Analyzing Tie-Strength across Different Media

Nils Jeners[1], Petru Nicolaescu[1], and Wolfgang Prinz[2]

[1] RWTH Aachen University, Informatik V, Aachen, Germany
{nils.jeners,petru.nicolaescu}@rwth-aachen.de
[2] Fraunhofer FIT, Sankt Augustin, Germany
wolfgang.prinz@fit.fraunhofer.de

Abstract. Human interactions are becoming more and more important in computer systems. The interactions can be classified according to predefined rules, based on the assumption that relations between humans differ greatly. The tie-strength notion is used throughout the social sciences literature in order to denote this classification. The present paper researches how the tie-strength between two persons can be computed automatically by applying structural data from different sources, e.g. email, shared workspaces. This data can provide a virtual copy of the ego-centric network of a user and therefore be utilized for social intelligent computing.

Keywords: Social media, social networks, email, shared workspaces, relationship modeling, tie-strength, sns.

1 Introduction

The basis for social intelligent computing is to understand human interactions, how communication and collaboration works. With the emergence and spreading of IT systems, the communication channels evolved rapidly. There are different types of communication mechanisms and many applications that support the communication, collaboration and coordination between people. Different channels such as telephone, email, social applications, shared workspaces, blogs, and forums are used nowadays to exchange professional and personal information. The user base for many of these systems is very large and is still growing rapidly.

There is a lot research in particular systems (such as social networking: Facebook [1], Twitter [2]; email [3], [4]; instant messaging and chat systems [5]; and shared workspace applications [6]) that are specialized in offering stable and well-designed functionalities (e.g., feeds, handling emails, instant messaging, handling video and image media, etc.). Providing unification of those different services have now started. However, there are still gaps containing important and useful aspects that should be considered in this regard.

Relationships between humans are the core of social systems. Their type differs across environments and depending on the people involved in a relationship. In the technical comprehension of this paper, two people have a relationship if they are exchanging information. Measuring the relationship intensity leads to the tie-strength

P. Herrero et al. (Eds.): OTM 2012 Workshops, LNCS 7567, pp. 554–563, 2012.

[7], the term that describes the mutual closeness between two people. Many different applications are used for communication or collaboration, but they all share some similarities, such as a sender, receiver and a message. Such applications allow humans to build relationships implicitly by communication or explicitly by asking for friendship. With the exchange of information or emotions, we build our own egocentric social network, embedded within other bigger networks.

The various channels used for social interactions contain much more data than the actual sender, receiver and the message, e.g. structural information (tags, priority). By managing this data and connecting it to a user interface in order to visualize it, it would be possible to support users in their communication and cooperation and increase the productivity. Our approach helps to understand the social environment of a user to form social intelligent user interfaces. The presented prototype monitors different communication channels, such as email and a shared workspace system, mines the gathered data and presents the results in an understandable and intuitive format.

The benefits that result from measuring the relation between persons (tie-strength) are multiple, such as:

1. Provide social intelligent user interfaces
2. Assessing personal or work relations
3. Identifying communication weaknesses between persons.
4. Identifying communities of practice.
5. Improving the media selection based on the media related tie-strength.

Considering the above mentioned facts, we build an ego-centric social network by identifying and weighting the most important communication factors of the media channels used by a person. The resulted information can be analyzed and the results containing the communication actors relevant to the person can be displayed in a simple and meaningful way. In order to better explore users' preferences and opinions on this particular topic, a questionnaire and user interviews were conducted. Moreover, for proving and evaluating the ideas described, a user-oriented software system has been designed and implemented.

2 Background and Related Work

The definition of a relation can be seen as an exchange of information between two people. A relation is grounded on communication and can have a certain quality [7]. Tie-strength can be defined as the tightness (intensity) of that relation. Examples of strong ties are close friends and examples of weak ties are acquaintances or co-workers which do not work on the same project.

