Passarella, A. (2013). Egocentric online social networks: Analysis of key features and prediction of tie strength in Facebook. *Computer Communications*, 36(10), 1130-1144.
- Aten, K., DiRenzo, M., & Shatnawi, D. (2017). Gender and professional e-networks: Implications of gender heterophily on job search facilitation and outcomes. *Computers in Human Behavior*, 72, 470-478.
- Backstrom, L., & Kleinberg, J. (2014). *Romantic partnerships and the dispersion of social ties: a network analysis of relationship status on facebook*. Paper presented at the Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing.
- Bakshy, E., Eckles, D., Yan, R., & Rosenn, I. (2012). *Social influence in social advertising: evidence from field experiments*. Paper presented at the Proceedings of the 13th ACM Conference on Electronic Commerce.
- Banerjee, S., & Bhardwaj, P. (2019). Aligning marketing and sales in multi-channel marketing: Compensation design for online lead generation and offline sales conversion. *Journal of Business Research*, 105, 293–305.
doi:10.1016/j.jbusres.2019.06.016
- Bapna, R., Gupta, A., Rice, S., & Sundararajan, A. (2017). Trust and the Strength of Ties in Online Social Networks: An Exploratory Field Experiment. *MIS Quarterly*, 41(1), 115-130.
- Bapna, R., Ramaprasad, J., Shmueli, G., & Umyarov, A. (2016). One-way mirrors in online dating: A randomized field experiment. *Management Science*, 62(11), 3100-3122.
- Bapna, R., Ramaprasad, J., & Umyarov, A. (2018). Monetizing freemium communities: Does paying for premium increase social engagement? *MIS Quarterly*, 42(3), 719-735.
- Bapna, R., & Umyarov, A. (2015). Do your online friends make you pay? A randomized field experiment on peer influence in online social networks. *Management Science*, 61(8), 1902-1920.

- Bapna, S., & Funk, R. (2021). Interventions for Improving Professional Networking for Women: Experimental Evidence From the IT Sector. *MIS Quarterly*, 45(2), 593-636.
- Bensaou, B. M., Galunic, C., & Jonczyk-Sédès, C. (2014). Players and purists: Networking strategies and agency of service professionals. *Organization Science*, 25(1), 29-56.
- Bernhard, R. (2013). *Studying Facebook via Data Extraction: The Netvizz Application*. Paper presented at the WebSci '13 Proceedings of the 5th Annual ACM Web Science Conference, New York: ACM.
- Bollen, K. A. (2012). Instrumental variables in sociology and the social sciences. *Annual Review of Sociology*, 38, 37-72.
- Borgatti, S. P., & Halgin, D. S. (2011). On network theory. *Organization Science*, 22(5), 1168-1181.
- Borgatti, S. P., Jones, C., & Everett, M. G. (1998). Network measures of social capital. *Connections*, 21(2), 27-36.
- Brown, J. J., & Reingen, P. H. (1987). Social ties and word-of-mouth referral behavior. *Journal of Consumer research*, 14(3), 350-362.
- Brynjolfsson, E., Collis, A., & Eggers, F. (2019). Using massive online choice experiments to measure changes in well-being. *Proceedings of the National Academy of Sciences*, 116(15), 7250-7255.
- Bucher, T., & Helmond, A. (2017). The affordances of social media platforms. *The SAGE Handbook of Social Media*, 233-253.
- Cadwalladr, C., & Graham-Harrison, E. (2018). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian*.
- Casciaro, T., Gino, F., & Kouchaki, M. (2014). The contaminating effects of building instrumental ties: How networking can make us feel dirty. *Administrative Science Quarterly*, 59(4), 705-735.
- Chen, H., De, P., & Hu, Y. J. (2015). IT-enabled broadcasting in social media: An empirical study of artists' activities and music sales. *Information Systems Research*, 26(3), 513-531.
