Integrating micro-level interactions with social network analysis in tie strength research: the edge-centered approach

Full Paper

O. Torro
Tampere University of Technology
Korkeakoulunkatu 8, 33720, Tampere
Finland
osku.torro@tut.fi

J. P. Gupta
Tampere University of Technology
Korkeakoulunkatu 8, 33720, Tampere
Finland
jayesh.gupta@tut.fi

H. Kärkkäinen
Tampere University of Technology
Korkeakoulunkatu 8, 33720, Tampere
Finland
hannu.karkkainen@tut.fi

H. Pirkkalainen
Tampere University of Technology
Korkeakoulunkatu 8, 33720, Tampere
Finland
henri.pirkkalainen@tut.fi

R. Vatrapu
Copenhagen Business School
Howitzvej 60
2000 Frederiksberg, Denmark
rv.digi@cbs.dk

R. R. Mukkamala
Copenhagen Business School
Howitzvej 60
2000 Frederiksberg, Denmark
rrm.digi@cbs.dk

A. Hussain
Copenhagen Business School
Howitzvej 60
2000 Frederiksberg, Denmark
ah.digi@cbs.dk

ABSTRACT

A social tie is a target for ongoing, high-level scientific debate. Measuring the tie strength in social networks has been an important topic for academic studies since Mark Granovetter's seminal papers in 1970's. However, it is still a problematic issue mainly for two reasons: 1) existing tie strength measurements may not reflect the true social connections of individuals accurately enough, and 2) many different methods to gather data from social media are not applicable anymore due to different data openness issues. In addition, we have only little empirical knowledge of the actual tie strengthening process in online social networks. Therefore, we suggest a new approach to tie strength research, which focuses on studying communication patterns (edges) more rather than actors (nodes) in a social network.

In this paper we build a social network analysis-based approach to enable the evaluation of tie strength based on reciprocal interaction from publicly available Facebook data, and suggest that this approach could work as a basis for further tie strength studies. Our approach makes use of weak tie theory, and enables researchers to study micro-level interactions (i.e. discussions, messages, relationships) with large-scale social network analysis (SNA). This study provides a way to find relevant actors from publicly available data in the context of tie strengthening process, and answers how to take this stream of research closer to computational social science.

AcademicMindtrek'17, September 20–21, 2017, Tampere, Finland

ACM ISBN 978-1-4503-5426-4/17/09…$15.00
https://doi.org/10.1145/3131085.3131095

CCS CONCEPTS

• **Human-centered computing** → Collaborative and social computing theory, concepts and paradigms

KEYWORDS

Social network, social tie, social network analysis, tie strength, open data

1 INTRODUCTION

Many organizations struggle with IT-enabled work whilst the online communication environment differs from face-to-face interaction. For example, individuals who do not know each other well are expected to work seamlessly together [1], [2]. To overcome these obstacles, we need a refined understanding about human interaction patterns, and more specifically, we need to understand the communication process that leads to tie strengthening, i.e. people getting to know each other better [3]. Tie strength research addresses these challenges in a specific way.

However, some of the leading scholars have argued that we lack understanding of how social networks (communities, organizations, society) evolve [4]. In this discussion, the theory of social ties (e.g. “strong and weak ties”) [5] has a crucial role as it explains, for example, how information is transferred and adopted in a network. However, Granovetter’s [5] theory has not yet reached its full potential, because it lacks integration to study online social networks in detailed level. Therefore, also new approaches to study online social networks are needed [6].

In this paper, we suggest a new approach to tie strength research that integrates publicly available micro-level data with large scale social network analysis (SNA). We believe that it is critical to study the actual reciprocal communication patterns of the individuals that could inform us on social ties in the network. Thus, we argue that the edge-centered approach, that studies “edges” (messages, conversations) instead of “nodes” (actors) in a social network, is important for two main reasons: 1) This approach would allow researchers to gain insights from the actual content of interactions from publicly available data, without compromising legal or privacy issues; 2) the approach complements methods which use electronic (likes, links shared, etc.), or self-reported data (questionnaires, etc.), which may be limited and inaccurate when used solely [4].

