Evaluating Association between Research and Social Networks

Mohammed N. Al-Kabi
Computer Science Department
IT Faculty
Zarqa University
P. O. Box 2000
13110 Zarqa -Jordan
malkabi@zu.edu.jo

Heider A. Wahsheh
Computer Science Department
College of Computer Science
King Khaled University
Abha, Saudi Arabia
heiderwahsheh@yahoo.com

Izzat M. Alsmadi
Computer Science Department
University of New Haven
West Haven, CT 06516, USA
ialsmadi@newhaven.edu

Abstract: *Web 2.0 enables its users to generate their contents which lead to the emergence of a new era of online social networks. Nowadays the percentage of user/humans of online social network exceeds 50%, where the top 3 online social networks are Facebook, LinkedIn, and Twitter respectively. Twitter is among the top 10 websites worldwide on the Web, with more than 300 million active users. This study aims to discover whether there is any correlation between research networks and social networks. The motivation behind this paper is to discover whether the researchers in Jordan use Social Networks that are designed for general social purposes like Facebook, Google+, etc. as tools of communication to discuss topics related to their specialty. More specifically whether IT researchers in Jordan use both Facebook and Twitter as part of their research networks. Google Scholar is used to identify IT researchers in Jordan. Networking information between those researchers is collected from Facebook and Twitter. Results showed that while most of those researchers have individual pages in social networks, yet those networks or accounts are largely used for social, possibly professional but not research purposes. While there are some other Social Networks such as ResearchGate and LinkedIn which are professional networks used by academics and students, however, researchers in general should have a better usage of social networks.*

Keywords: *Social networks; Research networks; Twitter; Facebook; LinkedIn; ResearchGate.*

1. Introduction

The World Wide Web (W3) evolved since its release in the early 1990s, where it starts with specialized hidden Websites for researchers and end up nowadays with open access Web sites. The current Web enables its users to generate their own contents that include text, audio, video and images. Web 2.0 hosts online communities such as social networks, Wikis, forums, blogs, etc. The rapid development of Web 2.0 makes it a huge reservoir of data that can be exploited to mine customers' sentiments and opinions, public's point of views about different aspects of life. The large spectrum of Web 2.0 and social media contents express users' opinions and sentiments. It includes blog articles, recommendations, online reviews, discussion forums, and different types of social media.

Micro-blogging service; Twitter.com is ranked 10th on the list of the top 500 sites on the Web according to Alexa (www.alexa.com; retrieved: June 6th, 2015). It is a well-known social network website that is used by many people around the world. Twitter allows people to communicate with each other using text messages. These text messages were limited to 140 characters, but since July 2015 this limit is increased to be 10000 (10K) characters. Several studies have been conducted regarding the use of Twitter to disseminate information in scientific conferences and the level of usage of social networks in general for research purposes [1-3].

This study aims to discover whether there is any correlation between real life research networks and their social networks. A sample dataset of IT researchers in Jordanian universities was collected. Google Scholar is used to identify the IT researchers working in 26 Jordanian universities. Those researchers who have no Google Scholar accounts are not included in this study.

The remaining of this paper is organized as follows: In section 2, we present a review of related work. Section 3 presents a summary for the used dataset, while section 4 presents the adopted methodology. Section 5 shows the experiment and results. Section 6 presents our conclusions and future plans in this direction.

2. Related Work

Sentiments analysis studies of Arabic reviews and comments witnessed an increase during the last five years. This section first presents some of these studies, and we could not find in a large list of studies any that investigate the correlation between research networks and social networks in the Arab countries. Rushdi-Saleh, Martín-Valdivia, Ureña-López, and Perea-Ortega [4] constructed an Opinion Corpus for Arabic (OCA), translate all Arabic reviews to English, and then they use machine learning algorithms to identify the polarities of these reviews.

Al-Kabi, Gigieh, Alsmadi, Wahsheh, and Haidar study [5] builds a specialized sentiment analysis tool for the Arabic language, whether the reviews are written in MSA or a dialectical Arabic. They used this tool to identify the polarity of different Arabic reviews and comments. This team improves their tool later [6], so that it can determine whether the Arabic comment or review is subjective or objective. In addition to its capability to determine the strength of Arabic comments/reviews.

Al-Kabi, Al-Qudah, Alsmadi, Dabour, and Wahsheh study [7] used two free online tools (SocialMention and Twendz) to identify the polarity of the 4050 Arabic and English reviews they were collected. These collected reviews include Emoticons. They conclude that SocialMention tool is more effective than Twendz tool. A similar study to [7] is conducted by [8] that includes only Arabic reviews and comments collected from two social network websites (Facebook and Twitter).

