Identifying Weak Ties from Publicly Available Social Media Data in an Event

Jayesh Prakash Gupta
Tampere University of Technology
Korkeakoulunkatu 8
33720,Tampere, Finland
+358465757377
jayesh.gupta@tut.fi

Karan Menon
Tampere University of Technology
Korkeakoulunkatu 8
33720,Tampere, Finland
+358401305046
karan.menon@tut.fi

Hannu Kärkkäinen
Tampere University of Technology
Korkeakoulunkatu 8
33720,Tampere, Finland
+358405026358
hannu.karkkainen@tut.fi

Jukka Huhtamäki
Tampere University of Technology
Korkeakoulunkatu 8
33720,Tampere, Finland
+358405854771
jukka.huhtamaki@tut.fi

Raghava Rao Mukkamala[1]
[1]Copenhagen Business School
Howitzvej 60
2000 Frederiksberg, Denmark
+4541852299
rrm.itm@cbs.dk

Abid Hussain[1]
[1]Copenhagen Business School
Howitzvej 60
2000 Frederiksberg, Denmark
+4538154478
ah.itm@cbs.dk

Ravi Vatrapu[1,2]
[2]Westerdals Oslo School of ACT
Christian Krohgs gate 32,
0186,Oslo, Norway
+4538154435
rv.itm@cbs.dk

Jari Jussila
Tampere University of Technology
Korkeakoulunkatu 8
33720,Tampere, Finland
+358407178345
jari.j.jussila@tut.fi

Henri Pirkkalainen
Tampere University of Technology
Korkeakoulunkatu 8
33720,Tampere, Finland
+3580504371567
henri.pirkkalainen@tut.fi

ABSTRACT

The concept of weak ties was introduced by Granovetter through the seminal paper titled "Strength of weak ties". Since then the role of weak ties in general and their specific role as occupying the structural hole has been explored in many different fields. In this study, we identify actual or potential weak ties using publicly available social media data in the context of an event. Our case study environment is community managers' online discussions in social media in connection to the yearly-organized Community Manager Appreciation Day (CMAD 2016) event in Finland. We were able to identify potential weak ties using the conversation based structural holes, making use of social network analysis methods (like clustering) and content analysis in the context of events. We add to the understanding of and useful data sources for the Strength of weak ties theory originated from Granovetter, and developed further by other researchers. Our approach may be used in future to make more sophisticated conference recommendation systems, and significantly automate the data extraction for making useful contact recommendations from them for conference participants.

AcademicMindtrek'16, October 17-18, 2016, Tampere, Finland.

DOI: http://dx.doi.org//10.1145/2994310.2994354

CCS Concepts

• Human-centered computing~Computer supported cooperative work • Human-centered computing~Social recommendation • Human-centered computing~Empirical studies in collaborative and social computing

Keywords

Weak Ties; Tie Strength; Structural Hole; Event; Conference; Social Media; Twitter; Facebook; Recommendation System.

1. INTRODUCTION

The concept of "tie strength" was introduced by Granovetter[1] through the seminal paper titled "Strength of weak ties". The tie strength concept, as well as the concepts of weak and strong ties which are closely related to tie strength, have received a lot of academic interest since 1970's, drawing from Granovetter's seminal constructs [1]. The above concepts have been demonstrated to significantly impact the transfer and dissemination of knowledge and information [2]. Rather recently, the related research has gained new interest due to the rise and maturation of different types of social media and social network sites. Social media has been found to provide new ways to identify, create, strengthen and manage ties (e.g. [3]), and even help automate this process in many important ways. Social media have also created social big data, which provides new opportunities for making use of big data in various contexts.

More specifically, in the recent years social media have provided new ways of networking with other people both in general, as well as even in co-located events like conferences [4]. In conferences, one of the aims of the participants is to meet new people that might share similar interests or could provide relevant information[5] . Quite naturally, this has resulted in designing e.g.

conference recommendation systems which aim to provide relevant connection recommendations for conference participants [4], [6].

The reason for the selection of conferences and events as the context in this paper is due to their important role in the dissemination and exchange of scientific, managerial and other types of information and knowledge, as well as in the importance of the facilitation of related networking and collaboration between conference participants. Increasingly important ways for networking people in events and conferences are various social networking services and social media at large (e.g.[7]). During the last few years, social media has been used in various ways to support networking and collaboration between people in conferences, e.g. by content analysis and visualization of social media data, as well as by detecting weak and strong ties from social media.

In addition to giving recommendations based on certain keywords (participant interests, competence areas), usually extracted manually from the conference participants, recently some studies have tried to incorporate other sources of data like co-occurrence data or participant's mobile device data to provide more relevant recommendations [4], [8], [9]. Also data from social media sites like Twitter has been used to gain better insights about e.g. conference topics and useful contacts [10]. However, there is little research on the use of tie strength based recommendation systems [9] in case of an event. This study explores the possibility of evaluating tie strength using publicly available social media data about the conference. This could thus provide a novel conference recommendation and networking approach.

This study differs from and contributes to earlier above types of studies and other studies first, by making use of publicly available social media data about an event, and more specifically, social media conversation data (not the non-public personal social media data from personal networks, as in previous research) - merely using the non-public and personal social media data, or e.g. conversational data from email networks (see [11]), the automated evaluation of tie strength and the detection of e.g. weak or strong ties for networking recommendations cannot be established in such context. Second, our study differs from earlier ones by making use of network-level (not individual level as commonly done in previous research) measures, drawing from structural hole theory of Burt [12], [13], whilst structural holes can help to access non-redundant ties, and can help to identify potential and actual weak ties.