Our research question is as follows: How to find relevant actors (nodes) and their mutual interactions (edges) from publicly available large social media data set in the context of tie strengthening process? To answer this research question and to demonstrate the edge-centered approach, we build a network of reciprocal interaction from Facebook data. We also describe specific set of network/data properties that help with possible generalization issues (related to e.g. tie strengthening in knowledge work). We believe that this approach helps researchers to get closer towards *computational social science* [7], and mitigate some of the problems with big data paradox and noise removal fallacy [8].

We begin our paper with a section that describes the concept and importance of social ties, how tie strength has previously been measured, and why the edge-centered approach to tie strength research is different. The next section describes a practical example of our approach; how to build a network of reciprocal interaction from publicly available Facebook data. Finally, in the last section, we discuss the contribution of our study, and provide important examples for future studies.

2 SOCIAL TIE

2.1 Concept and importance

“The network driven nature of the world” [9] has drawn the attention of scholars for a long time, especially since the rise of the Internet and big data [8]. Granovetter’s [5] theory of strong and weak ties has offered a solid base for this discussion [6]; today his original article “The Strength of Weak Ties” has more than 44.000 citations in Google Scholar database (Google Scholar).

The basic concept of social ties (strong or weak) ties is quite simple: Strong ties refer to people that are close to us, or people with whom we spend a lot of time, like colleagues. Weak ties are active only a short period of time and formed mainly between acquaintances or previously unknown individuals. Granovetter [5] highlighted the importance of weak ties, because new information (ideas, insights, rumors, etc.) is transferred to the network mainly through weak ties. Strong ties tend to decide if this new information is finally adopted or not. [5]

Strong ties also tend to form high-cohesion, mutually connected networks [5], [10], [11] consisting highly redundant knowledge that is both explicit and tacit [11], [12]. Therefore, strong ties are considered essential when dealing with complex information [13]. Strong ties also correlate with the amount of useful information obtained from social network [14], and motivation to share knowledge [15]. Besides more effective information transfer, stronger ties are closely related to the amount of organizational support [16], [17], organizational commitment [18], and leveraging innovation capabilities [19].

As Granovetter’s [5] theory has been widely studied in organizational context [2], [10]–[14], [19], [20]), it has also influenced how we perceive and understand our networks at societal level [6]. However, this discussion is still at very early stage. Indeed, we need to widen our understanding about social ties to gain deeper insights from human behavior itself [4].

2.2 Measuring tie strength

In the field of IT, tie strength has often been used in different social networks analysis, for example in context of social media [21]–[24], network evolution [25], and information transfer [25] [20], to mention a few. However, how to integrate tie strength research with social network analysis is an ongoing scientific debate. Leading scholars have argued that “despite these important issues, little is known about whether electronic data indeed are a valid proxy for the real social connections they purportedly measure.” [4, 15099].

Granovetter [5] argued that it could be intuitively recognized whether the tie is strong or weak. However, it seems evident that studying complex social networks is not that straightforward. In the recent years, the collective process of production, consumption, and diffusion of information on social media are starting to reveal a significant portion of human social life [26]. This vast amount of information has led to the creation of (1) *big data paradox*: we are now able to access large amount of social media data which is mostly publicly available, yet we may have very little detail about an individual from this big data directly. In order to use this publicly available social media data in a meaningful way to study any social theory (e.g. weak tie theory), it is necessary to filter out the most relevant data. However, the traditional data mining approach of the data preprocessing and noise removal cannot be directly applied to social media data. The social media data suffers also from (2) *noise* *removal fallacy*; firstly the removal of noisy data from social media data may actually remove valuable information and secondly, the definition of noise is complicated and depends completely on the task at hand. Thus, there is a need to develop approaches which can overcome these challenges. [8]

Furthermore, many of the previous studies [21], [27] crawled the data from the social media platforms. Such method are now against the terms of use in these social media platforms. These aspects severely hinders the applicability of the methods of tie strength evaluation from these previous studies. Therefore, it is also essential to utilize openly available social media data [23], as data proprietary may bring up some legal and privacy issues [4] . However, access to this micro-level data (embedded deeply in the both nodes and edges of social networks) may provide many possibilities for tie strength research [4], [6].