The measuring of relations or the quantification of tie-strength is a question that has already attracted much research [8] [9] [10] [11] [12] [13] [14]. Hanneman and Riddle [15] presented scales of relation measurement: Binary measures, multiple-category nominal measures (e.g., colleague, business relation, husband, friend, etc.), grouped ordinal measures (which can be encoded as 1, -1 and 0), full-rank ordinal measures, interval measures.

Choosing an interval measure appears to be appropriate for a scenario, based on computers and multiple media channels. Granovetter [7] proposed four dimensions of tie-strength: amount of time, intimacy, intensity and reciprocal services. He later extended the list by factors such as network topology, emotional support, education, race, and gender, etc. [16]. Gilbert and Karahalios [1] give a strong indication that in practice the interaction frequency, recentness of communication and communication reciprocity can accurately determine the tie-strength. The *frequency* of contacts is the number of the interactions between two people in a specific period of time. Granovetter proposes a scale from ”often” as being at least twice a week, to ”rarely”, representing less than once per year. ”Occasionally” is meant to be between the two. A high degree of *frequency* leads to a higher tie-strength. *Reciprocity* refers to the direction of the communication. This can be a one-way communication with one sender and one receiver, or a two-way communication where the sender and receiver alternate mutually. A high degree of *reciprocity* leads to a higher tie-strength. *Recentness* means the age of the last exchanged message. A high degree of *recentness* leads to a higher tie-strength.

Shared workspaces research concerning social network data analysis is mostly dealing with the objects present in one’s workspace and with people collaborating on the objects. Such an approach is presented in the paper of Nasirifard et al. [6], treating the extraction of an object-centric and a user-centric social network using the log files of a shared workspace system. The user-centric social network is computed by pairing the contacts which interact upon the same object. After building the network, the edges are weighted in order to determine the *CooperationIndex* between the users.

The background and related work of this field shows us how versatile the dimensions of tie-strength are. With the help of previous research we could apply the compatible dimensions to our given scenario. *Frequency*, *recentness* and *reciprocity* appear in every communication and cooperation media, so that a model can easily be adapted to heterogeneous media channels. Furthermore we have seen how networks are extracted from the various communication channels and how to overcome technical drawbacks. Therefore we could learn from existing research to go further and develop a new model.

3 Conceptual Model

In order to compute the interaction strength, a formalization of the notions involved was needed to establish the most determinant factors and to ease data handling. The mathematical framework can be used by integrated new media channels that might be added to the model to compute tie-strength.

In the following we will describe the final model with its components:

- $C = \{c_1, c_2, \ldots, c_n\}$ is denoted as the set of contacts that will be assigned to a user u.
- $M = \{m_1, m_2, m_3\}$ is denoted as the set of media channels.
- Each media channel is characterized by one or more objects. The objects are part of media channels.

- $O_{m_i} = \{o_1, o_2, \ldots, o_p\}$ is denoted as the set of objects of medium $m_i \in M$.
- Objects are the result of actions of persons. These actions determine the properties characteristic to each object and can be seen as events. Examples of different types of events are: incoming emails, uploaded documents etc.
- $E_{m_i} = \{e_1, e_2, \ldots, e_r\}$ is denoted as the set of events of medium $m_i \in M$. The importance of an event has been considered to be time dependent. Therefore, the impact value v of one event from medium m at time t can be described as:

$$v_m(t) = \lfloor v_m(0) * f(t) \rfloor,$$

  where $v_m(0)$ is the initial value of the event on medium m and $f(t)$ is the aging function which results an aging factor at a given time t.
- Based upon the events occurring between two or more contacts on the same object, relations can be described. Relations are characterized by attributes such as strength (which we want to determine). Two contacts $(p_1; p_2)$ are considered to be in a relation if at any time t, at least one event took place on the same object, involving p_1 and p_2.

The tie-strength (which is also referred to the interaction or relation strength), is considered to be a combination of various measurable properties of the events and objects exchanged using different electronic media channels and which defines the intensity of a certain type of relation between two persons. In this context, we study the various types of relations in daily life by analyzing the available electronic actions that a person initiates with one or more of his contacts. As we could see in the theoretical notions exposed, the tie-strength can be predicted in some extent. The questions which need to be answered are: How can the analysis of online interactions and their environment influence the tie-strength? Which are the major indices to measure the tie-strength and how are the indices mapped to the features of depicted media channel? The properties considered in this computation, as well as more information about the procedures used for each media channel that was taken into consideration will be detailed in the following.