Taking into consideration the above research gaps in literature, we have devised the following research questions to address the gaps:

1. Can actual or potential weak ties be identified using the structural hole perspective at a network level, particularly in the context of events and seminars?
2. To what extent can the structural holes be identified from conversation data and especially publicly available social media data in the context of events and seminars?

To address the above research questions, this paper has been structured as follows: we first introduce the main concepts of structural holes, tie strength, and weak and strong ties. We then describe how weak ties have been identified, and how they have been evaluated especially using social media data. We also describe the previous use of social media in event-related networking and weak tie detection. We then introduce our research methodology based on one social media lead user event and community, CMAD Finland (the popular event of community managers that are active in their social media use both generally and during the yearly CMAD event). We describe the results from CMAD2016 event in Finland with 270 participants, and finally, discuss the significance of the results as well as our contribution to earlier research.

2. WEAK TIE IDENTIFICATION IN AN EVENT SETTING USING SOCIAL MEDIA

2.1 Concept of structural holes

Structural holes appear in social networks. According to Easley and Kleinberg [14], a structural hole in an organization is "the 'empty space' in the network between two sets of nodes that do not otherwise interact closely." Structural holes appear in literature extensively in different forms, often in the context of social capital and creativity or the creation of new knowledge [13]. An individual bridging a structural hole is able to increase his/her social capital through accumulating non-redundant information from varied sources[12], [13].

Interestingly, opposite viewpoints to structural holes as source of social capital exist [15]: on one hand, areas in a social network where connections between actors are missing provide new opportunities for actors to form bridges, therefore structural holes serve as a potential source for social capital [12]. Alternatively, dense network structure can be perceived as high social capital as the removal of individual connections serving as bridges does not affect the overall network[1] [17], [18].

In this paper, we subscribe to the view that structural holes are a source of potential social capital and, importantly, creativity. With brokerage as the mechanism, structural holes are an important source of social capital (Burt, 2004): "Opinion and behavior are more homogeneous within than between groups, so people connected across groups are more familiar with alternative ways of thinking and behaving. Brokerage across the structural holes between groups provides a vision of options otherwise unseen, which is the mechanism by which brokerage becomes social capital." Different perspectives of the relationship between structural holes and social capital exist, however according to Burt (2000) the perspectives agree "on a social capital metaphor in which social structure is a kind of capital that can create for certain individuals or groups a competitive advantage in pursuing their ends".

2.2 Concept of weak ties

According to Granovetter [1], the tie strength can be defined as "a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie". Based on this definition he characterized two kinds of ties - strong ties and weak ties. Strong ties are people whom you trust and who can provide you emotional support for example family members or close friends [1], [19], [20]. On the other hand, weak ties are people with whom you just have acquaintance. Weak ties have been attributed to providing novel information in many different context across a range of different studies. (see example [2], [21], [22]) One of the most important attribute of weak ties is that they have access to and can provide access to non-redundant information and act as bridging ties.[1], [23] According to Burt's structural hole theory[12], [13], structural holes have access to non-redundant ties and are usually the weak ties. Burt provides

[1] Scale-free networks, for example, are very vulnerable to coordinated attacks once their structure is revealed [16].

empirical evidence that bridging ties are weak ties [12], [24]. However, he considers tie strength as mere correlation of the underlying principle of non-redundancy. Based on these commonalities, it can be seen that both the theories of Strength of Weak Ties and Structural Hole Theory have small differences in ornamentation but are based on how networks work.[24] Thus, it can be seen that identifying weak ties using the structural holes or identifying weak ties by calculating tie strength are interrelated and can be used in conjugation to identify the weak ties.

In this study we follow approach suggested by Marsden and Campbell [25] and aim to conceptualize tie strength and structural hole in order to identify weak ties in an event context. We do this by evaluating tie strength at interpersonal level (between the event participants) making use of communication frequency as a proxy for tie strength evaluation and identifying potential weak ties in the context of an event setting.

2.3 Weak tie identification from social media

The proliferation of social media and social network sites has given rise to novel ways to establish and manage ties online [3]. This has led to studies which use the social networking sites data to predict the tie strength of these online relationships. Many of these studies have used the online personal data to calculate the tie strength. For example, studies to calculate tie strength using Facebook have used data related to participant's Facebook profile and friends. While studies related to Twitter, have used the data about the participant's followers and followees to calculate tie strength [26]–[28]. These kind of studies have used supervised computational methods that required human annotations like rating friends or nominating top friends [29], [30]. A few studies have used the unsupervised computational methods (see e.g. [29]) yet have relied on online personal data. Hence, it is evident that the earlier studies have operationalized tie strength calculations related to a participant's friend network using various individual measures and predictors which can be obtained from social networking sites [3], [26], [28], [28], [29]. There are a few studies which have used the social media data like Facebook friend list to directly identify the structural holes based on the network created using the friend list from social media. (see e.g. [31], [32]) However, these studies have relied on using participant's personal data from social networking sites that may not be accessible in case of events like conferences.

In the last few years, there has also been research which has used publically available social media data for calculating tie strength for example [33], [34]. However, these studies are related to tracing the actual information flow in large scale social networks and not for identifying different kind of ties [33], [34]. Few studies have used an organization's internal social network site to evaluate tie strength in professional context [35]. The present study uses publicly available data about an event from two different social networking sites. Hence, this study tries to address a research gap/ limitations of previous studies of using personal data from a single social media source [26]–[28], [30], [35] for identifying weak ties using tie strength and structural hole identification in the context of an event.