We argue that studying the actual *content* of interaction from publicly available data would be especially beneficial in tie strength research. Different "tie strength components" has been suggested in the literature [21], [22], [28], but it needs validating which ones are the most important in the perspective of communication process [3]. However, it seems obvious that both *duration of communication* [5], [29], [30] and *frequency of interaction* [5], [29] are essential mechanisms when people get to know each other better. Therefore, our analysis is based on assumption that 1) the duration of communication and 2) the frequency of interaction *facilitate* tie strengthening process, rather than being an outcome of communication process (like, e.g., trust or emotional support). We suggest that studying tie strengthening process, and integrating the content of these micro-level interactions to large scale SNA could be based on *the network of reciprocal interaction*. Thus, we propose an edge-centered approach that enables identification and analysis of actual communication patterns of individuals rather than emphasizing on the nodes solely.

3 BUILDING A NETWORK OF RECIPROCAL INTERACTION WITH THE EDGE-CENTERED APPROACH: INDIE GAME DEVELOPERS COMMUNITY

In this section we provide an example of how to apply our approach in practice. We build a network of reciprocal interaction from publicly available Facebook data. At first, we describe the selection of the data source (Indie Game Developers community), and why different network properties are important when building a network in context of tie strengthening (i.e. tie strengthening is actually expected). Then we will explain explicitly how data collection and analysis was done. Finally, we present a visualization from the actual network.

3.1 Selection and description of data source (Indie Game Developers community)

To build and to demonstrate the new approach to enabling tie strength evaluation from large publically available social media data, we first devised a set of criteria for finding a suitable and sufficiently challenging social media based large community. With this, we followed *theory based sampling* [31] as it focuses specifically on "phenomenon of interest". We believe that our data set suits well to study tie strengthening as phenomena in the context of, for example, knowledge work.

First, (1) the community in question has to be large in terms of number of overall members, and number of active members, as well as number of posts and discussions between members. This creates a challenging task while there is, supposedly, a lot of noise in the form of low frequency interactions which do not necessarily lead to tie strengthening. This creates a challenge for filtering out relevant discussions and posts, most probably increasing tie strength. Secondly, (2) the community needs to have been active for several years, so that we can expect that ties have already been formed and strengthened within the community between members during its lifecycle. Third, (3) the community can be expected to such that it includes both simple problem solving type of activities which do not require tie strengthening as such, as well as more goal-oriented long-term activities, such as competence and capability building and longer-term learning, that benefit from the building of relationships within the community and the strengthening of ties. This creates a further challenge and need to filter out irrelevant interactions from most relevant reciprocal interactions - we could not expect to notice very significant tie strengthening to be happening in very loose communities such as photo or video sharing communities. Furthermore, (4) the community should not be very centralized and top-down managed, because more in-depth tie strengthening requires frequent interactions which are reciprocal by nature [5], [29]. Finally, (5) for practical reasons enabling the content analyses in later phases, the community should be open for access and have open wall or API for enabling the collecting of data, and be mainly or solely English-language to allow the use of content analysis and text mining algorithms in the analyses of interactions and the tie-strengthening.

After reviewing different types of communities, both work- and non-work- related ones, such as, and reviewing their suitability and requirement for the sufficient challenge for filtering the relevant discussions from less relevant ones, we ended up in selecting a Facebook-based Indie Game Developers community as the source for collecting the data and creating and preliminarily testing the created approach.

3.2 Data collection

The Facebook data for the indie game developers was collected using the Social Data Analytics Tool (SODATO) [32], [33]. Full historical fetch of the Facebook group was done from 05-01-2008 to 29-04-2017 using SODATO. SODATO enables the systematic collection, storage, and retrieval of the entire corpus of social data for Facebook walls and groups. For this study the data from 01-01-2015 to 31-12-2015 was used. The details for this data corpus are provided in the Table 1.

Table 1: Details of Data Corpus

Content Attribute	Value	Actor Attribute	Value
Time period	Start: 2015-01-01	Total Unique Actors	337425
	End: 2015-12-31		
Total Page Likes	--	Unique Posters	24553
Posts	26290	Unique Commenters	103581
Comments	205729	Unique Comment Reply Actors	9613
Comment Replies	23084	Unique Wall Post Likers	327
Likes on Posts	332886	Unique Comment Likers	16043
Likes on Comments	35152	Unique Comment Reply Likers	5817

3.2 Data analysis

3.2.1 Social Network Creation. Facebook data in general allows forthright analysis. In this case study Facebook posts, comments and comment reply (reply to a comment) were the entities used in the analysis. A tailored Python script was written to analyze the above mentioned entities in Facebook data. The script further converted the refined data into the following network. The network shows interconnections between people communicating on Facebook. More explicitly, with interconnections, we mean users initiating Facebook posts, comments and comment replies to aforementioned Facebook entities. The Python script uses NetworkX library (version 1.11) to construct the network and serialize it in Graph Exchange XML Format or GEXF (version 1.2). The Python script was used to create the network graph files.

