"How well do we know each other?": detecting tie strength in multidimensional social networks

Luca Pappalardo, Giulio Rossetti, Dino Pedreschi
KDD Laboratory
ISTI-CNR and University of Pisa
Pisa, Italy
Email: luca.pappalardo@isti.cnr.it, giulio.rossetti@isti.cnr.it, pedre@di.unipi.it

Abstract—The advent of social media have allowed us to build massive networks of *weak* ties: acquaintances and nonintimate ties we use all the time to spread information and thoughts. Conversely, strong ties are the people we really trust, people whose social circles tightly overlap with our own and, often, they are also the people most like us. Unfortunately, the majority of social media do not incorporate explicitly tie strength information in the creation and management of relationships, and treat all users the same: friend or stranger, with little or nothing in between. In the current work, we address the challenging issue of detecting on online social networks the strong and intimate ties from the huge mass of such mere social contacts. In order to do so, we propose a novel multidimensional definition of tie strength which exploits the existence of multiple online social links between two individuals. We test our definition on a multidimensional network constructed over users in Foursquare, Twitter and Facebook, analyzing the structural role of strong and weak links, and the correlations with the most common similarity measures.

Index Terms—Multidimensional Social Networks; Link Mining; Tie Strength

I. INTRODUCTION

In the last few years, the advent of social networking sites has completely redefined the way we conceive our social relationships, creating the sensation of having broken the constraints of time and geography that limited people's social world. In these virtual environments establishing new friendships is immediate and effortless, so it is reasonable to think that the number of our social bonds could approach to infinite, removing the social boundaries of our modern, technological era. However, what social networks have allowed us to do is to build massive networks of *weak* ties: acquaintances and nonintimate ties we use all the time to reach out to persons, business requests, speaking engagements, or ideas and advice. Despite such enormous quantity of acquaintances, recent works have revealed two major aspects of both online and real social networks:

  i) people still have the same circle of intimacy as ever [1], [2], [3],
  ii) the formation of friendships is strongly influenced by the geographic distance, breaking down the illusion of living in a "global village" [8], [9].

People tend to interact intensely with a small subset of individuals, carrying out a social grooming in order to maintain and nurture strong, intense ties. Strong ties are the people we really trust, people whose social circles tightly overlap with our own and, often, they are also the people most like us. Although such trusted friendships are not so important in the spreading of information [4], new ideas [10], or in finding a job [11], they can affect emotional and economic support [12], [13] and often join together to lead organizations through times of crisis [14]. Unfortunately, the majority of social media do not incorporate tie strength in the creation and management of relationships, and treat all users the same: friend or stranger, with little or nothing in between. A first attempt to take into consideration the social role of a friendship was done by Facebook and Google+ by the introduction of the "circles". Users can use circles as a way to organize their contacts in a sort of address book, creating different groups for relatives, work colleagues, close friends and so on. However, such conceptual organization of contacts does not provide quantitive information about the real strength of the ties, but only the nature of relationships between users and the context in which they take place. For example, the presence of a user in the circle of work colleagues does not necessarily imply the existence of a strong tie, and does not provide any explicit quantification of the importance of the relationship.

In the current work, we address the following issue: how to define a tie strength measure that is capable to discriminate between intimate ties and mere online social contacts?