2.4 Social media and events

In the past few years' social media and social networking sites have provided a new way of networking with other people even in co-located events like conferences [4]. In such conferences, one of the aims of the participants is to meet new people who may share similar interests or may provide relevant information [5]. This has resulted in a need to build conference recommendation systems which may provide relevant recommendations to the participants [5], [6]. Generally, these kind of systems have relied on giving recommendation based on certain keywords which may be extracted from the conference participant's registration form or some other information participant information provided at the time of conference registration [6], [9]. Recently some studies have tried to incorporate other sources of data like bibliographic data, co-occurrence data, participant's mobile device data and also data from sites like epinions.com, Flickr to provide more relevant recommendations[4], [8], [9]. On the other hand, data from social media sites like Twitter and Facebook has been used by the conference organizers to gain better insight about the conference and help in better planning for the future conferences [10]. However, there is a limited research on the use of weak ties to provide recommendation systems [9] in case of an event. In the past there have been a few studies to calculate tie strength in the context of an event, but these studies do not use any real empirical data or social media data [8]. The present study explores the possibility of identifying weak ties using publically available social media data about a conference. This approach may provide a novel method for conference recommendation in the future.

3. RESEARCH METHOD AND APPROACH

3.1 Case description

Our case study environment is community managers' online discussions in social media in connection to yearly-organized Community Manager Appreciation Day (CMAD 2016) event in Finland. The most recent event took place on January 25, 2016 in Jyväskylä, Finland. CMAD events have been organized globally since 2010 and they originate from Jeremiah Owyang's blog to recognize and celebrate the efforts of community managers around the world using social media to improve customer experiences [4]. The organizing committee of the fifth CMAD event (CMAD 2016) in Finland included 17 people participating in the planning meetings with a supporting online community of 238 members. Total of 270 people participated in the CMAD 2016 event.

It can be argued that discussions in social media represent only a small part of the overall communication between community members in professional communities, because many professionals are not in social media (e.g. Twitter or Facebook) or are not actively using social media in professional context. As a consequence, data-driven approaches can be seen as a limitedly useful in studying professional communities. In this case, however, majority of the community members belonging to the community of community managers can be considered as advanced lead users of social media and online community management approaches, with most of them being highly active in Twitter and Facebook.

Furthermore, related to learning events and conferences, it has been observed that most of the activity take place during the learning event or conference, with little communication before and after [36], making it questionable to draw any legitimate conclusions from data collected before and after the conference. However, based on previous studies of community managers in Finland [10], [37], we argue that community managers communicate with each other also between events, and have also participated actively in planning the event, and assume that by collecting data based on these community member's discussions from Twitter and Facebook we can capture sufficient and representative amount of data to draw conclusions.

3.2 Data Collection

3.2.1 Social media data

The social media data for the event CMAD 2016 was collected from Twitter and Facebook. The detail corpus statistics for both Facebook and Twitter data are given in Table 1 and Table 2 respectively.

Table 1.Facebook data corpus

Content Attribute	Value	Actor Attribute	Value
Time period	Start: 2013-02-04 End: 2016-05-23	Total Actors	8798
Total Page Likes	--	Total Unique Actors	374
Posts	555	Unique Posters	81
Comments	2925	Unique Commenters	199
Comment Replies	149	Unique Comment Reply Actors	53
Likes on Posts	2529	Unique Wall Post Likers	327
Likes on Comments	2536	Unique Comment Likers	204
Likes on Comment Replies	104	Unique Comment Reply Likers	38

Full historic fetch of the two Facebook pages (CMADFI 2014 & CMADFI 2015) from 01-01-2014 to 26-05-2016 was conducted using the Social Data Analytics Tool; [38], [39]. SODATO enables the systematic collection, storage, and retrieval of the entire corpus of social data for Facebook walls and groups.

Table 2. Twitter data corpus

Content Attribute	Value	Actor Attribute	Value
Time period	Start: 2013-01-21 End: 2016-04-18	Total Users	12454
Total Tweets	12454	Total Unique Users	1651
Original Tweets	7568	Unique Original Tweet Users	858
Retweets	4886	Unique Retweet Users	1262

Twitter data was collected in two phases. First, in order to list all tweets sent before, during, and after CMADFI 2016, we accessed Flockler[2], a social media-driven content management system that is used to run the CMADFI website. Flockler provides a web application programming interface (API) that allowed us to collect all tweets related to CMADFI 2016. In addition, in order to collect a full set of Twitter data including the full set of metadata that Twitter provides for each tweet, we distilled tweet ids from Flockler data and implemented a tailored batch script that uses Twitter REST API[3] to access full tweet data including tweet sender, Twitter users mentioned in each tweet as well as hashtags related to tweets. The batch script exports tweet data in JSON for further processing. For this study the social media data from 1st September, 2015 to 30th April, 2016 was used for performing all the analysis.

3.2.2 Survey data

The second source of data was collected from the conference participants directly. Self-reported data was necessary means of data for us to interpret the social media data against our theoretical framing. The survey was operationalized based on the theoretical descriptions of Granovetter [1] and adapting operationalized scale by [40]. The following survey items (Table 3) were operationalized.

[2] https://flockler.com/

[3] https://dev.twitter.com/rest/public

Table 3. Survey questions

Survey Question
1. How novel (on an average) was the information you received from the CMAD 2016 participants amongst the following groups?
2. Which 3 - 5 CMAD 2016 participants do you consider as source of most novel information or ideas?