- Chen, X., van der Lans, R., & Phan, T. Q. (2017). Uncovering the importance of relationship characteristics in social networks: Implications for seeding strategies. *Journal of Marketing Research (JMR)*, 54(2), 187-201.
- Chu, S.-C., & Kim, Y. (2011). Determinants of consumer engagement in electronic word-of-mouth (eWOM) in social networking sites. *International journal of Advertising*, 30(1), 47-75.
- Cobb-Clark, D. A., & Schurer, S. (2012). The stability of big-five personality traits. *Economics Letters*, 115(1), 11-15.
- Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695(5), 1-9.
- Dewan, S., & Ramaprasad, J. (2014). Social media, traditional media, and music sales. *MIS Quarterly*, 38(1), 101-122.
- Dunbar, R. I. (1993). Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences*, 16(4), 681-694.

- Ebbes, P., & Netzer, O. (2018). *Using Social Network Activity Data to Identify and Target Job Seekers*. SSRN (3200214).
- Ebbes, P., Papies, D., & Van Heerde, H. J. (2011). The sense and non-sense of holdout sample validation in the presence of endogeneity. *Marketing Science*, 30(6), 1115-1122.
- Ebbes, P., Papies, D., & van Heerde, H. J. (2016). Dealing with Endogeneity: A Nontechnical Guide for Marketing Researchers. In C. Homburg, M. Klarmann, & A. Vomberg (Eds.), *Handbook of Market Research*: Springer International Publishing AG.
- Eckles, D., & Bakshy, E. (2020). Bias and high-dimensional adjustment in observational studies of peer effects. *Journal of the American Statistical Association*, 1-11.
- Elkington, D. (2013). Lead Response Management Study. Retrieved from http://www.leadresponsemanagement.org/lrm_study
- Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook “friends:” Social capital and college students’ use of online social network sites. *Journal of computer-mediated communication*, 12(4), 1143-1168.
- eMarketer. (2020). The Global Media Intelligence Report. Retrieved from <https://www.globalwebindex.com/reports/gmi-report>
- Everett, M., & Borgatti, S. P. (2005). Ego network betweenness. *Social Networks*, 27(1), 31-38.
- Fawcett, T. (2004). ROC graphs: Notes and practical considerations for researchers. *Machine Learning*, 31(1), 1-38.
- Feld, S. L. (1981). The focused organization of social ties. *American Journal of Sociology*, 86(5), 1015-1035.
- Forret, M. L. (2018). Networking as a job-search behavior and career management strategy. *The Oxford Handbook of Job Loss and Job Search*, 275.
- Forret, M. L., & Dougherty, T. W. (2001). Correlates of networking behavior for managerial and professional employees. *Group & Organization Management*, 26(3), 283-311.
- Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, 1(3), 215-239.
- Friedkin, N. (1990). A Guttman scale for the strength of an interpersonal tie. *Social Networks*, 12(3), 239-252.
- Gandini, A. (2016). Digital work: Self-branding and social capital in the freelance knowledge economy. *Marketing Theory*, 16(1), 123-141. doi:10.1177/1470593115607942
- Garg, R., & Telang, R. (2018). To be or not to be linked: Online social networks and job search by unemployed workforce. *Management Science*, 64(8), 3926-3941.
- Gee, L. K. (2019). The more you know: information effects on job application rates in a large field experiment. *Management Science*, 65(5), 2077-2094.
- Gee, L. K., Jones, J., & Burke, M. (2017). Social networks and labor markets: How strong ties relate to job finding on Facebook’s social network. *Journal of Labor Economics*, 35(2), 485-518.
- Gerber, A. S., & Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation*: W. W. Norton & Company.

- Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. (2016). *Impact evaluation in practice* (Second Edition ed.): The World Bank.
- Gilbert, E., & Karahalios, K. (2009). *Predicting tie strength with social media*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.
- Gilbert, E., & Karahalios, K. (2015). United States of America Patent No.: U. S. Patent.
- Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6), 1360-1380.