Because computing the tie-strength is the main focus of this paper, the relations component plays a very important role in the overall architecture of the AMICIS system. Analysis has been performed in order to identify the best way in which the overall strength can be computed. The final approach, taken after analyzing different options (i.e., linear vs. non-linear equations), was chosen in the form of a linear equation, the overall tie-strength being the sum of the individual tie-strengths computed for each media channel. The individual tie-strength components are then computed as the weighted sum of the relevant attributes specific to each media channel. Moreover, as it can be observed in the equation below, the tie-strength is a function dependent on time, its value decreasing with the time passing. Therefore, the overall tie-strength is considered to be the sum of the calculated tie-strength of each media with a certain aging effect.

3.1 Email

The computation of the individual tie-strength for email was done by identifying the influential properties of emails and applying them as factors of the tie-strength model. In order to compute a value which represents the tie-strength between two people, the first step is to identify the communication partners and then calculate the tie-strength between them. The relations in email can directly be deduced from the fields *from, to, cc* (and *bcc*). This makes eight possible (not-directed) relation types: *from*-{*to*, *cc*, *bcc*}, *to*-{*to*, *cc*, *bcc*}, and *cc*-{*cc*, *bcc*}.

For the calculation of the tie-strength, we assume that:

- Recent emails have a greater significance than older emails (*date*).
- The *direction* of email communication can be unidirectional (one-way) or bidirectional (two-way). Bidirectional emails reflect a higher importance. Unidirectional emails might be noise, e.g. newsletters and notification mails.
- Email *threads* are emails between two or more people with the same subject. Longer *threads* reflect a higher degree of interest.
- The total *number of emails* exchanged between two people in a certain period of time may reflect a strong relation between the two people.
- The lack of communication between two people for a longer period of time hint to a weakening of the relation or to a weak tie in general.

3.2 Shared Workspace

For our analysis we used a shared workspace system. It contains of different types of information such as documents, calendar, URLs, threaded discussions, and member profiles. The content is organized in a folder hierarchy, which can be used to grant access to a specific group of people. A typical example for such a shared workspace is the BSCW system [17].

The relations of a shared workspace system can be deduced by transforming the document-centric network in a people-centric one [6]. The three roles are depending on events respectively activities (*write, read*) on a certain object. *Creator, modifier,* and *reader* perform activities with an object and thus they have a relation to the other actors of this object.

For the calculation of tie-strength for the BSCW system, we assume that:

- A recent interaction on an object has greater significance than older ones (*date).*
- The *direction* indicator can range from non-directional (*read-read*), over unidirectional (*write-read*) to bidirectional (*write-write*). Bidirectional interactions lead to collaboration and reflect a higher degree of involvement, e.g. a direct work relation such as colleague-colleague relation within a team or a supervisor-subordinate relation.
- The *frequency* is the total number of events in a certain period of time of a certain document. A high frequency reflects a strong relation between the users.

- The lack of events respectively actions on an object for a longer period of time may hint that the object has become obsolete, i.e. the collaboration has stopped or rather paused. Therefore it points to a weak tie between the collaborators.

The data used in order to unify different media channels has been analyzed in order to find a model that would correspond to unifying social relationship data from various sources.

The main sources are represented each by an entity with the main attributes relevant to the tie-strength extraction. The objects belonging to each of the sources (e.g., email, BSCW document, etc.) determine an event. An event involves a relation between two contacts and it is characterized by a strength (which we have to determine). The relation between the contacts is directed, having an emitter and a receiver, depending on the initiator of the event and on its target. Events between other contacts are also computed, for a good overview of the relation network which is linked to the user. The contacts are extracted from the event's receiver/emitter address.