Question 1 asked the participants to rate novelty of the information from three separate groups of participants on scale of 1-7. These three groups were: participants who survey respondent knew well; participants who survey respondent met face to face but did not know well; and participants who survey respondent have not had face to face interaction with. Question 1 was used to identify the different sources and quality of information in general. On the other hand, Question 2 focused on identifying novel information sources for individual survey respondent. Due to practical problem of recalling names of survey participants, we limited the number of participant names to five.

An online survey link was shared to all the CMAD 2016 participants through the CMAD Facebook group wall and also by the official twitter handle of CMAD. 25 survey responses were received from a total of 270 participants. The survey was available in English and Finnish and was based on the CMAD 2016 event only.

3.3 Data analysis

Twitter and Facebook data in general allows forthright analysis. In case of Twitter, the used Rest API arranges the tweet data in a format that is easy to process programmatically. This means that the users (e.g. @jyshgupta) and hashtags (e.g. #cmadfi) are represented with a specific syntax and structure. Similarly, in case of Facebook, posts, comments, comment reply (reply to a comment) and likes were the entities used in the analysis.

A tailored Python script was written to analyze the above mentioned entities in both Twitter and Facebook data. The script further converted the refined data into two networks:

- The first network represents interconnections between people communicating over Twitter. More explicitly, with interconnections, we point to users mentioning each other in tweets through comments and discussions.
- The second network shows interconnections between people communicating on Facebook. More explicitly, with interconnections, we mean users initiating Facebook posts, comments and comment replies as well as “Likes” 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 create the network graph files.

Gephi, an interactive visualization and exploration platform available in open source [41], was used to analyze and visualize the networks. Gephi was used to layout the networks, calculate metrics for network nodes, analyze networks for possible sub-networks or clusters (Modularity Class metric) calculated with Gephi’s implementation of community detection algorithm [42] and adjust the visual properties of the visualized network according to the analysis. In this particular case, the weighted degree (sum of weighted indegree and outdegree) and modularity class (clustering) are the metrics that were of interest in 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 [41]. Hence the nodes that are interconnected will be placed close to each other [37], [43].

4. FINDINGS

4.1 Descriptive analysis

Based on the responses received for Question 1 of the survey it was observed that the most novel information (on an average) was received from the participants whom the respondents had not met face to face (average rating on a scale of 1 to 7 was 5.13), in comparison to participants that respondents knew well (average rating on a scale of 1 to 7 was 4.00), or had met face to face but did not know well (average rating on a scale of 1 to 7 was 4.65).

4.2 Analysis based on correlating social media data with survey data

The visualization of the CMAD participants' conversation on Twitter and Facebook during the period of the study is shown in Figure 1 and 2. The nodes in the visualization represent the CMAD participants. While their interests are made visible by connections to other participants, the greater the interest the greater the size of the connection (line width in Figure 1 and 2). 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.

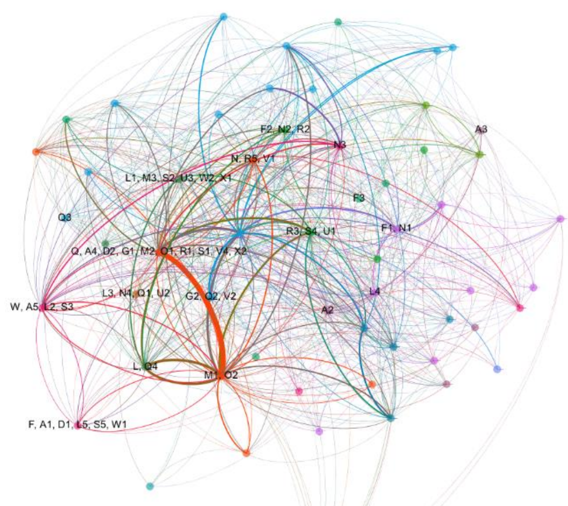

Figure 1. Force driven network of people based on tweets

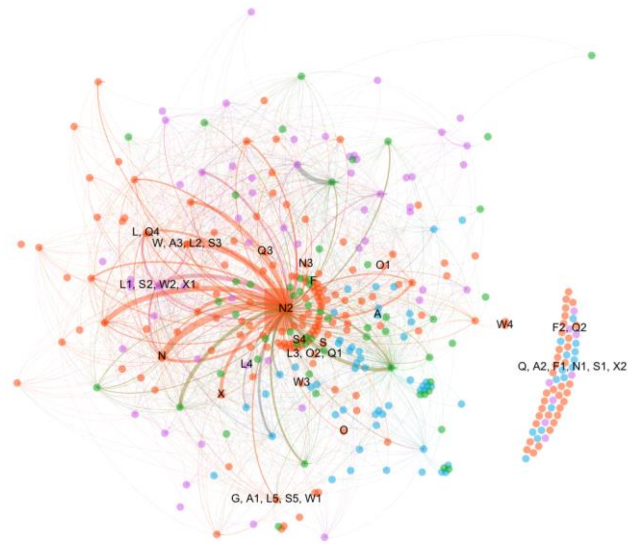

Figure 2. Force driven network of people based on Facebook conversations

The labeled nodes in the network graphs shown in Figure 1 and 2 represent the survey respondents (alphabetical letters A to X) and also their novel source of information as provided in the survey response for Question 2 (for example, survey respondent is labeled as A while his/her novel information sources are labeled as A1, A2, A3, A4 and A5). In case of Twitter based network (shown in Figure 1), 25 different clusters were identified. On the other hand, in case of Facebook (shown in Figure 2) 4 different clusters were identified.