- Gustavsson, J. P., Weinryb, R. M., Göransson, S., Pedersen, N. L., & Åsberg, M. (1997). Stability and predictive ability of personality traits across 9 years. *Personality and individual differences*, 22(6), 783-791.
- Hahn, E., Gottschling, J., & Spinath, F. M. (2012). Short measurements of personality—Validity and reliability of the GSOEP Big Five Inventory (BFI-S). *Journal of Research in Personality*, 46(3), 355-359.
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4), 1029-1054.
- Harris, R. D., & Tzavalis, E. (1999). Inference for unit roots in dynamic panels where the time dimension is fixed. *Journal of Econometrics*, 91(2), 201-226.
- Hayes, J. L., King, K. W., & Ramirez Jr, A. (2016). Brands, friends, & viral advertising: A social exchange perspective on the ad referral processes. *Journal of Interactive Marketing*, 36, 31-45.
- Haynes, J., & Marshall, L. (2018). Beats and tweets: Social media in the careers of independent musicians. *New Media & Society*, 20(5), 1973-1993.
- Hein, S., Schlereth, C., & Mueller-Klockmann, T. (2019). Long-Term Brand Equity Measurement - Status Quo and Challenges. *Transfer*, 3, 6-11.
- Hildebrand, C., & Bergner, A. (2020). Conversational robo advisors as surrogates of trust: onboarding experience, firm perception, and consumer financial decision making. *Journal of the Academy of Marketing Science*, 1-18.
- Hinz, O., Hill, S., & Kim, J.-Y. (2016). TV's Dirty Little Secret: The Negative Effect of Popular TV on Online Auction Sales. *MIS Quarterly*, 40(3), 623-644.
- Hinz, O., Skiera, B., Barrot, C., & Becker, J. U. (2011). Seeding strategies for viral marketing: An empirical comparison. *Journal of Marketing*, 75(6), 55-71.
- Hinz, O., Spann, M., & Hann, I.-H. (2015). Research note—can't buy me love... or can I? Social capital attainment through conspicuous consumption in virtual environments. *Information Systems Research*, 26(4), 859-870.
- Hoelzer, J. (2019). *The Impact of Different Tie Strengths on the Effectiveness of Social Media Marketing*. (Bachelor of Science Bachelor Thesis), WHU - Otto Beisheim School of Management,
- Hoisl, K., & Mariani, M. (2017). It's a man's job: Income and the gender gap in industrial research. *Management Science*, 63(3), 766-790.
- Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*: Cambridge University Press.

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112): Springer.
- Johnston, M. W., & Marshall, G. W. (2016). *Sales Force Management: Leadership, Innovation, Technology* (12th ed.): Routledge.
- Jones, J. J., Settle, J. E., Bond, R. M., Fariss, C. J., Marlow, C., & Fowler, J. H. (2013). Inferring tie strength from online directed behavior. *PLOS ONE*, 8(1), e52168.
- Kahanda, I., & Neville, J. (2009). Using Transactional Information to Predict Link Strength in Online Social Networks. *ICWSM*, 9, 74-81.
- Kane, G., Alavi, M., Labianca, G., & Borgatti, S. (2014). What's different about social media networks? A framework and research agenda. *MIS Quarterly*, 38(1), 275-304.
- Kannan, P. K., & Li, H. A. (2017). Digital marketing: A framework, review and research agenda. *International Journal of Research in Marketing*, 34(1), 22-45. doi:10.1016/j.ijresmar.2016.11.006
- Karahanna, E., Xu, S. X., Xu, Y., & Zhang, N. A. (2018). The needs-affordances-features perspective for the use of social media. *MIS Quarterly*, 42(3), 737-756.
- Katona, Z., Zubcsek, P. P., & Sarvary, M. (2011). Network effects and personal influences: The diffusion of an online social network. *Journal of Marketing Research*, 48(3), 425-443.
- King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis*, 9(2), 137-163.
- Koch, O. F., & Benlian, A. (2017). The effect of free sampling strategies on freemium conversion rates. *Electronic Markets*, 27(1), 67-76.
- Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15), 5802-5805.
- Kuhn, K. M. (2016). The Rise of the "Gig Economy" and Implications for Understanding Work and Workers. *Industrial and Organizational Psychology*, 9(1), 157.
- Kuwabara, K., Hildebrand, C. A., & Zou, X. (2018). Lay theories of networking: How laypeople's beliefs about networks affect their attitudes toward and engagement in instrumental networking. *Academy of Management Review*, 43(1), 50-64.
- Lanz, A., Goldenberg, J., Shapira, D., & Stahl, F. (2019). Climb or jump: status-based seeding in user-generated content networks. *Journal of Marketing Research*, 56(3), 361-378.
- Laurie, H., & Lynn, P. (2009). *The use of respondent incentives on longitudinal surveys*: Wiley Online Library.
- Lee, J. J. (2010). Heterogeneity, brokerage, and innovative performance: Endogenous formation of collaborative inventor networks. *Organization Science*, 21(4), 804-822.
- Leonardi, P. M. (2013). When does technology use enable network change in organizations? A comparative study of feature use and shared affordances. *MIS Quarterly*, 749-775.
- Leskovec, J., & McAuley, J. J. (2012). *Learning to discover social circles in ego networks*. Paper presented at the Advances in Neural Information Processing Systems.

- Leskovec, J., & Sosič, R. (2016). SNAP: A general-purpose network analysis and graph-mining library. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 8(1), 1-20.
- Leung, M. D. (2014). Dilettante or renaissance person? How the order of job experiences affects hiring in an external labor market. *American Sociological Review*, 79(1), 136-158.
- Li, Z., Fang, X., Bai, X., & Sheng, O. R. L. (2017). Utility-based link recommendation for online social networks. *Management Science*, 63(6), 1938-1952.
- Liben-Nowell, D., & Kleinberg, J. (2007). The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology*, 58(7), 1019-1031.
- Lin, N. (1999). Building a network theory of social capital. *Connections*, 22(1), 28-51.
- Lo Presti, A., Pluviano, S., & Briscoe, J. P. (2018). Are freelancers a breed apart? The role of protean and boundaryless career attitudes in employability and career success. *Human Resource Management Journal*, 28(3), 427-442.
- Lonati, S., Quiroga, B. F., Zehnder, C., & Antonakis, J. (2018). On doing relevant and rigorous experiments: Review and recommendations. *Journal of Operations Management*, 64, 19-40.
- Lousdal, M. L. (2018). An introduction to instrumental variable assumptions, validation and estimation. *Emerging themes in epidemiology*, 15(1), 1-7.
- Malev, M. (2020). Is the coronavirus pandemic an engine for the digital transformation? Retrieved from <https://dmexco.com/stories/is-the-coronavirus-pandemic-an-engine-for-the-digital-transformation/>
- Manago, A. M., Taylor, T., & Greenfield, P. M. (2012). Me and my 400 friends: The anatomy of college students' Facebook networks, their communication patterns, and well-being. *Developmental Psychology*, 48(2), 369.
- Marsden, P. V., & Campbell, K. E. (1984). Measuring tie strength. *Social Forces*, 63(2), 482-501.
- Mayring, P. (2010). Qualitative Inhaltsanalyse. In G. Mey & K. Mruck (Eds.), *Handbuch Qualitative Forschung in der Psychologie* (pp. 601–613). Wiesbaden: VS Verlag für Sozialwissenschaften.
- McCarthy, J. E., & Levin, D. Z. (2019). Network residues: The enduring impact of intra-organizational dormant ties. *Journal of Applied Psychology*.
- McGlaughlin, F., Doyle, J., & Balegno, S. (2012). Benchmark Report - 2012 Lead Generation: Key industry trends for generating leads of the highest quality. Retrieved from https://content.marketingsherpa.com/data/public/reports/benchmark-reports/BMR-Lead_Generation.pdf
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1), 415-444.