Actually, it does not exist a formal, unique and shared definition of tie strength, and literature has often provided very personal interpretations of Granovetter's intuition: "the strength of a tie is a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding) and the reciprocal service which characterize the tie" [5]. The most frequently used measurements of tie strength in social networks are based on the number of conversations between users [1], or, in the mobile phone context, on the duration of calls [4]. However, in our opinion these common approaches suffer two major shortcomings. Firstly, the number and intensity of conversations strongly depends from user to user, making it difficult to understand which of these conversations are dedicated to intimate relationships. Secondly, they do not take into account that strong ties must be powered by a form of social grooming, that is mainly based on geographical proximity and face-to-face contacts.
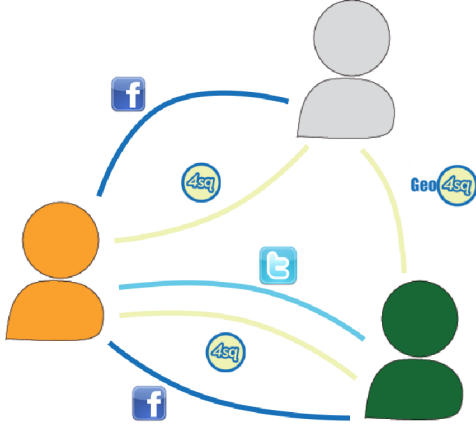
Fig. 1. A schematization of our 4-dimensional social network

Network	Nodes	Edges	Weighted
Foursquare	5783	42691	No
GeoFoursquare	4901	17987	Yes
Facebook	2081	5618	No
Twitter	3745	31638	No
Complete Network	7500	97934	-

TABLE I
BASIC STATISTICS OF THE FOUR DIMENSIONS AND THE WHOLE
MULTIDIMENSIONAL NETWORK.

In order to overcome such shortcomings and extend current techniques, we propose a new definition for the strength of a tie, which exploits the existence of multiple online social links between two individuals. Indeed, while weak ties often rely on a few commonly available media [15], strong ties tend to diversify communicating through many different channels [16]. Moreover, the patterns of homophily tend to get stronger as more types of relationships exist between two people, indicating that homophily on each type of relation cumulates to generate greater homophily for multidimensional than monodimensional ties [7]. To model this behavior, we introduce a strength function and test its meaningfulness on a 4-dimensional social network.

II. RELATED WORK

The concept of tie strength was introduced by Mark Granovetter in his seminal paper "The Strength of Weak Ties" [5]. He proposed four main factors shaping the strength of a tie: amount of time, intimacy, intensity and reciprocal services. Subsequent research expanded the list adding demographic and socio-economic status [17], emotional support [18] and network topology [19]. In [20], authors used survey data from three metropolitan areas to discover the predictors of tie strength. Onnela et al. [4] utilized the duration of calls as a measure for tie strength, and observed that social networks are robust to the removal of the strong ties but fall apart after a phase transition if the weak ties are removed. Gilbert and Karahlios [21] presented a predictive model that maps social media data to tie strength, reaching the 85% accuracy

in distinguishing between strong and weak ties.

Multidimensional network analysis is a relatively recent field. The authors in [22] analyzed the degree distributions of the various dimensions, highlighting the need for analytical tools for the multidimensional study of hubs. A framework for the analysis of multidimensional networks is introduced in [6], defining a large set of measures capturing the interplay of the dimensions both at the global and at the local level. Rossetti et al [23] addressed the link prediction problem in the context of multidimensional networks.

III. MULTIDIMENSIONAL TIE STRENGTH

On the vast online world, two individuals can interact and share interests through several social networking platforms. They can be coworkers on LinkedIn, friends on Facebook or Google+, followers/following on Twitter, they can frequent the same venues on Foursquare, or all of these things together. To express this kind of information, as done by the authors of [6], we choose as model the one offered by multidimensional networks.

Definition 1 (Multidimensional Network). A multidimensional network is a network in which two nodes can be connected, at the same time, by multiple edges that belong to different dimensions.
We model such structure with an edge-labeled undirected multigraph denoted by a tuple $G = (V, E, L)$ where: V is a set of nodes; L is a set of labels; E is a set of labeled edges, i.e. a set of triples (u, v, d) where $u, v \in V$ are nodes and $d \in L$ is a label. Henceforth, we use the term *dimension* to indicate *label*.

Since strong ties tend to diversify communicating through many different channels [16], it makes sense to define a tie strength measure that exploits the multidimensional nature of online interactions. In order to do this, we extend traditional approaches adding three other features.