Al-Kabi, Abdulla, and Al-Ayyoub [9] constructed a labeled dataset of Arabic reviews/comments collected from Yahoo!-Maktoob web site. They present results of their analysis to different aspects of the collected Arabic reviews *K*-nearest neighbour classifiers) to determine the polarities of Arabic reviews. Furthermore, they study the effect of t and use them to test two supervised classifiers (Support Vector Machine (SVM), and Naive Bayes (NB)).

The effect of stemming, feature correlation and n-gram models on Arabic sentiment analysis are studied by Duwairi and El-Orfali study [10]. They investigate the effectiveness of three classification algorithms (Support Vector Machine, Naive Bayes, and *K*-nearest neighbour classifiers) to determine the polarities of Arabic reviews. Furthermore, they study the effect of the nature of the Arabic dataset on sentiment analysis.

Many social network analysis studies in social sciences were conducted 50 years before the emergence of Web 2.0 and online social media era like those conducted by [11-14]. One study is conducted by Watts, and Strogatz [12] dedicated to a small world network of movie actors where the appearance of actors in a single movie indicates that they know each other.

One of the pre-Web 2.0 era studies that investigate the structure of scientific collaboration networks is conducted by Newman [14], where two or more researchers are considered connected if they have a joint scientific work. Newman [14] uses a number of specialized databases, such as MEDLINE, NCSTRL, etc. to construct explicit networks. His study presents a network of human acquaintances that contains more than one million people, where any two researchers are linked if they participate in at least one research. He constructed four collaboration graphs for scientists in four different fields.

Ebner and Reinhardt study [15] tries to discover how Twitter can serve scientific communities and conferences. Furthermore, this aims to find whether Twitter adds any scientific value to conferences. To achieve its goals the use of Twitter at ED-MEDIA 2009 e-Learning conference is explored. The statistical analyses of tweets using ED-MEDIA hashtag (#edmedia) reveal an arbitrary growth in both the number of tweets and the number of Twitter participants during the conference.

Another study to show how this microblogging tool (Twitter) can be used during conferences is conducted by Reinhardt, Ebner, Beham, and Costa [16]. Those authors show in their study how Twitter enhance the knowledge of Twitter groups and communities, by presenting topics and exchanging information. They found that Twitter is used by conference speakers and attendees for various purposes, and mainly it is used to communicate and share resources. Furthermore, they found that Twitter is used in conferences by few participants to follow parallel sessions.

Letierce, Passant, Decker and Breslin [17] study is based on a survey that showed Twitter among the top four most popular applications (Twitter, personal email, Skype and project mailing lists) used for disseminating information, and showed that Twitter is among the top three applications that are used by Semantic Web researchers to disseminate information. The main goal of their study is to know how Twitter is used for disseminating scientific messages by the particular community (Semantic Web researchers). They targeted Semantic Web conferences since most conferences offer official hashtags that they disseminate via their twitter accounts, websites or brochures, and since conferences indicate a particular timeline when such scientific content can be shared on Twitter. They found that users who have authority during a conference get a high authority score or both a high authority and hub value score on Twitter. Furthermore, they discover that the researchers used messages targeted to peer researchers.

Hadgu and Jäschke study [18] presents a method to identify Twitter accounts of computer science researchers, and demonstrate how they use of Twitter for computer science discipline. Seeds set of computer science conferences was used to collect Twitter users used in their study. They empirically analyze the data of identified computer science users such as their popularity, age, and influence. They generate some features such as tweets profile, friend/follower ratio, tweets with a URL and tweet content. They used three classification methods Support Vector Machines (SVM), Random Forest (RF) and Regression Trees (CART) to evaluate their approach that yields accepted results.

3. Research-Social Network the Dataset

Google Scholar is used in this study to collect the main seeds for our dataset as one of the main powerful sources about researchers and publications. On June 21st, 2015 we found there are 2190 researchers who

are affiliated to Jordanian universities and have accounts on Google Scholar. Those who are IT researchers constitute only 192 researchers out of total 2190. IT researchers in this study are those who are in Google Scholar and are also affiliated to IT faculties, and those who are affiliated with Computer Engineering, Computer Networks, Computer and Network Security, Electrical Engineering, and Management Information Systems departments.

Table 1 shows the list of 26 Jordanian universities sorted in descending order according to the number of IT Google scholars. This list includes 113 researchers that are affiliated with 11 public universities, and 79 researchers that are affiliated with 15 private universities. That means 58.8% (113/192 %) of those researchers are affiliated with public universities, and 41.2% (79/192 %) of those researchers are affiliated with private universities.