- Moe, W. W., & Ratchford, B. T. (2018). How the explosion of customer data has redefined interactive marketing. *Journal of Interactive Marketing*, 42, A1-A2.

- Montgomery, J. M., Nyhan, B., & Torres, M. (2018). How conditioning on posttreatment variables can ruin your experiment and what to do about it. *American Journal of Political Science*, 62(3), 760-775.
- Nahapiet, J., & Ghoshal, S. (1998). Social capital, intellectual capital, and the organizational advantage. *Academy of Management Review*, 23(2), 242-266.
- NetworkX. (2019). NetworkX - Software for Complex Networks. Retrieved from <https://networkx.github.io/>
- Nielsen. (2016). What are connected shoppers doing and not doing online? Retrieved from <https://www.nielsen.com/eu/en/insights/article/2016/what-are-connected-shoppers-doing-and-not-doing-online/>
- Nikitkov, A., & Sainty, B. (2014). The role of social media in influencing career success. *International Journal of Accounting & Information Management*, 22(4), 273-294.
- Oestreicher-Singer, G., & Zalmanson, L. (2013). Content or community? A digital business strategy for content providers in the social age. *MIS Quarterly*, 37(2), 591-616.
- Oldroyd, J., Mcelheran, K., & Elkington, D. (2011). The short life of online sales leads. *Harvard Business Review*, 89(3), 28.
- Patterson, L. (2007). Marketing and sales alignment for improved effectiveness. *Journal of Digital Asset Management*, 3(4), 185–189. doi:10.1057/palgrave.dam.3650089
- Porter, C. M., & Woo, S. E. (2015). Untangling the networking phenomenon: A dynamic psychological perspective on how and why people network. *Journal of Management*, 41(5), 1477-1500.
- Portes, A. (1998). Social capital: Its origins and applications in modern sociology. *Annual Review of Sociology*, 24(1), 1-24.
- Putnam, R. D. (2000). Bowling alone: America's declining social capital. In *Culture and politics* (pp. 223-234): Springer.
- Reingen, P. H., & Kernan, J. B. (1986). Analysis of referral networks in marketing: Methods and illustration. *Journal of Marketing Research*, 23(4), 370-378.
- Rishika, R., & Ramaprasad, J. (2019). The Effects of Asymmetric Social Ties, Structural Embeddedness, and Tie Strength on Online Content Contribution Behavior. *Management Science*, 65(7).
- Rosenbaum, P. R. (2020). Modern algorithms for matching in observational studies. *Annual Review of Statistics and Its Application*, 7, 143-176.
- Rostila, M. (2011). The facets of social capital. *Journal for the Theory of Social Behaviour*, 41(3), 308-326.
- Rotabi, R., Kamath, K., Kleinberg, J., & Sharma, A. (2017). *Detecting strong ties using network motifs*. Paper presented at the Proceedings of the 26th International Conference on World Wide Web Companion.
- Sagarin, B. J., West, S. G., Ratnikov, A., Homan, W. K., Ritchie, T. D., & Hansen, E. J. (2014). Treatment noncompliance in randomized experiments: Statistical approaches and design issues. *Psychological Methods*, 19(3), 317-333.
- Sanderson, E., & Windmeijer, F. (2016). A weak instrument F-test in linear IV models with multiple endogenous variables. *Journal of Econometrics*, 190(2), 212-221.

- Säuberlich, F., Smith, K., & Yuhn, M. (2005). Analytical Lead Management in the Automotive Industry. In D. Baier, R. Decker, & L. Schmidt-Thieme (Eds.), *Data Analysis and Decision Support* (pp. 290–299). Berlin/Heidelberg: Springer-Verlag.
- Schlereth, C., Barrot, C., Skiera, B., & Takac, C. (2013). Optimal Product-Sampling Strategies in Social Networks: How Many and Whom to Target? *International Journal of Electronic Commerce*, 18(1), 45-72.
- Schlereth, C., & Skiera, B. (2012). Measurement of consumer preferences for bucket pricing plans with different service attributes. *International Journal of Research in Marketing*, 29(2), 167-180.