The first one takes into account the intensity of interaction and the similarity of the nodes in a single dimension:

Definition 2 (Node interaction and similarity).

$$h_d(u,v) = w_d(u,v)\frac{|\Gamma_d(u) \cap \Gamma_d(v)|}{min(|\Gamma_d(u)|, |\Gamma_d(v)|)} \qquad (1)$$

where $w_d \colon V \times V \times L \to \mathbb{N}^+$ is a weight function representing the intensity of the interaction between the nodes in the dimension d, and Γ_d is the set of neighbors of a node. In order to capture whether they belong to the same circle of friendships, and whether such circle is prominent for one of them, the intensity of interactions is influenced by the percentage of common neighbors with respect to the more selective node (the one with less friends). The idea is that, on each dimension, the strength of a tie is defined by the frequency of the interaction among the users, weighted by the value of "friendship selectiveness" of the more selective user (i.e. the probability that the community shared by the two users is the main one for the user with the smaller set of friends).
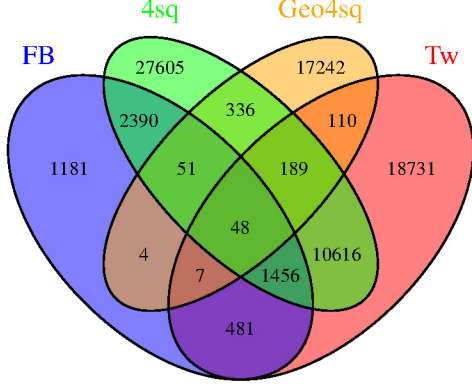
Fig. 2. A Venn diagram showing how the edges of the four dimensions overlap in the multidimensional network (FB = Facebook, 4sq = Foursquare, Tw = Twitter, and Geo4sq = inferred co-occurrence network).
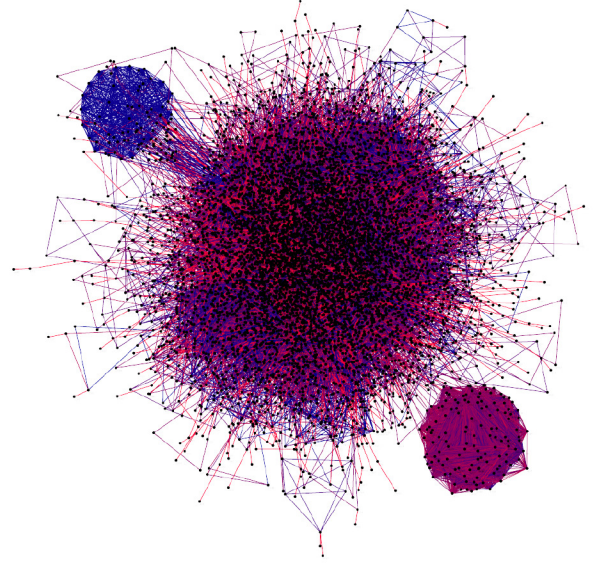

Fig. 3. A global visualization of the network N. The colors of the edges in a color gradient from blue to red indicate the strength of ties, from strong to weak respectively.

The second feature regards the relevance of a dimension for the connectivity of a user: the removal of the links belonging to a dimension should not affect significantly the capacity to reach his real strong connections.

Definition 3 (Connection Redundancy).

$$\varphi_d(u,v) = (1 - DR(u,d))(1 - DR(v,d)) \qquad (2)$$

The dimension relevance DR [6] is the fraction of neighbors that become directly unreachable from a node if all the edges in a specified dimension were removed. We give an higher score to the edges that appear in several dimensions, so we are interested in the complement of those values. If the two nodes are linked in more than one dimension, the score is raised until a maximum of 1.