3.2.1 Social Network Visualization (Fig. 1). Gephi, an interactive visualization and exploration platform available in open source [34], was used to analyze and visualize the networks. Gephi was used to layout the networks, calculate metrics for network nodes, filter out the network using the different Gephi filters in conjugation (like self-loop, mutual edge, degree range) and adjust the visual properties of the visualized network according to the analysis. The layout of the networks in this study is the outcome of a force driven layout algorithm in which nodes repel each other and the edges connecting the nodes act as springs pulling the nodes back together [34]. Hence the nodes that are interconnected will be placed close to each other.

In this study, the python script to create network graph resulted in a network which contained 16260 nodes and 123137 edges. This graph file (.gexf) was opened in Gephi and the filtering of the network was done. First, the Gephi filter 'Giant Component' was applied. This filter keeps only the graph component with most number of nodes. Second, the Gephi filter 'Mutual Edge' was applied to the network. Mutual Edge filter keeps only edges that have mutual or reciprocal edges. This step is necessary as the tie strengthening process *requires reciprocal interaction between actors* (nodes). Thus, filtering out all actors which did not have any reciprocal interaction was executed. Third, the Gephi filter Self Loop was applied to the network. Self Loop filter removes all the self loops from the network. This step is essential as the tie strength is always between two dyads and not for an individual node. Finally, the Gephi filter Degree Range was used. This filter keep nodes which lie within the defined degree range.

The filtered network was visualised using the force driven layout algorithm. The visualization of the Facebook group Indie Game Developers during the period of the study is shown in Figure 1. This visualisation contains 313 nodes and 1072 edges which was obtained by applying the Gephi filters as explained earlier. The degree range for the Gephi Degree Range filter was set to be between 100 and 645 resulting in the filtered network shown in Figure 1. The nodes in the visualization represent the Indie Game group members. While their interaction are made visible by connections to other members, the greater the interaction the greater the size of the connection (line width in Figure 1). The color of node represents the cluster of nodes in the network, which is based on a community-detection algorithm that analyzes the network to find group of nodes that are particularly tightly interconnected. The size of each node represents the degree of interaction (sum of indegree and outdegree) of each group member. The larger the size of the node the greater the amount of interaction of the member.

To put it simply, these 313 nodes and 1072 edges (presented in Figure 1.) show the amount of individuals (nodes) and their communication patterns (edges). This visualization has been cleared of noise, meaning that less than 100 messages exchanged between individuals was not taken into account. This level of granularity may be adjusted as wanted, of course. Noise removal is directed to low-frequency interactions (edges with low weight), in which tie strengthening is not expected as total amount of communication remains low. Therefore, different tie strengthening patterns are expected to manifest within these 1072 edges. The same approach could be applied to significantly larger data sets, and that would leverage generalization capabilities significantly when analyzing the content of these messages.

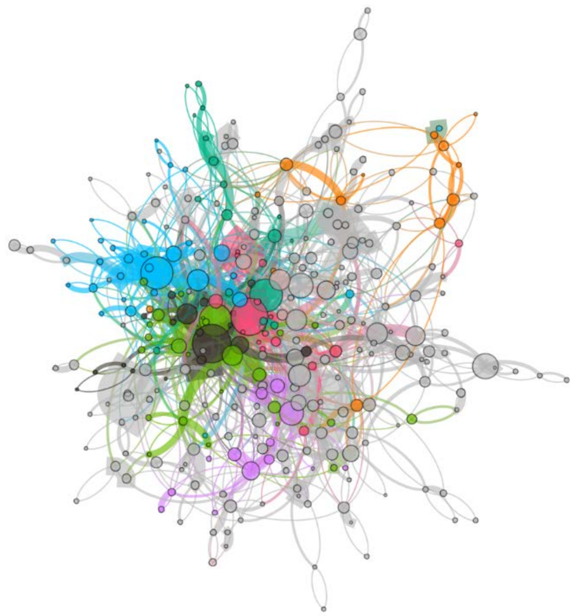

Figure 1: Force driven network of people based on Facebook conversations.