4 Prototype

The prototype (AMICIS – A Multisource social Interaction and Communication analysIStool) was implemented as a local browser-based application. The system imports the data from the provided user accounts (i.e., email and BSCW) and stores all occurred events after a specific date (which was chosen by the user as a starting point) for the data retrieval. The contacts of the user and the egocentric social network associated are computed during this process. When the data retrieval is completed, the tie-strength between each contact and the user of the application is computed. The initial import is only required at the first start of the application. The system updates after a certain time interval: New events are gathered, and the tie-strength values are updated. If no new events appear, the tie-strength values are also automatically updated, in order to reflect the time decay feature of the events.

The data can be visualized in several ways. A table containing the contacts ordered decreasingly according to the tie-strength is displayed on the top of the main page. The user is able to analyze the topmost contacts in terms of tie-strength. The tie-strength is also presented in a unified view for all the contacts from the different sources, separate views for each data source being also available. The social network can be visualized in a concentric circle view, the user being represented in the center and its contacts being spread around the user as bubbles in circles with the distance to the center and distinct diameter proportional to the tie-strength value (cf. [18]).

The tie-strength can be also presented without the time decay influence, such that the users can better differentiate between the contacts presented in the two settings and thus obtain more accurate results when evaluating the system.

The tie-strength between the user and its individual contacts is represented using a timeline feature, in which the events with the respective contact are represented with their value on a time axis. The value which is displayed is the current value at the time of the visualization, therefore both the event importance and the time decay of

the events can be observed. Also, more than one contact can be added to the timeline at the same time, making it easy to compare the tie-strength between contacts as it can be seen in Figure 1.

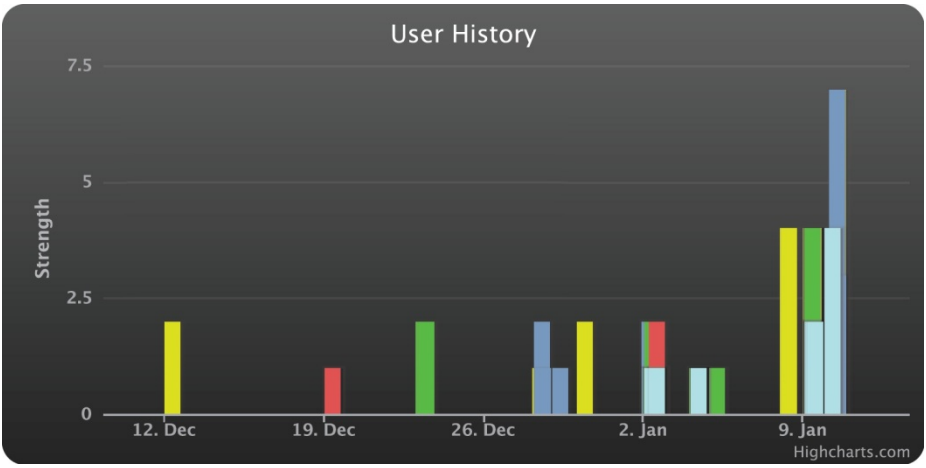

Fig. 1. Screenshot of the tie-strength of certain contacts (one color stands for one person; five people belonging to a group)

5 Evaluation

The user evaluation has been conducted during several days, with different users, the majority being students and researchers. A total number of eleven users have participated in the final evaluation, half of them having completed also the user studies performed in earlier stages of the project.

The objectives of the user evaluation were to test the usability of the AMICIS system, to find out how accurate the discovered relations are. Further we wanted to find out the benefits of such a tool, gather missing features and test the percentage of noise in the discovered relations. It was important to establish whether this tool could help, respectively assist the collaboration and communication.

In order to cover all the above mentioned aspects, the user evaluation was performed using a semi-structured interview performed in two stages. The first part was conducted before the actual usage of the tool. The second part was performed after the tool has been used. A questionnaire with open as well as closed questions regarding different objectives of the project was used. In the first part of the interview, we requested information about the general usage of the source channels in the social context, i.e. number of messages received per day, estimation of the number of contacts in the network, etc. After this step, the users were given a set of instructions to use the system. After the users accomplished the given tasks, they were free to use the tool without any specific tasks.