We merge these aspects taking into account the multidimensionality of a tie: a greater number of connections on different dimensions is reflected in a greater chance of having a strong tie [7]:

Definition 4 (Multidimensional Tie Strength). *Let $u,v \in V$ be two nodes and L the set of dimension of a multidimensional network $G = (V, E, L)$. The strength function $\mathbf{str} : V \times V \to \mathbb{R}$ between two users u, v is defined as:*

$$\mathbf{str}(u,v) = \sum_{d \in D} h_d(u,v)(1 + \varphi_d(u,v)) \qquad (3)$$

The measure proposed, given its formulation, could be used to estimate the strength of ties even in monodimensional networks: in that scenario the φ_d function assume a value equal to zero and the overall sum became the value of h_d. This scoring function is our final measure of tie strength.

IV. EXPERIMENTS

We constructed a multidimensional network $G = (V, E, L)$ by collecting friendships existing between the same 7500
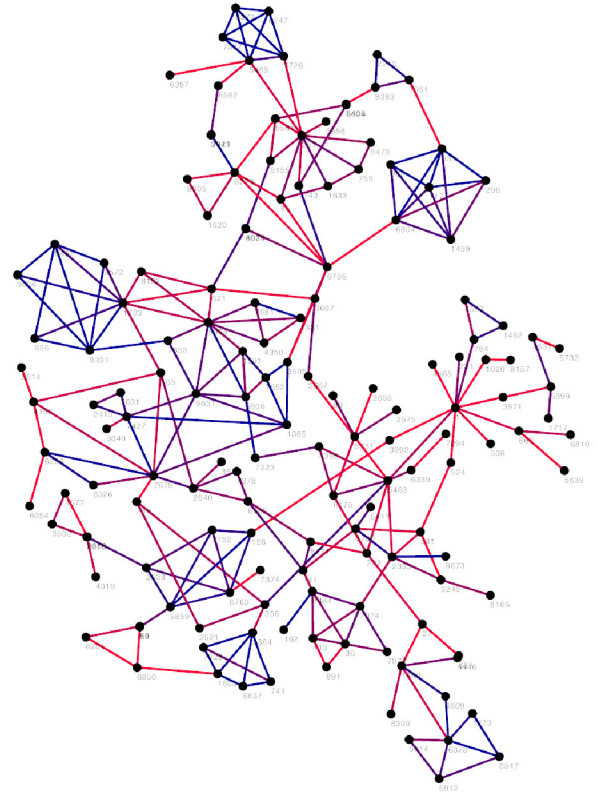

Fig. 4. A portion of the network N. The colors of the edges in a color gradient from blue to red indicate the strength of ties, from strong to weak respectively.

individuals[1] in three online social networks (Foursquare, Twitter and Facebook). Moreover, we inferred a co-occurrence

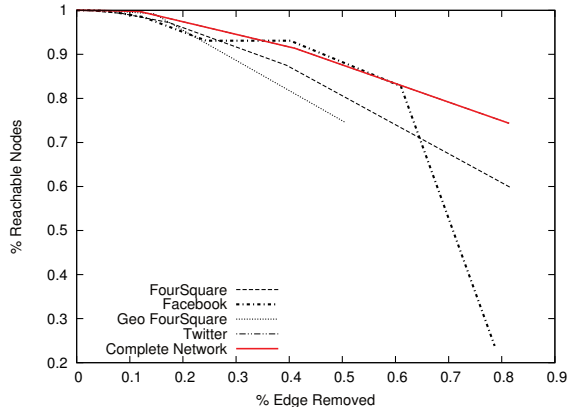[1] All considered users are geographically located in the city of Osaka (Japan).