Table 1. Distribution of IT Jordanian Researchers among the 26 Universities.

I	University Name	No. of Researchers	Type
1	The University of Jordan (JU)	26	Public
2	Jordan University of Science and Technology (JUST)	22	Public
3	Yarmouk University (YU)	22	Public
4	Al-Ahliyya Amman University (AAU)	18	Private
5	Balqa Applied University (BAU)	15	Public
6	Princess Sumaya University for Technology (PSUT)	15	Private
7	Hashemite University (HU)	8	Public
8	Applied Science University (ASU)	7	Private
9	Isra University (IU)	6	Private
10	Jerash University (JPU)	5	Private
11	Al-Zaytoonah University of Jordan (ZUJ)	5	Private
12	Zarqa University (ZU)	5	Private
13	University of Petra (UOP)	5	Private
14	World Islamic Sciences and Education University (WISE)	4	Public
15	German-Jordanian University (GJU)	4	Public
16	Mutah University (MU)	3	Public
17	Al-Hussein Bin Talal University (AHU)	3	Public
18	Tafila Technical University (TTU)	3	Public
19	AL- al Bait University (AABU)	3	Public
20	Middle East University (MEU)	3	Private
21	Philadelphia University (PUJ)	3	Private
22	Amman Arab University (AAUJ)	2	Private
23	Jadara University (JPU)	2	Private
24	Irbid National University (INU)	1	Private
25	Ajloun National University (ANU)	1	Private
26	American University of Madaba (AUM)	1	Private
	Total	192	

It shown in our methodology, we search for each researcher in Twitter and Facebook social networks to determine whether the researcher under consideration has an account in the two social networks or not. Our dataset is divided into two subsets; Facebook and Twitter. Table 2 presents the summarization of researcher's information list.

Table 2. Researchers' Information List

Social Network	Researchers without accounts	Undetermined Accounts	Private Accounts	Public Accounts
Facebook	79	21	37	55
Twitter	134	12	11	35

The researchers list presented in table 2 is divided into four parts; some researchers do not have social networks accounts. We were unable to determine if some accounts were actually belonging to the researchers under study. Although some accounts were publicly accessible, other accounts were private and not accessible due to privacy constraints. Our dataset was based on public accounts which were retrieved in July 2015.

Manually for each researcher we extracted the list of Twitter followers and followings, as well as the list of friends in the Facebook.

In order to find if the researcher appears in the list of follower and following/ friend list, we develop researcher friend tool (RFT). RFT is capable of finding the intersection set between researchers list and researcher's friends/follower/ followings set. RFT computes the ratio of this intersection over the researcher's friends. Figure 1 presents RFT algorithm.

Input:
R: Researcher name.
GL: Google Scholar list.
RL: Researcher list.
FL: Friends/Follower/ Followings list.
E: Element in the list.

Output:
FL′: intersection set between researchers list and researcher's friends/follower/ followings set.

Initialization:
FL′ = 0.

Begin
1: For each *R* ϵ *GL* has *FL*
2: For each *R*, compare *E* of *FL* with *RL*:
3: If E ϵ *RL* Then:
4: *FL′= FL′+ E*
5: Else
6: Next *E*
7: End For
8: Write FL′ to file
9: End for
End

Figure 1. RFT Algorithm.

4. Methodology

The main objective of this study is to evaluate the correlation between research networks and social networks. To achieve this goal data is collected from

Facebook and Twitter social networks and is studied and analyzed in accordance with data about the same individuals from Google scholars. We proposed the following phases in our methodology:

1. Collect the data manually of 192 IT researchers in Jordan from Google Scholar Database. These researchers are considered as the main seeds' set to start building our dataset.
2. Conducting a social network (Facebook & twitter) search using the constructed seeds' set to find first whether researchers have accounts on (Facebook & twitter) or not.
3. For each researcher account in the above two social networks, we extracted the list of followers and followings (based on Twitter), as well as the list of friends (based on Facebook) within available users' privacy considerations.
4. Develop researcher friend tool (RFT), which is capable of retrieving the intersection set between researchers list and researcher's friends/follower/ followings sets.
5. Evaluate the intersection sets using a statistical software package (SigmaStat version 3.5) to evaluate the statistical measures (e.g. correlations, associations) with significant levels.

5. Experiment and Results

Our experiments aim to evaluate the correlation between research networks and social networks, using RFT outputs with SigmaStat statistical software version 3.5.