- Seibert, S. E., Kraimer, M. L., & Liden, R. C. (2001). A social capital theory of career success. *Academy of Management Journal*, 44(2), 219-237.
- Shapiro, C., & Varian, H. R. (1999). *Information rules: a strategic guide to the network economy*: Harvard Business Press.
- Shevchuk, A., & Strebkov, D. (2018). Safeguards against opportunism in freelance contracting on the Internet. *British Journal of Industrial Relations*, 56(2), 342-369.
- Shmargad, Y., & Watts, J. K. (2016). When online visibility deters social interaction: The case of digital gifts. *Journal of Interactive Marketing*, 36, 1-14.
- Shmueli, G. (2010). To explain or to predict? *Statistical Science*, 25(3), 289-310.
- Shmueli, G., & Koppius, O. R. (2011). Predictive analytics in information systems research. *MIS Quarterly*, 35(3), 553-572.
- Shriver, S. K., Nair, H. S., & Hofstetter, R. (2013). Social ties and user-generated content: Evidence from an online social network. *Management Science*, 59(6), 1425-1443.
- Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1), 99-118.
- Smith, T. M., Gopalakrishna, S., & Chatterjee, R. (2006). A three-stage model of integrated marketing communications at the marketing-sales interface. *Journal of Marketing Research*, 43(4), 564-579.
- Sorge, A., & Hofstede, G. (1983). Culture's Consequences: International Differences in Work-Related Values. *Administrative Science Quarterly*, 28(4), 625.
doi:10.2307/2393017
- Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In D. W. Andrews & J. H. Stock (Eds.), *Identification and Inference for Econometric Models* (pp. 80-108). New York: National Bureau of Economic Research.
- Stolz, S. (2014). *About the Effects of Network Configuration Measures on Information Exchange in Online Social Networks*. (Master of Science Master Thesis), Maastricht University.
- Stolz, S., & Schlereth, C. (2020). *Predicting Tie Strength with Ego Network Structures*. Retrieved from: <https://data.mendeley.com/datasets/hr9tjzj72v/>
- Stolz, S., & Schlereth, C. (2021). Predicting Tie Strength with Ego Network Structures. *Journal of Interactive Marketing*, 54, 40-52.
- Stolz, S., Schlereth, C., & Lanz, A. (2020). *How Conversions in Freemium Businesses Impact User Activity*. Paper presented at the 42nd INFORMS Marketing Science Conference, Duke University, Fuqua, America (Online Session).

- Stolz, S., Wisskirchen, K., Schlereth, C., & Hoffmann, A. (2021). The Emerging Business of Online Lead Generation. *Marketing Review St. Gallen*, 4, 32-39.
- Strong, E. K. (1925). *The psychology of selling and advertising*: McGraw-Hill book Company, Incorporated.
- Stuart, T. E., & Sorenson, O. (2007). Strategic networks and entrepreneurial ventures. *Strategic Entrepreneurship Journal*, 1(3-4), 211-227.
- Sun, C., Shi, Z. J., Liu, X., Ghose, A., Li, X., & Xiong, F. (2019). The Effect of Voice AI on Consumer Purchase and Search Behavior. *SSRN (3480877)*, Working Paper.
- Sun, T., Shi, L., Viswanathan, S., & Zheleva, E. (2019). Motivating effective mobile app adoptions: Evidence from a large-scale randomized field experiment. *Information Systems Research*, 30(2), 523-539.
- Sun, T., & Taylor, S. J. (2020). Displaying things in common to encourage friendship formation: A large randomized field experiment. *Quantitative Marketing and Economics, Not assigned to an issue*, 1-35. doi:<https://doi.org/10.1007/s11129-020-09224-9>
- Sundararajan, A., Provost, F., Oestreicher-Singer, G., & Aral, S. (2013). Research commentary—information in digital, economic, and social networks. *Information Systems Research*, 24(4), 883-905.