The content of the different identified clusters was analyzed. Based on this content analysis (done by reading the content of tweets and Facebook posts), different potential sub-communities were identified (based on the themes of the discussion discovered from content analysis) and the different clusters were labeled. In case of Twitter data, it was possible to identify these potential sub communities and is shown in Table 4.

Table 4. Cluster identification based on modularity class using Twitter data

Modularity class	Cluster name
1	Personal branding
2	Employee advocacy
3	Drawings and infographics
5	**Broadly about cmadfi event
6	Community manager
7	Communications
10	*Reporting on CMADFI event
15	Customer service
16	Project
17	*Outsider greetings
18	Tekes
20	Knowledge management
21	Jyväskylän energia

The content analysis of Twitter data revealed different potential sub communities which were based on the different conversational data-based clusters (modularity class) identified. For example, modularity class 20 was related to discussion theme about knowledge management while modularity class 1 was related to theme of personal branding. On the other hand, the content analysis of the Facebook data did not reveal any specific themes and could not be used to identify any sub communities based on the different identified clusters.

Out of the total of 25 survey participants, 15 responses were received for the Question 2 related to the participants who were the most novel source of information for the survey respondent. Based on these responses the Table 5 was created. This table shows the calculated modularity class (from social media data) of the survey respondent and the most novel information sources identified by each respondent. In Table 5, column "Survey respondents" refers to the 15 individual survey participants, coded by alphabetical letters; columns I-V refer to the clusters of novel information sources identified by survey respondents. The "Modularity class number" refers to the different modularity class- based clusters of survey respondents which were identified during the analysis and can be seen in Table 4 for the Twitter data. The green color in Table 5 was used to show the novel sources which had different modularity class than the survey respondent.

Table 5. Correlating modularity class of the novel information sources with social media data

Survey Respondents ↓	Modularity Class of Respondents↓		Novel Source of Information & their Modularity Class↓									
			I		II		III		IV		V	
	T	F	T	F	T	F	T	F	T	F	T	F
L	20	1	20	0	16	1	1	1	7	0	16	1
R	7	-	5	-	16	-	2	-	5	-	20	-
Q	5	1	1	1	7	0	6	1	20	1	--	--
S	16	1	5	1	20	0	16	1	2	1	16	1
O	20	1	7	1	16	1	1	-	1	-	--	--
W	16	1	16	1	20	0	0	1	7	1	--	--
A	6	1	16	1	6	1	6	1	5	-	16	-
M	1	-	5	-	5	-	20	-	--	--	--	--
N	16	1	5	1	5	1	20	1	--	--	--	--
U	16	-	2	-	1	-	20	-	--	--	--	--
V	2	-	20	-	7	-	2	-	5	-	--	--
X	5	1	20	0	5	1	2	-	--	--	--	--
F	7	1	5	1	7	0	--	--	--	--	--	--
G	16	1	7	-	16	-	3	-	--	--	--	--
D	16	-	16	-	5	-	--	--	--	--	--	--

(green)	Different Modularity Class than the respondent
--	Did not provide a response
-	Data not available
T	Twitter data
F	Facebook data

From the Table 5, it can be seen that in total the 15 survey respondents provided a total of 55 individual novel information sources which correspond to the individual cells of the Table 5. It can be observed from the table that in case of Twitter 44 from a total of 55 individual novel information sources (approximately 80%) belong to a different modularity class. On the other hand, in case of Facebook, only 7 from a total of 55 individual novel information sources (approximately 12%) belong to a different modularity class. It can also be observed from the Table 5 that all the 55 individual novel information sources are present in the Twitter data, only 31 from a total of 55 individual novel sources are present in the Facebook data.

5. DISCUSSION AND CONCLUSIONS

In this exploratory study we followed an approach which has been followed in many previous studies to operationalize tie strength. This was then used to create two different people networks from social media conversation data of Twitter and Facebook about the event. The combination of cluster analysis of the networks complemented with the content analysis of the identified clusters helped in detecting the different theme-based potential sub communities in the CMAD 2016 event based on the Twitter network. This helped us in identifying the different structural holes based on the structure of the network. For example, from the network structure shown in Figure 1 nodes like W, N1, D2, A3 and other similar kind of nodes can be seen to be the non-redundant bridging nodes between the different identified sub communities. These nodes would satisfy Burt's structural hole definition and could be labeled as the actors bridging structural holes.

Based on the individual survey responses for the question 2, it was observed that the most novel individual sources of information for survey respondents were also the non-redundant bridging nodes based on the network shown in Figure 1. Thus, by combining the findings from Figure 1 and Table 5, the identified actors bridging the structural holes point towards the weak ties or potential weak ties in this case. These weak ties act as the bridging ties between the different potential sub communities of the whole CMAD community that were identified in Table 4.

This exploratory study used publicly available data from two different social media channels that were used in the CMAD16 event. In our case the content analysis of Twitter was helpful in identifying the actors bridging structural holes while no discernible themes could be identified from the content analysis of the Facebook data. Thus, Facebook data in this case did not provide any help in identifying the structural holes. From our findings we can state that when collecting the social media data, we should be aware of the various patterns and purposes of the use of various social media in an event: the temporal pattern of use for the specific event, the reason and context of the use of various social media channels, and also the manner in which the specific social media channels are used for event networking.