4 CONTRIBUTIONS AND CONCLUSION

The objective of our study was to find relevant actors (nodes) and their mutual interactions (edges) from publicly available large social media dataset in the context of the tie strengthening process. The focus of this study was to overcome the limitations of earlier studies which use data gathering methods that are not applicable (due legality or privacy issues), or other studies which use measurements that may not reflect the true social connections of individuals accurately enough. In this paper, we presented the edge-centered approach which overcomes many of these earlier limitations and demonstrated this approach with Facebook data. Our findings demonstrate that this approach could benefit future tie strength studies

4.2 Contributions to research

At first, the edge-centered approach in tie strength research allows researchers to build a network of reciprocal interaction from open social media data without compromising current legal or privacy issues with social media data. Secondly, getting rid of "low frequency" connections between individuals mitigate some of the problems with big data paradox and noise removal fallacy. Indeed, the formation of strong ties requires time and frequent interaction to develop, making low frequency connections less relevant from tie strengthening perspective [5], [30]. Therefore, this approach "exploits survivorship bias" when focusing on interaction patterns of possible strong ties after they have been formed.

We believe the edge-centered approach would bring up many interesting possibilities when, for example, integrating large-scale SNA with text mining and machine learning algorithms which are able to analyze and discover patterns from large amount of natural language in automated manner [35]. Basically, the edge-centered approach could be applied to any social media dataset with discussion data. Social network data and message level content analysis has been used to, for example, identifying leadership in online communities (what kind of language is used, sociability of individuals, etc.) [36]. However, studies that analyze content of messages use often manual coding (like previously mentioned study), not automated text-mining or machine learning algorithms [35].

We also believe that this approach would allow researchers to test well-established sociological theories and assumptions from human behavior. For example, Granovetter [5] mentioned that "the strength of a tie is (probably a linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie" [5, 1361]. There is also a plethora of similar assumptions in other theories as well that could be tested by analyzing large amount of different communication patterns in social networks.

For example, *the norm of reciprocity* is critical "starting mechanism" for the strengthening of social relationships at their early stage. This mechanism includes also simple social exchange, like providing respect, appreciation, etc. [17], which could be spotted from text-based interaction as well. This kind of confirmatory research (theory testing) would enable significant theoretical contributions but could also guide organizations in managing the interaction of their employees in the most effective manner. It would be interesting to see what kind of communication patterns survive (and create stronger ties between individuals) and what fade away.

Furthermore, our approach complements research methods which rely electronic (likes, links shared, etc.), or self-reported data (questionnaires, etc.) which may suffer from inaccuracy, recall- and cognitive biases, or the errors of perception [4]. Studying these communication patterns objectively would bring us closer to *social physics* [39] by the means of *computational social science* [7].

4.1 Contributions to practice

Analyzing conversational data in online social networks would provide interesting insights of how different "interaction layers" are developing over time. As many virtual teams are struggling with unsuccessful communication [1], information systems (IS) scholars have argued that more research is needed to understand the content of communication at network level [36]. Visualizing these interactions would enhance sense-making significantly, making this kind of information more useful for many organizations and decision makers.

Combining ample volumes of texts with large-scale SNA would produce insightful results. For example, it would be beneficial for organizations to know how emotional support is transferred in distributed works setting, as self-disclosure is higher (people "open up" more easily) in text based interaction, but on the other hand, lack of social cues and low synchronicity is a source of potential conflicts and misunderstandings [37]. Preventing or intervening these conflicts beforehand would be useful for large organizations with complex intra-organizational social networking services with a lot of interaction data. Accordingly, it would be possible to spot sources of excitement occurring in social networking services; who enjoys discussing with whom, how communication patterns between individuals are evolving (like emotional content, language formality, social exchange), etc., and that would help when organizations struggle to find proper group dynamics in virtual teams.

4.2 Conclusion and limitations

We did not use any content analysis or text-mining techniques in this study. We consider this as a limitation. However, we argue that this should be the next step in refinement of the edge-centered approach. We also believe the approach presented in this paper is applicable as such.