Fig. 5. The stability of the networks to strong link removal. The curves correspond to removing first the high-strength links, moving toward the weaker ones. (It is interesting to notice that the Facebook network, being smaller and more sparse, tends to collapse quickly).
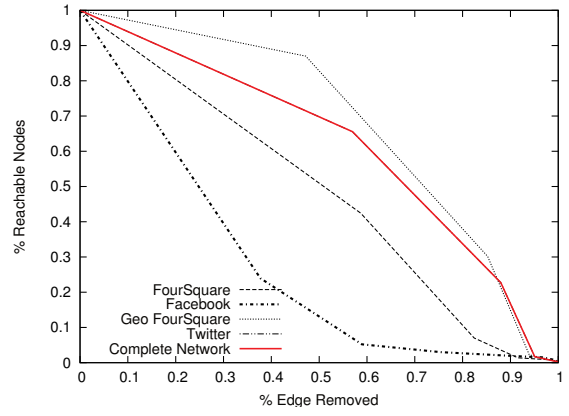

Fig. 6. The stability of the networks to weak link removal. The curves correspond to removing first the low-strength links, moving toward the stronger ones.

network linking two users if they made a Foursquare checkin in the same venue within a time interval of 15 minutes, during a time span of one month. The number of co-occurrences between two individuals was taken as the weight for the corresponding edge. Figure 1 presents a schematic example of our 4-dimensional network, whereas Table I summarizes some characteristics of the multidimensional network and of its dimensions.

In order to test the meaningfulness of our definition and analyze the structural role of strong and weak links, we calculated the strength measure on G and, using the scores obtained, inferred a weighted network $N = (V, E_N)$, collapsing all the edge between two nodes into one. Figure 3 shows a global visualization of N, from which three main clusters clearly emerge, with the one on the left representing people communicating in many different social networking platforms. Furthermore, our measure seems to be consistent with the *"strength of weak ties"* hypothesis [5], with strong ties connecting local communities, and weak ones acting as bridge between them (Figure 4). To test more rigorously this aspect, we studied the resilience of N and the individual networks to the removal of either strong and weak links. Since weak ties act as bridges between different communities, we expect that their removal made the network structure fall apart quickly [4]. Indeed, the deletion of strong ties do not infect considerably the connectivity of the networks, with the 70% of the nodes still reachable in N removing almost all the strong arcs (Figure 5). Conversely, the removal of weak ties rapidly *"destroys"* the networks, splitting them into several small connected components (Figure 6). Our definition is therefore capable to discriminate between intimate circles and the edges acting as bridges between them.

Figure 2 shows a Venn diagram representing the number of ties belonging to each possible intersection of the dimensions. It clearly shows that there are only 48 bonds appertaining to all the 4 dimensions. Such links represent a sort of *"super strong"* ties, i.e. those having a high probability of being real

and intimate friendships.

With the purpose of investigate if the proposed measure assigns a strength value correctly, we analyzed how its score correlates with three well-known network measure: Jaccard, Adamic-Adar and Edge Betweenness.

A. Strength vs. Jaccard

Comparing the values assigned by our measure with the corresponding Jaccard coefficient, we want to verify the existence of a correlation between the strength of a tie and the similarity of the individuals involved. We plot the tie strength against the Jaccard coefficient, both for the network N and the single dimensions. As shown in Figure 7, weak ties tend to have a small Jaccard coefficient, whereas those with higher strength seem more similar. However, there are cases in which an high similarity does not reflect in higher strength. This is because the Jaccard coefficient is defined as the ratio between the common neighbors and all the friends of the two users, whereas we choose to give more credit to the most selective user moving from a symmetric to a non-symmetric measure (equation 1). Our choice tends to penalize similar nodes (with respect to the Jaccard's measure) if they show a significant variance in the cardinality of the sets of non shared neighbors.