To our knowledge, this is the first study to identify actual or potential weak ties using merely publicly available social media data, as well as to do that in an event context. This study provides understanding about the use of new social media related data sources for identifying actual or potential weak ties in event context which may be used e.g. to automate the process of identifying different ties (see e.g. [29], [30]) to connect professionals in events. This introduces means to support matchmaking in a way that increases the collaborative creativity between actors with complementing interests. In more general, we also contribute to the understanding of identifying weak ties by identifying actors bridging structural holes from social media by making use of network level measures in addition to the previously used interpersonal level measures (e.g. [23], [26], [28]), as suggested by Marsden and Campbell [25]. This is the first one to make use of the structural hole theory from such data, which thus adds also to current understanding of the strength of weak ties theory originated from Granovetter's seminal work. This being an exploratory study, we used two widely used social media channels to evaluate tie strength, thus being able to see how different channels perform in the evaluation of tie strength, and finding out how various contextual factors in their use impact the usefulness and usability of public social media data in the conference context.

One of the motives of attending events such as conferences, seminars is to meet and establish ties with potentially useful contacts, the identification of weak ties may be helpful in achieving this goal. However, the usefulness of the novel information or in a more general sense, the usefulness of the identified weak ties, as such, without combining it for instance with content analysis based understanding of the interests and capabilities of identified weak ties, may vary a lot based on the kind of event it is. For example, in case of a focused conference which has a clearly defined theme and is very narrow in the range of topics that are addressed in it, the identification of weak ties itself may be sufficient to establish a potentially useful contact. As the participants in such an event are very well aware of the topic and theme of the event, suggesting an identified weak tie, already in itself, may be useful as both the participant and their weak ties would likely share the same common interests which are addressed in such a focused event. On the other hand, in the case of large events with a wide range of themes, there may be large number of weak ties which could be identified. However, in this case many of the identified weak ties, making use of the social media data, may not be relevant and useful, without e.g. prioritizing them by their identified interests, skills and expertise that are of interest to a conference participant. As the range of topics and themes addressed in such event is very large, it would be more useful to narrow down the list of all identified weak ties further to potentially the most useful weak ties. This, however, will require further analysis and use of other data sources. For example, in case of a large or relatively widely focused academic conferences (such as HICSS, CHI or Academic Mindtrek), the bibliographic academic reference data of the event participants, providing understanding of participants' research interests and capabilities, may be combined with their social media data to suggest potentially most useful weak ties. The combination of different data sources can vary based on the type of event and its context but in the case of such large events, in particular, it may significantly help in identification of potentially most useful weak ties. Conceptual design implications for a such a big data based recommender system is presented below.

5.1 Design implications for conference recommendation systems based on social data and weak tie identification

Based on the theoretical concepts in social network analysis [44] and social set analysis [45] and informed by the empirical findings from the exploratory study, our current work implications are concerned with the design, development and evaluation of a recommender system. The system seeks to support both weak ties from social relationships as well as cohorts from social associations inferred from the combined datasets of events and social media channels. Figure 3 presents the big data analytics schematics for the proposed system.

1. Systematically collect big social data about organizations and events from Facebook, Twitter etc. using the Social Data Analytics Tool [38] as well as research and commercial tools.
2. Technically combine organizational and event process data with social data so that the resulting dataset is *legally compliant*, *ethically correct, privacy adherent, and data security ensured.*
3. Big Social Data Analytics: Phase One: Adopt current methods, techniques and tools from Computational Social Science to model and analyze
   - **3.1.** Interaction Analysis: Social Network Analysis, Complex Systems Dynamics, Event Study Methodology from Finance, Data Mining from Computer Science
     - Who is doing what, when, where, how and with whom?
     - Social media users and organizational stakeholders
   - **3.2.** Conversation Analysis: Computational Linguistics & Machine Learning
     - What are the things human actors (and fraudulent accounts/robots) saying?
     - Social media users and organizational stakeholders discussing/mentioning various topics/keywords of organizational/societal relevance/irrelevance and expressing their subjective feelings etc.
4. Applying graph and set theoretical methods and techniques [45], [46]
5. Software realization of the empirical findings from traditional (graph theoretical) and novel (set theoretical) approaches to Computational Social Science as a tools for Organizations
6. Generation of instrumental benefits for organizations and individuals in terms of meaningful facts (sensible data), actionable insights (applicable information), valuable outcomes (constructive knowledge) and sustainable impacts (wisdom).