The edges present the discussions, the messages, and all the actual content of interaction in the network. Therefore, analyzing the edges provides interesting insights in tie strengthening perspective, especially if the amount of data gets larger and allows more generalization from conversational data. However, at first, the actual tie strengthening process needs to be justified by theory. We need to understand the core tie strength components that are visible (and what are not) in text based interaction. We strongly encourage researchers towards this kind of interdisciplinary research. Many theorists believe that the key to understand interpersonal relations is hidden behind these communication patterns. Building and maintaining relationships is all about communication, also in an online environment, as also described by Walther [40, 443]; "Although many people perceive that social media messages are trivial and banal, so is the stuff by which relationships are maintained". Gaining deeper insights from the essence of human online communication would benefit many organizations, as well as our society as a whole.

5 ACKNOWLEDGMENTS

This work has been funded by Academy of Finland (Suomen Akatemia) under grant 295893 for the research project COBWEB.

REFERENCES

[1] L. Dubé and D. Robey, "Surviving the paradoxes of virtual teamwork," *Inf. Syst. J.*, vol. 19, no. 1, pp. 3–30, Jan. 2009.

[2] J. Kotlarsky and I. Oshri, "Social ties, knowledge sharing and successful collaboration in globally distributed system development projects," *Eur. J. Inf. Syst.*, vol. 14, no. 1, pp. 37–48, Mar. 2005.

[3] O. Torro and H. Pirkkalainen, "Strengthening Social Ties via ICT in the Organization," 2017.

[4] S. Wuchty, "What is a social tie?," *Proc. Natl. Acad. Sci.*, vol. 106, no. 36, pp. 15099–15100, Sep. 2009.

[5] M. S. Granovetter, "The Strength of Weak Ties," *Am. J. Sociol.*, vol. 78, no. 6, pp. 1360–1380, May 1973.

[6] S. Aral, "The Future of Weak Ties," *Am. J. Sociol.*, vol. 121, no. 6, pp. 1931–1939, May 2016.

[7] D. Lazer *et al.*, "Computational Social Science," *Science*, vol. 323, no. 5915, pp. 721–723, Feb. 2009.

[8] R. Zarafani, M. Ali Abbasi, and H. Liu, *Social Media Mining - An Introduction*. .

[9] A.-L. Barabási, "Scale-Free Networks: A Decade and Beyond," *Science*, vol. 325, no. 5939, pp. 412–413, Jul. 2009.

[10] D. Krackhardt, "The Strength of Strong Ties: The Importance of Philos in Organizations," in *Networks and Organizations: Structure, Form,and Action*, Harvard Business School Press, 1992, pp. 216–239.

[11] M. Tortoriello, R. Reagans, and B. McEvily, "Bridging the Knowledge Gap: The Influence of Strong Ties, Network Cohesion, and Network Range on the Transfer of Knowledge Between Organizational Units," *Organ. Sci.*, vol. 23, no. 4, pp. 1024–1039, Aug. 2011.

[12] D. Z. Levin and R. Cross, "The Strength of Weak Ties You Can Trust: The Mediating Role of Trust in Effective Knowledge Transfer," *Manage Sci*, vol. 50, no. 11, pp. 1477–1490, Nov. 2004.

[13] M. T. Hansen, "The Search-Transfer Problem: The Role of Weak Ties in Sharing Knowledge across Organization Subunits," *Adm. Sci. Q.*, vol. 44, no. 1, pp. 82–111, 1999.

[14] D. Constant, L. Sproull, and S. Kiesler, "The Kindness of Strangers: The Usefulness of Electronic Weak Ties for Technical Advice," *Organ. Sci.*, vol. 7, no. 2, pp. 119–135, Apr. 1996.

[15] Vilma Vuori and Jussi Okkonen, "Knowledge sharing motivational factors of using an intra-organizational social media platform," *J. Knowl. Manag.*, vol. 16, no. 4, pp. 592–603, Jul. 2012.

[16] P. M. Blau, *Exchange and Power in Social Life*. Transaction Publishers, 1964.

[17] A. W. Gouldner, "The Norm of Reciprocity: A Preliminary Statement," *Am. Sociol. Rev.*, vol. 25, no. 2, pp. 161–178, 1960.

[18] R. P. Settoon, N. Bennett, and R. C. Liden, "Social exchange in organizations: Perceived organizational support, leader–member exchange, and employee reciprocity," *J. Appl. Psychol.*, vol. 81, no. 3, pp. 219–227, 1996.