B. Strength vs. Adamic Adar

As done with the Jaccard coefficient, we compare our measure with Adamic-Adar. This measure considers how the mutual neighbors of two nodes are selective in establishing connections: the more selective the friendships are, the more likely the two individuals belong to the same friendship community. As we can see in Figure 8, it seems that the strength increases together with the Adamic score in Facebook, Twitter and the network N. It does not happen with Foursquare, presumably because of the peculiar typology of the service that it offers. Anyway, the trend shown by the figure suggests the following conclusion: two nodes belonging to selective circles of friendships have a greater chance to establish a strong bond.
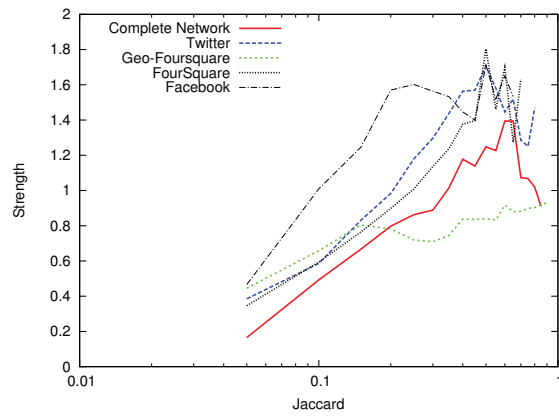
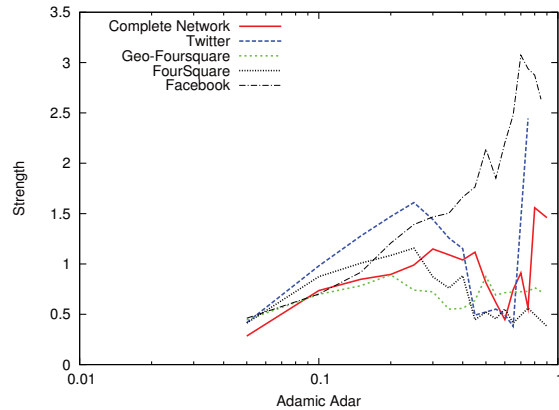Fig. 7.   Relation between Jaccord coefficient and strength values.

Fig. 9.   Relation between Edge Betweenness coefficient and strength values.

Fig. 8.   Relation between Adamic-Adar coefficient and strength values.

C. Strength vs. Edge Betweenness

The edge betweenness is a measure of edge's centrality, equal to the number of shortest paths that pass through that edge. An edge with an high betweenness is likely a bridge between two different communities and, by definition, a weak link. We compare our strength function against this score computed over the single dimensions only. The computation of this measure on the network N is meaningless because, in such network, an edge could establish paths that are not real. As expected, Figure 9 shows that when the edge betweenness increases, the value of strength seems to decrease.

V. CONCLUSION

In this work we have introduced a measure of tie strength for multidimensional networks. Supported by a validation on a 4-dimensional social network, we found that the strength of a tie is strictly related to the number of interactions among the individuals involved. Moreover, it is also related to the number of different contexts in which those connections take place. In the future, we plan to investigate how the information provided by the tie strength can be exploited to tackle well-known problems such as link prediction and community discovery.
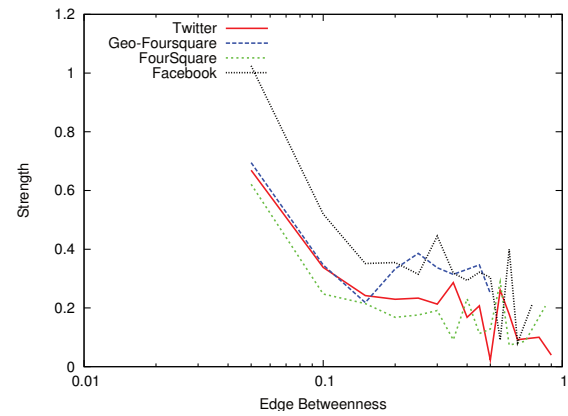
REFERENCES