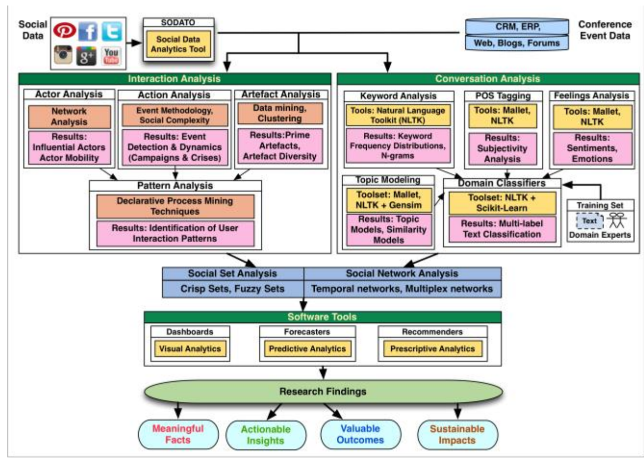

Figure 3. Schematic diagram

5.2 Limitations and Future Research

The exploratory study has certain limitations. First, in this study, we studied only some potential approaches mainly related to the analyses of conversation based networks or clusters for identifying potential weak ties. However, in the future these approaches should be used to identify different kind of ties. Second, due to the limited amount of respondents in our survey, we cannot yet draw any statistically significant results, but the results should be considered as preliminary. We will extend the survey in further event studies to increase the accuracy and generalizability of results. Third, the full data corpus for Twitter was collected after the event using the REST API of Twitter. While now focusing on social media conversational data, the above omission restricted the use of Twitter follower/followee network data in our study as we did not have any timestamp data about when a particular CMAD participant became Twitter follower of another participant. The use of this data could have enabled the operationalization of another approach for identifying the weak ties in an event setting, and we will make use of this approach in our further studies.

Despite these limitations, we want to emphasize how important the study of structural holes and bridging ties is not only to the event context but also to the wider innovation practice itself. As cross-disciplinary and cross-sectoral studies are increasingly emphasizing the role of disruptive ideas and creative-thinking that come outside of our strong and close networks [47]–[49], the identification and effective utilization of bridging ties could not only increase the amount of novel and new information but lead to new innovations as well.

In addition to the above, this study leaves room for future studies in many other areas, as well. First, there are many ways for identifying weak ties and this study has only used the concept of conversation based structural holes. This can be complemented with tie strength based weak tie identification. In future studies we will combine more dimensions and measures to evaluate tie strength in an event context. Second, this study uses data from social media channels, while the future studies can try to incorporate data from other sources like bibliographic data of scientific publications, location data and various other data sources with the social media data. Finally, incorporating big social data in the form of large collection of Twitter data and public Facebook walls of events may enable developing automated tie strength evaluation methods in case of events.

6. ACKNOWLEDGMENTS

This work has been funded by Academy of Finland (Suomen Akatemia) under grant 295893, 295894 and 295895 for the research project COBWEB.

7. REFERENCES

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

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

[3] H. Ahn and J.-H. Park, “The structural effects of sharing function on Twitter networks: Focusing on the retweet function,” *J. Inf. Sci.*, vol. 41, no. 3, pp. 354–365, Jun. 2015.

[4] A. X. Zhang, A. Bhardwaj, and D. Karger, “Confer: A Conference Recommendation and Meetup Tool,” in *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*, 2016, pp. 118–121.

[5] F. Xia, N. Yaw Asabere, J. J. Rodrigues, F. Basso, N. Deonauth, and W. Wang, “Socially-aware venue recommendation for conference participants,” in *Ubiquitous Intelligence and Computing, 2013 IEEE 10th International Conference on and 10th International Conference on Autonomic and Trusted Computing (UIC/ATC)*, 2013, pp. 134–141.

[6] M. Hornick and P. Tamayo, “Extending Recommender Systems for Disjoint User/Item Sets: The Conference Recommendation Problem,” *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 8, pp. 1478–1490, Aug. 2012.

[7] W. Reinhardt, J. Maicher, H. Drachsler, and P. B. Sloep, “Awareness support in scientific event management with ginkgo,” in *Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies*, 2011, p. 40.

[8] M. Gupte and T. Eliassi-Rad, “Measuring tie strength in implicit social networks,” in *Proceedings of the 4th Annual ACM Web Science Conference*, 2012, pp. 109–118.

[9] Y. Zhong, J. Yang, and R. Nugroho, “Incorporating Tie Strength in Robust Social Recommendation,” in *Big Data (BigData Congress), 2015 IEEE International Congress on*, 2015, pp. 63–70.

[10] H. Aramo-Immonen, J. Jussila, and J. Huhtamäki, “Exploring co-learning behavior of conference participants with visual network analysis of Twitter data,” *Comput. Hum. Behav.*, vol. 51, pp. 1154–1162, 2015.

[11] D. Fisher, B. Hogan, A. J. Brush, M. Smith, and A. Jacobs, “Using social sorting to enhance email management,” *Hum.-Comput. Interact. Consort.*, 2006.

[12] R. S. Burt, “Structural hole,” *Harv. Bus. Sch. Press Camb. MA*, 1992.

[13] R. S. Burt, “Structural Holes and Good Ideas,” *Am. J. Sociol.*, vol. 110, no. 2, pp. 349–399, 2004.

[14] D. Easley and J. Kleinberg, *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press, 2010.

[15] R. S. Burt, “The network structure of social capital,” *Res. Organ. Behav.*, vol. 22, pp. 345–423, 2000.

[16] B. A.-L. Barabási and E. Bonabeau, “Scale-free,” *Sci. Am.*, 2003.

[17] J. S. Coleman, “Social capital in the creation of human capital,” *Am. J. Sociol.*, pp. S95–S120, 1988.

[18] J. S. Coleman, “Family, school, and social capital,” *Int. Encycl. Sociol. Educ.*, pp. 623–625, 1997.

[19] P. V. Marsden and K. E. Campbell, “Measuring Tie Strength,” *Soc. Forces*, vol. 63, no. 2, pp. 482–501, Dec. 1984.

[20] D. Krackhardt, “The strength of strong ties: The importance of philos in organizations,” *Netw. Organ. Struct. Form Action*, vol. 216, p. 239, 1992.

[21] S. S. Levine and M. J. Prietula, “How Knowledge Transfer Impacts Performance: A Multilevel Model of Benefits and Liabilities,” *Organ. Sci.*, vol. 23, no. 6, pp. 1748–1766, Sep. 2011.