SigmaStat is a statistical software package that compares effects among groups. SigmaStat analyzes the data, performs the proportions, and conducts the regression and correlation analysis [19]. The statistical significance was declared when *P*-value equals to or less than (≤) 0.05.

Three statistical methods are applied using SigmaStat software; Kruskal–Wallis one-way analysis of variance, Newman–Keuls method, and Pearson product-moment correlation coefficient test. The Kruskal–Wallis one-way analysis of variance by ranks is a non-parametric test to check if the groups are constructed from the same distribution or not. This test compares two or more independent groups, which may have variance in their sizes. This test aims to analyze the specific group pairs for stochastic dominance [20].

The Newman–Keuls method is used to find sample means that significantly different from each other among multiple comparisons. This method uses different critical values for different pairs of mean comparisons [21]. Pearson product-moment correlation coefficient is a linear correlation measurement between two variables: *X* and *Y* [22].

Table 3 shows the results yielded from applying Kruskal–Wallis one-way analysis of variance by ranks on Facebook.

Table 3. The results of Kruskal-Wallis One Way Analysis of Variance on Ranks

Group	N	Missing	Median	25%	75%
Friend as Researcher (FR)	55	0	0.457	0.000	3.331
Facebook Friend (FF)	55	0	99.543	96.669	100.000
$H = 81.786$ with 1 degrees of freedom. ($P \leq 0.001$)					

The results of Kruskal–Wallis one-way analysis of variance by ranks showed the differences in the median values among the treatment groups. The gained (P-value ≤ 0.001), which indicates that there is a statistically significant difference, so all or part of research network partners should be also Facebook friends.

Table 4 presents the results of Newman–Keuls method, where q value represents the difference between two sample means and dividing it by the standard error.

Table 4. The results of Newman–Keuls method

Comparison between two groups	*q value*	*P-value*
FR % vs FF %	12.749	< 0.05

Newman–Keuls *P*-value results indicate positive answer that all or part of research network partners should be also Facebook friends.

Applying Pearson Product Moment Correlation measure on Twitter yields a *P*-value greater than 0.05 that indicates that there is no significant relationship between research network partners and Twitter follower/following groups.

A study [18] is considered to the closest to our study, and it used ten features that include friend/follower ratio, and applied classification methods to evaluate their approach. Therefore, we compare the performance of their supervised algorithms that classified the relation features on the social network and with the performance of our classification algorithm. The Classification in this study is based on Facebook data only, due to the scarcity of twitter data. Therefore, positive results indicate that all or part of research network partners should be also Facebook friends.

Table 5 shows the performance comparison of the algorithms, where we use Accuracy, Precision and Recall measurements. The formulas of these three metrics are shown below:

$$Accuracy_i = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$Recall_i = \frac{TP}{TP + FN} \quad (2)$$

$$Precision_i = \frac{TP}{TP + FP} \quad (3)$$

Where the four main elements are: True Positive (*TP*), True Negative (*TN*), False Positive (*FP*) and False Negative (*FN*).

Table 5. Comparison of Algorithms Performance.

Algorithm	***Accuracy***	***Precision***	***Recall***
Our SVM	66.6%	0.667	0.667
SVM in [18]	91%	0.9	0.89
Our Logistic Regression	50%	0.4	0.5
Logistic Regression in [18]	89%	0.88	0.87

We found out that information or features collected in this study are not enough to make significant judgements on the kind of possible correlation between research and social networks.

6. Conclusion and Future Work

In Internet social networks different individuals interact with each other and exchange messages, videos, images, etc. The relations between those individuals can come from real life relations (e.g. relatives, co-students, employees, friendship, spouse) or can from the Internet or virtual world.

Academics and students work together in research groups or teams where those teams can exist in different physical locations. The new telecommunication tools make it possible to effectively communicate with messages, text, audio, video and with high or real time streaming.

Previous research studies showed conflicting results on the amount of common interactions between researchers using social networks such as Twitter and Facebook. We conducted a study to evaluate the level of interactions between IT researchers in Jordan in social networks, particularly, Twitter and Facebook. Results showed that in social networks communication between researchers of the study group is not significant. Results showed also that collected data does not include enough features to have strong confidence on our research findings.

This study relied on Facebook and Twitter that are designed for general social purposes. Therefore, in the future we plan to explore other general purpose social networks, and collect data about more researchers. It is useless to collected data from ResearchGate and LinkedIn, since it is designed for research social network and the users put their research papers attached by their co-authors. Furthermore, we plan to extend our dataset in future in two dimensions; size and features about researchers.