The relations discovered were considered accurate by 60% and very accurate by 40% of the users. The computation of the results with the time decay feature considered, were as expected for 50% of the users. This corresponds to the requirement study that indicated difficulties in finding a good decay function. The other 50% of the users, found the presented relations relevant, but they expected a bigger amount of relations to be present with contacts to whom they felt very close, but with whom they did not talk for a longer period of time. In general, the closest relations were accurate, but the older relations were harder to classify by the users. The users for which an exact match was not found used some other communication channels in their daily interactions (instant messaging, communication features from social network applications). 90% of the users considered that the decay of the relations with the time passing feature is relevant. 90% of the users appreciated the recommendation which helps them to see which relations need reinforcement. The recommended contacts initially had a relatively strong tie, but a lower value recently.

80% of the users appreciated the automatic discovery of the relation strength with their contacts as well as the unified view of all their relations in one place. 60% of the users would use the AMICIS system for a longer period of time and 50% of the users would use the AMICIS system in order to help them organize their contacts.

Some of the users considered that the relations presented initially had a high level of noise. Most of them considered that the blacklist feature is very good and appropriate in removing the noise, but they also wanted a feature for the automatic removal of only incoming edges from the relation graph. Therefore a modification was made to the AMICIS system and a button has been added for the automatic removal of just incoming links. This feature was appreciated by the users from the next evaluation sessions; however it also blacklisted relevant contacts, for which the events initiated by the user were not present in the history, because they were outside the considered time interval for the initial import.

All in all, we consider that based on the model presented, such a tool can successfully identify and compute the tie strength with a high degree of accuracy and with real benefits to the users. Some possible improvements, the conclusions of this work and the research issues which we consider remaining open, are presented in the section below.

6 Conclusion and Outlook

The present paper addresses questions regarding the methods which can be used in order to determine the tie-strength for individual users across different media. The unification of different media channels in influencing the social context in which the users interact has been studied, as well as the indicators and the practices which can influence the relation strength. Based on those attributes or indicators, the tie-strength has been finally computed.

As the evaluation results have shown, more sources should be analyzed for obtaining better approximations of the tie-strength for individual users. Because the users have very different behaviors when coming to the use of online applications for interacting with other people, resulting from differences in taste, habit, environment, popularity, needs, etc., the more data sources are integrated into the system, the more interactions may be analyzed. Therefore, by applying this, a more accurate computation of the relation strength and a better coverage of the relations may be achieved.

Another important point which should be further developed is that even though the system proposes to investigate the relation strength as its results from the online interactions of an individual user, the emotional intensity should be also considered in measuring the tie-strength. The time variable and the relation frequency may determine an important part of the relation strength, but the affective component has also a big influence in the user perception. A semantic analysis of the contents of the events exchanged, through tag extraction and language pattern analysis would be an important step in the direction of measuring the affective importance of events.

The users have hinted in the evaluation process that detection of groups and contextual computations of relation intensity (e.g., the relations and their strength in university groups, between the participants at a workshop or conference, between friends, etc.) would also improve the user experience and the quality of recommendations the system can provide.

This paper contributes a new model which incorporates two heterogeneous media channels. One communication system (email) and one cooperation system (BSCW) are harmonized in terms of tie-strength. Surprisingly for the users the two systems lead to two different egocentric networks which intersect in only single contacts. From this fact we can deduce that it is necessary to consider all used media to cover the whole network of one person. Therefore we are planning to incorporate two more classes of applications such as social networking sites and instant messaging systems. From our user study we learned that these two classes cover mainly private and leisure communication. But nevertheless they are important for the user.

We have tried to bridge the existing research gaps conducted for studying interpersonal relations in daily life (offline) and the studies on the relations which are established and/or maintained using online systems. After studying the attributes considered in measuring the relation strength in the literature, which proved to be quite different across different studies, the paper presents a series of attributes which proved to be most relevant in on-line tie-strength analysis. Here, the approach was to consider individual indicators suitable for each media channel and to provide the general model capable of unifying the values coming from the individual systems, in order to obtain one value for the relation strength.

Acknowledgments. This work is partly supported by the European Commission through the Seventh Framework Programme for Research and Technological Development (GRANATUM project).

References

1. Gilbert, E., Karahalios, K.: Predicting tie-strength with social media. In: Proceedings of the 27th International Conference on Human Factors in Computing Systems (CHI 2009), pp. 211–220. ACM, New York (2009)
2. Gilbert, E.: Predicting tie-strength in a new medium. In: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW 2012), pp. 1047–1056. ACM, New York (2012)
3. Bird, C., Gourley, A., Devanbu, P.T., Gertz, M., Swaminathan, A.: Mining email social networks. In: Proceedings of the International Workshop on Mining Software Repositories, Shanghai (2006)
4. Culotta, A., Bekkerman, R., McCallum, A.: Extracting Social Networks and Contact Information from Email and the Web. In: Proceedings of CEAS-1 (2004)
5. Resig, J., Dawara, S., Homan, C.M., Teredesai, A.: Extracting social networks from instant messaging populations. In: Proc. of the 7th ACM SIGKDD Workshop on Link KDD (2004)
6. Nasirifard, P., Peristeras, V., Hayes, C., Decker, S.: Extracting and Utilizing Social Networks from Log Files of Shared Workspaces. In: Camarinha-Matos, L.M., Paraskakis, I., Afsarmanesh, H. (eds.) PRO-VE 2009. IFIP AICT, vol. 307, pp. 643–650. Springer, Heidelberg (2009)
7. Granovetter, M.: The Strength of Weak Ties. American Journal of Sociology 78(6), 1360–1380 (1973)
8. Feld, S.: The Focused Organization of Social Ties. American Journal of Sociology 86(5), 1015–1035 (1981)
9. Marsden, P.V., Campbell, K.E.: Measuring Tie-strength. Social Forces 63(2), 482–501 (1984)
10. Cummings, J., Kiesler, S.: Who Works with Whom? Collaborative Tie-strength in Distributed Interdisciplinary Projects. In: Proceedings of the 3rd International e-Social Science Conference, Ann Arbor, Michigan, October 7-9 (2007)
11. Burt, R.: Structural Holes: The Social Structure of Competition. Harvard University Press (1995)
12. Roth, M., Flysher, G., Matias, Y., Leichtberg, A., Merom, R.: Suggesting Friends Using the Implicit Social Graph Categories and Subject Descriptors. Human Factors 371(9620), 233–241 (2010)
13. Ting, I.-H., Wu, H.-J., Chang, P.-S.: Analyzing Multi-source Social Data for Extracting and Mining Social Networks. In: 2009 International Conference on Computational Science and Engineering, vol. 4, pp. 815–820. IEEE (2009)
14. Haythornthwaite, C.: The strength and the impact of new media. In: Proceedings of the 34th Annual Hawaii International Conference on System Sciences, January 3-6, p. 10 (2001)
15. Hanneman, R.A., Riddle, M.: Introduction to Social Network Methods. Network, University of California, vol. 46, pp. 2890–2896 (2005)
16. Granovetter, M.: The Strength of Weak Ties: A Network Theory Revisited. In: Marsden, P.V., Lin, N. (eds.) Sociological Theory, vol. 1, pp. 201–233. John Wiley & Sons (1983)
17. Appelt, W.: WWW Based Collaboration with the BSCW System. In: Bartosek, M., Tel, G., Pavelka, J. (eds.) SOFSEM 1999. LNCS, vol. 1725, p. 66. Springer, Heidelberg (1999)
18. Erickson, T., Smith, D.N., Kellogg, W.A., Laff, M.R., Richards, J.T., Bradner, E.: Socially translucent systems: Social proxies, persistent conversation and the design of 'Babble.' In: Human Factors in Computing Systems: The Proceedings of CHI 1999. ACM Press (1999)