[22] M. E. Sosa, “Where Do Creative Interactions Come From? The Role of Tie Content and Social Networks,” *Organ. Sci.*, vol. 22, no. 1, pp. 1–21, Feb. 2011.

[23] M. Granovetter, P. V. Marsden, and N. Lin, “Social structure and network analysis,” *Soc. Struct. Netw. Anal.*, 1982.

[24] S. P. Borgatti and D. S. Halgin, “On network theory,” *Organ. Sci.*, vol. 22, no. 5, pp. 1168–1181, 2011.

[25] P. V. Marsden and K. E. Campbell, “Reflections on Conceptualizing and Measuring Tie Strength,” *Soc. Forces*, vol. 91, no. 1, pp. 17–23, Sep. 2012.

[26] 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, Sep. 2013.

[27] 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.

[28] 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.

[29] R. Xiang, J. Neville, and M. Rogati, “Modeling Relationship Strength in Online Social Networks,” in *Proceedings of the 19th International Conference on World Wide Web*, New York, NY, USA, 2010, pp. 981–990.

[30] I. Kahanda and J. Neville, “Using Transactional Information to Predict Link Strength in Online Social Networks.,” *ICWSM*, vol. 9, pp. 74–81, 2009.

[31] Y. Liu, J. Venkatanathan, J. Goncalves, E. Karapanos, and V. Kostakos, “Modeling What Friendship Patterns on Facebook Reveal About Personality and Social Capital,”

ACM Trans Comput-Hum Interact, vol. 21, no. 3, p. 17:1–17:20, Jun. 2014.
[32] M. Rezvani, W. Liang, W. Xu, and C. Liu, “Identifying Top-k Structural Hole Spanners in Large-Scale Social Networks,” in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, New York, NY, USA, 2015, pp. 263–272.
[33] Zhan Shi, Huaxia Rui, and A. B. Whinston, “Content Sharing in a Social Broadcasting Environment: Evidence from Twitter,” *MIS Q.*, vol. 38, no. 1, pp. 123-A6, Mar. 2014.
[34] S. Aral and D. Walker, “Tie strength, embeddedness, and social influence: A large-scale networked experiment,” *Manag. Sci.*, vol. 60, no. 6, pp. 1352–1370, 2014.
[35] A. Wu, J. M. DiMicco, and D. R. Millen, “Detecting Professional Versus Personal Closeness Using an Enterprise Social Network Site,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2010, pp. 1955–1964.
[36] W. Reinhardt, M. Ebner, G. Beham, and C. Costa, “How people are using Twitter during conferences,” *Creat. Innov. Competencies Web Proc. 5th EduMedia*, pp. 145–156, 2009.
[37] H. Aramo-Immonen, H. Kärkkäinen, J. J. Jussila, S. Joel-Edgar, and J. Huhtamäki, “Visualizing informal learning behavior from conference participants’ Twitter data with the Ostinato Model,” *Comput. Hum. Behav.*, vol. 55, pp. 584–595, 2016.
[38] A. Hussain and R. Vatrapu, “Social data analytics tool (sodato),” in *Advancing the Impact of Design Science: Moving from Theory to Practice*, Springer, 2014, pp. 368–372.
[39] 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*, Springer, 2014, pp. 99–118.
[40] A. Petróczi, T. Nepusz, and F. Bazsó, “Measuring tie-strength in virtual social networks,” *Connections*, vol. 27, no. 2, pp. 39–52, 2007.
[41] M. Bastian, S. Heymann, M. Jacomy, and others, “Gephi: an open source software for exploring and manipulating networks.,” *ICWSM*, vol. 8, pp. 361–362, 2009.
[42] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, “Fast unfolding of communities in large networks,” *J. Stat. Mech. Theory Exp.*, vol. 2008, no. 10, p. P10008, 2008.
[43] J. Jussila, J. Huhtamäki, K. Henttonen, H. Kärkkäinen, and K. Still, “Visual Network Analysis of Twitter Data for Co-organizing Conferences: Case CMAD 2013,” in *2014 47th Hawaii International Conference on System Sciences*, 2014, pp. 1474–1483.
[44] E. Otte and R. Rousseau, “Social network analysis: a powerful strategy, also for the information sciences,” *J. Inf. Sci.*, vol. 28, no. 6, pp. 441–453, Dec. 2002.
[45] R. Vatrapu, R. R. Mukkamala, A. Hussain, and B. Flesch, “Social Set Analysis: A Set Theoretical Approach to Big Data Analytics,” *IEEE Access*, vol. 4, pp. 2542–2571, 2016.
[46] N. M. Tichy, M. L. Tushman, and C. Fombrun, “Social network analysis for organizations,” *Acad. Manage. Rev.*, vol. 4, no. 4, pp. 507–519, 1979.
[47] E. Enkel, O. Gassmann, and H. Chesbrough, “Open R&D and open innovation: exploring the phenomenon,” *Rd Manag.*, vol. 39, no. 4, pp. 311–316, 2009.
[48] C. Baldwin and E. von Hippel, “Modeling a paradigm shift: From producer innovation to user and open collaborative innovation,” *Organ. Sci.*, vol. 22, no. 6, pp. 1399–1417, 2011.
[49] L. Dahlander, L. Frederiksen, and F. Rullani, “Online communities and open innovation,” *Ind. Innov.*, vol. 15, no. 2, pp. 115–123, 2008.