- Takac, C., Hinz, O., & Spann, M. (2011). The social embeddedness of decision making: opportunities and challenges. *Electronic Markets*, 21(3), 185.
- Toubia, O., & Stephen, A. T. (2013). Intrinsic vs. image-related utility in social media: Why do people contribute content to twitter? *Marketing Science*, 32(3), 368-392.
- Tröster, C., Parker, A., Van Knippenberg, D., & Sahlmüller, B. (2019). The coevolution of social networks and thoughts of quitting. *Academy of Management Journal*, 62(1), 22-43.
- Trusov, M., Bodapati, A. V., & Bucklin, R. E. (2010). Determining influential users in internet social networks. *Journal of Marketing Research*, 47(4), 643-658.
- Utz, S. (2016). Is LinkedIn making you more successful? The informational benefits derived from public social media. *New Media & Society*, 18(11), 2685-2702.
- Utz, S., & Breuer, J. (2016). Informational benefits from social media use for professional purposes: Results from a longitudinal study. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 10(4).
- Van den Born, A., & Van Witteloostuijn, A. (2013). Drivers of freelance career success. *Journal of Organizational Behavior*, 34(1), 24-46.
- Van Hove, G., Van Hooft, E. A., & Lievens, F. (2009). Networking as a job search behaviour: A social network perspective. *Journal of Occupational and Organizational Psychology*, 82(3), 661-682.
- Vissa, B. (2012). Agency in action: Entrepreneurs' networking style and initiation of economic exchange. *Organization Science*, 23(2), 492-510.
- Viswanath, B., Mislove, A., Cha, M., & Gummadi, K. P. (2009). *On the evolution of user interaction in facebook*. Paper presented at the Proceedings of the 2nd ACM Workshop on Online Social Networks.

- Voigt, S., & Hinz, O. (2016). Making digital freemium business models a success: Predicting customers' lifetime value via initial purchase information. *Business & Information Systems Engineering*, 58(2), 107-118.
- Wasko, M. M., & Faraj, S. (2005). Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly*, 29(1), 35-57.
- Weiler, M., Stolz, S., Lanz, A., Schlereth, C. & Hinz, O. (2022). Social Capital Accumulation Through Social Media Networks: Evidence from a Randomized Field Experiment and Individual-Level Panel Data. *MIS Quarterly*, *Forthcoming*.
- Weiler, M., & Hinz, O. (2019a). Social Capital Accumulation through Social Media Networks and Its Benefits in a Project-based Labor Market: Evidence from a Randomized Online Field Experiment. In *The Value and Use of Accumulating Social Capital: New Insights from Social Media Networks (Dissertation)* (pp. 185-239). Frankfurt, Germany: Goethe University.
- Weiler, M., & Hinz, O. (2019b). Without each other, we have nothing: a state-of-the-art analysis on how to operationalize social capital. *Review of Managerial Science*, 13(5), 1003-1035.
- Wiese, J., Min, J.-K., Hong, J. I., & Zimmerman, J. (2015). *You never call, you never write: Call and sms logs do not always indicate tie strength*. Paper presented at the Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing.
- Wisskirchen, K. (2020). *Online Lead Generation Companies: A Segmentation*. (Master of Science), WHU - Otto Beisheim School of Management.
- Wolff, H.-G., & Moser, K. (2009). Effects of networking on career success: A longitudinal study. *Journal of Applied Psychology*, 94(1), 196.
- Wu, L. (2013). Social network effects on productivity and job security: Evidence from the adoption of a social networking tool. *Information Systems Research*, 24(1), 30-51.
- Xant.ai. (2019). The truth behind successful outbound sales cadences. Retrieved from <https://www.xant.ai/resources/the-truth-behind-successful-outbound-sales-cadences/>
- Xiang, R., Neville, J., & Rogati, M. (2010). *Modeling relationship strength in online social networks*. Paper presented at the Proceedings of the 19th International Conference on World Wide Web.
- Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, 112(4), 1036-1040.