[1] B. Goncalves, N. Perra, A. Vespignani, *Modeling Users' Activity on Twitter Networks: Validation of Dunbar's Number*, PLoS One, Vol. 6 (28 May 2011).
[2] Cameron Marlow, *Maintained Relationships on Facebook*, Weblog of Cameron Marlow, http://overstated.net/2009/03/09/maintained-relationships-on-facebook.
[3] R. I. M. Dunbar, *The social brain hypothesis*, Evol. Anthropol., Vol. 6, No. 5. (1998), pp. 178-190.
[4] J. P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, A. L. Barabási, *Structure and tie strengths in mobile communication networks*, Proceedings of the National Academy of Sciences, Vol. 104, No. 18. (1 May 2007), pp. 7332-7336.
[5] M. S. Granovetter, *The Strength of Weak Ties*, American Journal of Sociology, Vol. 78, No. 6. (1973), pp. 1360-1380.
[6] M. Berlingerio, M. Coscia, F. Giannotti, A. Monreale, D. Pedreschi, *Foundations of Multidimensional Network Analysis*, In Proceedings of the International Conference on Advances in Social Networks Analysis and Mining - 2011.
[7] M. McPherson, L. S. Lovin, and J. M. Cook, *Birds of a feather: Homophily in social networks*, Annual Review of Sociology, vol. 27, no. 1, pp. 415-444, 2001.
[8] J. Goldenberg, M. Levy, *Distance Is Not Dead: Social Interaction and Geographical Distance in the Internet Era*, CoRR abs/0906.3202: (2009).
[9] D. Mok, B. Wellman, and J. A. Carrasco, *Does distance still matter in the age of the internet?*, Urban Studies, vol. 46, 2009.
[10] R. S. Burt, *Structural Holes and Good Ideas*, American Journal of Sociology, 110(2), 349399, 2004.
[11] M. Granovetter, *Getting a Job: A Study of Contacts and Careers*, University Of Chicago Press, 1974.
[12] C. Schaefer, J. C. Coyne, et al., *The Health-related Functions of Social Support*, Journal of Behavioral Medicine, 4(4), 381406. 1990.
[13] J. H. Fowler and N. A. Christakis, *Dynamic spread of happiness in a large social network: longitudinal analysis over , 20 years in the framingham heart study*, BMJ, vol. 337, no. dec04_2, pp. a2338+, Jan. 2008.
[14] D. Krackhardt, R. N. Stern, *Informal Networks and Organizational Crises: An Experimental Simulation*, Social Psychology Quarterly, 51(2), 123-140, 1988.
[15] C. Haythornthwaite, B. Wellman, *Work, Friendship, and Media Use for Information Exchange in a Networked Organization*, J. Am. Soc. Inf. Sci., 49(12), 1101-1114, 1998.
[16] C. Haythornthwaite, *Strong, Weak, and Latent Ties and the Impact of New Media*, Information Society, 18(5), 385401, 2002.

[17] N. Lin, W. M. Ensel, et al., *Social Resources and Strength of Ties: Structural Factors in Occupational Status Attainment*, American Sociological Review, 46(4), 393405, 1981.

[18] B. Wellman, S. Wortley, *Different Strokes from Different Folks: Community Ties and Social Support*, The American Journal of Sociology, 96(3), 558588, 1990.

[19] R. Burt, *Structural Holes: The Social Structure of Competition*, Harvard University Press, 1995.

[20] P. V. Marsden, K. E. Campbell, *Measuring Tie Strength*, Social Forces, 63(2), 482501, 1990.

[21] E. Gilbert, K. Karahalios, *Predicting tie strength with social media*, in Proceedings of the 27th international conference on Human factors in computing systems, ACM, 2009, pp. 211-220.

[22] M. Szell, R. Lambiotte, and S. Thurner, *Trade, conflict and sentiments: Multi-relational organization of large-scale social networks*, arXiv.org, 1003.5137, 2010.

[23] Giulio Rossetti, Michele Berlingerio, Fosca Giannotti, *Scalable Link Prediction on Multidimensional Networks*, ICDM Workshops 2011: 979-98.