[19] K. Rost, "The strength of strong ties in the creation of innovation," *Res. Policy*, vol. 40, no. 4, pp. 588–604, May 2011.

[20] S. Retzer, P. Yoong, and V. Hooper, "Inter-organisational knowledge transfer in social networks: A definition of intermediate ties," *Inf. Syst. Front.*, vol. 14, no. 2, pp. 343–361, Apr. 2012.

[21] E. Gilbert and K. Karahalios, "Predicting Tie Strength with Social Media," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2009, pp. 211–220.

[22] E. Gilbert, "Predicting Tie Strength in a New Medium," in *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, New York, NY, USA, 2012, pp. 1047–1056.

[23] J. P. Gupta *et al.*, "Identifying weak ties from publicly available social media data in an event," 2016, pp. 11–19.

[24] C. J. Hutto, S. Yardi, and E. Gilbert, "A Longitudinal Study of Follow Predictors on Twitter," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2013, pp. 821–830.

[25] G. Kossinets, J. Kleinberg, and D. Watts, "The Structure of Information Pathways in a Social Communication Network," in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2008, pp. 435–443.

[26] C. A. Davis *et al.*, "OSoMe: the IUNI observatory on social media," *PeerJ Comput. Sci.*, vol. 2, p. e87, Oct. 2016.

[27] R. L. Fogués, J. M. Such, A. Espinosa, and A. Garcia-Fornes, "BFF: A tool for eliciting tie strength and user communities in social networking services," *Inf. Syst. Front.*, vol. 16, no. 2, pp. 225–237, Apr. 2014.

[28] A. Petroczi, F. Bazsó, and T. Nepusz, "Measuring tie-strength in virtual social networks," *Connections*, vol. 27, no. 2, pp. 39–52, 2006.

[29] P. Blumstein and P. Kollock, "Personal Relationships," *Annu. Rev. Sociol.*, vol. 14, no. 1, pp. 467–490, 1988.

[30] J. B. WALTHER, "Interpersonal Effects in Computer-Mediated Interaction: A Relational Perspective," *Commun. Res.*, vol. 19, no. 1, pp. 52–90, Feb. 1992.

[31] M. Q. Patton, *Qualitative Research & Evaluation Methods*. SAGE, 2002.

[32] A. Hussain and R. Vatrapu, "Social Data Analytics Tool (SODATO)," in *Advancing the Impact of Design Science: Moving from Theory to Practice*, 2014, pp. 368–372.

[33] A. Hussain, R. Vatrapu, D. Hardt, and Z. A. Jaffari, "Social Data Analytics Tool: A Demonstrative Case Study of Methodology and Software," in *Analyzing Social Media Data and Web Networks*, M. Cantijoch, R. Gibson, and S. Ward, Eds. Palgrave Macmillan UK, 2014, pp. 99–118.

[34] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: An Open Source Software for Exploring and Manipulating Networks," *ResearchGate*. [Online]. Available: https://www.researchgate.net/publication/221297890_Gephi_An_Open_Source_Software_for_Exploring_and_Manipulating_Networks. [Accessed: 02-Jun-2017].

[35] S. Debortoli, O. Müller, I. Junglas, and J. vom Brocke, "Text Mining For Information Systems Researchers: An Annotated Topic Modeling Tutorial," *Commun. Assoc. Inf. Syst.*, vol. 39, no. 1, Jul. 2016.

[36] S. Faraj, S. Kudaravalli, and M. Wasko, "Leading Collaboration in Online Communities," *MIS Q*, vol. 39, no. 2, pp. 393–411, Jun. 2015.

[37] D. Derks, A. H. Fischer, and A. E. R. Bos, "The role of emotion in computer-mediated communication: A review," *Comput. Hum. Behav.*, vol. 24, no. 3, pp. 766–785, May 2008.

[38] A. Majchrzak, L. Cherbakov, and B. Ives, "Harnessing the power of the crowds with corporate social networking tools : how IBM does it," *MIS Q. Exec. Res. J. Devoted Improv. Pract.*, vol. 8, no. 2, 2009.

[39] A. Pentland, *Social Physics: How Social Networks Can Make Us Smarter*. Penguin, 2015.

[40] J. B. WALTHER, *Theories of computer-mediated communication and interpersonal relations*. .