References

[1] Alexa Top 500 Global Sites. Available at: http://www.alexa.com/topsites [Online; accessed June 06 - 2015].

[2] Social Networking Statistics. Available at: http://www.statisticbrain.com/social-networking-statistics/ [Online; accessed June 04 - 2015].

[3] Twitter Usage/ Company Facts. https://about.twitter.com/company [Online; accessed June 04 - 2015].

[4] Rushdi-Saleh M., Martín-Valdivia M. T., Ureña-López L. A., and Perea-Ortega J. M., "Bilingual Experiments with an Arabic-English Corpus for Opinion Mining." In Proceedings of Recent Advances in Natural Language Processing, 2011, pp. 740–745.

[5] Al-Kabi M., Gigieh A., Alsmadi I., Wahsheh H., and Haidar M. "An opinion analysis tool for colloquial and standard Arabic." In The Fourth International Conference on Information and Communication Systems (ICICS 2013), 6 pages. Irbid, Jordan, (April 23-25, 2013).

[6] Al-Kabi M. N., Gigieh A. H., Alsmadi I. M., Wahsheh H. A., Haidar M. M., "Opinion Mining and Analysis for Arabic Language." International Journal of Advanced Computer Science and Applications (IJACSA), SAI Publisher, 5(5), 2014, pp. 181-195.

[7] Al-Kabi M., Al-Qudah N. M., Alsmadi I., Dabour M., Wahsheh H., "Arabic / English Sentiment Analysis: An Empirical Study." In The Fourth International Conference on Information and Communication Systems (ICICS 2013), 6 pages. Irbid, Jordan, (April 23-25, 2013).

[8] Khasawneh R. T., Wahsheh H. A., AL-Kabi, M. N., Alsmadi I. M. "Sentiment Analysis of Arabic Social Media Content: A Comparative Study." The 8th International Conference for Internet Technology and Secured Transactions (ICITST-2013), December 9-12, 2013, London, UK, 6 pages.

[9] Al-Kabi M.N., Abdulla N.A., Al-Ayyoub M., "An analytical study of Arabic sentiments: Maktoob case study." In Proceedings of 8th International Conference for Internet Technology and Secured Transactions (ICITST), 2013, 9-12 Dec. 2013, pp.89-94.

[10] Duwairi R.; El-Orfali M. "A study of the effects of preprocessing strategies on sentiment analysis for Arabic text." Journal of Information Science, vol. 40 (4), 2014, pp. 501-513.

[11] Wasserman S., & Faust K., "Social Network Analysis: Methods and Applications." Cambridge: Cambridge University Press (1994).

[12] Watts D. J., and Strogatz S. H., "Collective dynamics of 'small-world' networks." Nature 393, (4 June 1998), pp. 440-442.
[13] Scott J., "Social Network Analysis: A Handbook." SAGE Publications Ltd; 2nd edition (March 1, 2000), ISBN-13: 978-0761963394.
[14] Newman M. E. J., "The structure of scientific collaboration networks." In Proceedings of the National Academy of Sciences USA 98 (2001), pp. 404–409.
[15] Ebner M., and Reinhardt W., "Social networking in scientific conferences - Twitter as tool for strengthen a scientific community." In Proceedings of the 1st International Workshop on Science 2.0 for TEL (EC-TEL 2009) Nice, France. Berlin: Springer, October 2009.
[16] Reinhardt W., Ebner M., Beham G., Costa C., "How people are using Twitter during conferences." In Proceedings of the 5th EduMedia conference (Salzbrug, Austria, 2009), pp. 145-156.
[17] Letierce J., Passant A., Breslin J. G., "Understanding how Twitter is used to spread scientific messages." In Proceedings of the Web Science Conference 2010, Raleigh, NC, USA, (26--27 April 2010), pp. 1-8.
[18] Hadgu A. T., Jäschke R., "Identifying and analyzing researchers on twitter." In Proceedings of the 2014 ACM conference on Web science (WebSci '14), ACM, New York, NY, USA, 2014, pp. 23-32.
[19] SigmaStat, Retrieved July, 25, 2015, https://en.wikipedia.org/wiki/SigmaStat
[20] Kruskal–Wallis one-way analysis of variance, Retrieved July, 27, 2015, https://en.wikipedia.org/wiki/Kruskal%E2%80%93Wallis_one-way_analysis_of_variance
[21] Newman–Keuls method, Retrieved July, 27, 2015, https://en.wikipedia.org/wiki/Newman%E2%80%93Keuls_method
[22] Pearson product-moment correlation coefficient, Retrieved July, 27, 2015, https://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient