

Identifying Unreliable Sources of Skill and Competency Information

Maryam Fazel-Zarandi

Department of Computer Science
University of Toronto
Toronto, Canada
mfazel@cs.toronto.edu

Mark S. Fox

Department of Mechanical and Industrial Engineering
University of Toronto
Toronto, Canada
msf@eil.utoronto.ca

Abstract—Organizations need to accurately understand the skills and competencies of their human resources in order to effectively respond to internal and external demands for expertise and make informed hiring decisions. In recent years, however, human resources have become highly mobile, making it more difficult for organizations to accurately learn their competencies. In such environment, organizations need to rely significantly on third parties to provide them with useful information about individuals. These sources and the information they provide, however, vary in degrees of trust and validity. In a previous paper, we developed an ontology for skills and competencies and modeled and analyzed the various sources of information used to derive the belief in an individual's level of competency. In this paper, we present an approach based on social network analysis for identifying unreliable sources of competency information. We explore the conditions under which evaluations given by an individual or a group about another can be trusted. We evaluate this approach using recommendation data gathered by crawling user profiles in LinkedIn.

Keywords—expert profiling; collusion detection; skill and competency management

I. INTRODUCTION

Organizations need to accurately understand the skills and competencies of their human resources in order to effectively respond to internal and external demands for expertise and make informed hiring decisions. Wrong decisions in this respect may result in significant loss of value and high turnover of poorly matched human resources [1]. This is particularly important for most services organizations, especially those with a medium or large number of employees that provide a variety of products and services to multiple and changing clients. The replacement of the traditional system of life-long employment with more flexible work roles [2], the necessary multi-skilling of workers [3], and the increased competition due to globalization, however, have made the management of human assets quite a challenge in recent years. Human resources have become highly mobile, making it more difficult for organizations to accurately learn their skills and competencies.

In such environment, organizations need to rely significantly on third parties to provide them with accurate skill and competency information about individuals. In particular, references, letters of recommendation, and evaluations in the form of 360 reviews or performance appraisals are important sources used to assess individuals [4], both in hiring new employees and in consideration of promotions to higher ranks.

These sources and the information they provide, however, vary in degrees of trust and validity. In relying on these sources we must recognize that people have personal goals, change over time, and the nature of their relationships may affect their judgments. In other words, not all sources are honest in exchanging information about others even if they have been trustworthy in previous interactions. In particular, *collusion* occurs when several individuals co-operate in dishonest behavior by lying to or withholding information from a third party [6, 7]. In the case of hiring, for example, individuals can collude with those who recommend them in order to get hired. Additionally, in the case of multi-rater reviews, individuals can collude to distort their ratings positively across all group members [8]. In fact, studies have shown that collusion and selection of raters who influence results positively, is an important weakness which affects the usefulness of such reviews [9]. The incentive for such collusive behavior is that it delays the arrival of information about individuals' skills and competencies that could be potentially damaging to their career and future wages [10]. It may also be the case that some individuals act together to give negative ratings to an individual outside their group (e.g., the case of retaliation). As such, to construct accurate expertise profiles, we need to identify when the information provided by an individual's network and peers can be considered as credible.

The focus in this paper is on identifying unreliable sources of skill and competency information. We are interested in understanding the conditions under which evaluations given by an individual or a group about another can be trusted. In this direction, we present an approach based on social network analysis. We evaluate this approach using recommendation data gathered by crawling user profiles in LinkedIn.

II. APPROACH

Individuals and organizations are "embedded within networks of interconnected relationships" [11]. The social network approach provides a set of methods, models, and theories for studying these relationships. Wasserman and Faust [12] state the following as some of the principles underlying this approach that make it different from other approaches:

- Actors and their actions are viewed as interdependent rather than independent units.
- Relational ties between actors are channels for flow of resources.

- Network structure provides opportunities for or constraints on individual action.

In other words, the most distinguishing feature of the network perspective is its emphasis on relationships and their implications [12, 13]. These relationships can be characterized as strong or weak, multiplex, symmetric or asymmetric, etc.

Based on these principles, social network analysis is an appropriate technique and theory for identifying unreliable sources of skill and competency information. More specifically, Brass et al. [6] state that the ongoing social relationships between actors provide the constraints and opportunities that may help explain and detect unethical behavior in general. In particular, they argue that

“as the strength, multiplexity, symmetry, and status equality of a relationship increase, frequency of interaction and trust provide increased opportunities and payoffs for unethical behavior, whereas empathy, psychological proximity, and the cost of losing a strong, multiplex relationship constrain unethical behavior” (pg. 19).

In other words, in relationships characterized by strong multiplex ties, decisions to engage in an unethical activity are made in light of the history and future of the relationship. There may also be indirect influence that others may have on these decisions. In particular, with additional individuals in the community we need to consider the effects of *surveillance* and *reputation* [6]. An advantage of the network perspective is that it allows the study of these effects through the consideration of the structure of relationships in the entire network. Note that since it is impossible to determine individual's intent, we can only focus on observable behaviors that produce results similar to those produced by a colluding group [7].

It is perhaps worth emphasizing that the environment we are considering is one in which individuals are expected to invest in acquiring reputation under normal circumstances. The bases for this assumption are that organizations have memories, individuals within a particular field typically communicate with one another, and evaluative belief about the actions of other individuals spreads within communities. If this is not the case, then individuals will always choose to collude since there is no further scope for potential gains or losses.

III. IDENTIFYING UNRELIABLE REFERENCES

References and letters of recommendation are important sources used to assess individuals, both in hiring new employees and consideration of promotions to higher ranks. In order to identify unreliable references, we introduce a network of social relations *RelNet*. This network has two distinct kinds of edges: 1) person-person edges connect two individuals and indicate ties in a social network; and 2) person-organization edges connect an individual to an organization and indicate membership or previous interactions of the person with the organization. Note that *RelNet* is a weighted network, where weights on links indicate the strength of the relationship.

Consider an organization *Org* seeking recommendations (or evaluations) for individual *C* from recommenders R_1, \dots, R_n , and a network of relations *RelNet* with the following weights:

- $0 < w_i \leq 1$ indicate the strength of the link between recommender *i* and the candidate;

- $0 \leq v_i \leq 1$ indicate the strength of the link between recommender *i* and the organization ($v_i = 0$ indicates no previous interactions);

- $0 \leq u_{ij} \leq 1$ indicate the strength of the link between recommenders *i* and *j*.

For each recommender *i*, the relative value r_i of its ties to *C* compared to *Org* is equal to the ratio of the weights of those links ($r_i = w_i/v_i$, $v_i \neq 0$). Intuitively, r_i is the penalty of breaking the *R_i-C* link compared to the *R_i-Org* link from *R_i*'s perspective. Now, assume $0 \leq k \leq n$ recommenders give high recommendations. We want to identify the potential collusive set in favor of the candidate (indicated by *PCS⁺*) and/or against the candidate (indicated by *PCS⁻*). To this end, we need to consider the following three cases.

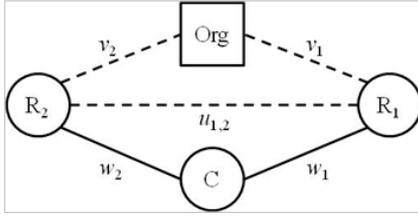
1. $k = n$: Assume $0 \leq m < n$ recommenders have $r_i \leq 1$ and the remaining $(n - m)$ recommenders have $r_i > 1$. In this case, $\{C, R_{m+1}, \dots, R_n\}$ is a potential collusive set (*PCS⁺*). Furthermore, for any of the m remaining recommenders, if the cost of breaking ties with all (in the worst case) of the members of the potential collusive group by giving a low recommendation is higher than the cost of breaking ties with the organization, then we can also add this individual to *PCS⁺*. In other words, for any recommender $1 \leq i \leq m$, if $\sum_{j=m+1}^n u_{i,j} + w_i > v_i$, then this recommender is also added to *PCS⁺*.
2. $k = 0$: In this case all recommenders have given low recommendations for *C*. As such, we need to consider the case of collusion between the recommenders against the candidate. For every pair of recommenders *i* and *j*, if $u_{i,j} > w_i$ and $u_{i,j} > w_j$ then these two recommenders are added to the potential collusive set against the candidate (*PCS⁻*).
3. $0 < k < n$: In this case there is a discrepancy between the recommendations given for *C*. For the k recommenders who gave high recommendations, we consider Case 1 above with a network consisting of *Org*, *C*, and only these k recommenders. For the remaining $(n - k)$ recommenders who gave low recommendations, we consider Case 2 above with a network consisting of *Org*, *C*, and only these recommenders.

A. Hiring Example

Assume the organization seeks recommendation for candidate *C* from recommender R_1 . The recommender is connected to the candidate and the organization via links with strength $0 < w_1 \leq 1$ and $0 \leq v_1 \leq 1$, respectively ($v_1 = 0$ indicates no previous interactions). In this example, R_1 and *C* can collude against *Org*: R_1 can provide *Org* unreliable information about *C*'s skills and competencies, thus influencing *Org*'s decision about hiring *C*. However, if R_1 acts dishonestly toward *Org*, it risks potential loss of a relationship. In order to determine whether R_1 's recommendation for *C* is reliable, we consider the cost of breaking the *R₁-C* tie compared to the *R₁-Org* tie from R_1 's perspective, i.e., the relative value $r_1 = w_1/v_1$ of the two ties as perceived by R_1 . We can see that as r_1 increases, the possibility of *C* and R_1 colluding against *Org* also increases.

Now assume another recommender R_2 is added, with $0 \leq u_{1,2} \leq 1$ indicating the weight of the link between the two recommenders. Here, in addition to the cost of breaking a tie,

the organization can also consider the discrepancy between the two recommendations.



1. Both R_1 and R_2 give high recommendations. If both r_1 and r_2 are greater than 1, then there is a possible chance of collusive behavior by C - R_1 - R_2 since both R_1 and R_2 value their ties to C more than their ties to Org . If $r_1 > 1$ and $r_2 < 1$, then there is still a possible chance of collusive behavior by C - R_1 - R_2 if $u_{1,2} + w_2 > v_2$. R_2 considers that in the worst case he may risk breaking ties with both R_1 and C if he chooses to give a low recommendation (minimize maximum regret).
2. Both R_1 and R_2 give low recommendations. If $u_{1,2} > w_1$ and $u_{1,2} > w_2$, then there is a possible chance of collusive behavior by R_1 - R_2 against the candidate.
3. R_1 gives a high recommendation and R_2 gives a low recommendation. If $r_1 > 1$, then there is a possibility of pair-wise collusion between R_1 and C .

B. Measuring Tie Strength

To identify possible collusions, we need to quantify the strength of ties between individuals and organizations. Since *RelNet* is a social-affiliation network, measures for person-person and person-organization links are required.

Granovetter [15] initially defined social tie strength as “a combination of the amount of time, the emotional intensity, the intimacy, and the reciprocal services which characterize the tie” (pg. 1361). Over the years, this list of indicators of tie strength has been expanded by other scholars (e.g., [16, 17]). In a review of the existing literature, Gilbert and Karahalios [18] found seven dimensions of tie strength proposed in different works, namely, intensity, intimacy, duration, reciprocal services, structural, emotional support, and social distance. In recent years, due to the increasing popularity of online social networking services, predicting relationship strength based on profile similarity and online interaction activity has also gained the attention of researchers (e.g., [18-20]). These models usually make use of 1) the theory of homophily, i.e., individuals form ties with others who are similar [21], 2) the balance theory, i.e., individuals form ties based on the configuration of ties the two persons have with third individuals [22], and/or 3) duration and frequency of interactions and the words exchanged between the two individuals as measures of intensity and intimacy. Xiang et al. [20], for example, model tie strength as a hidden effect of nodal profile similarities by considering attributes such as the schools and companies the users attended and the online groups that they joined.

The aforementioned studies have mainly considered person-person links. In fact, person-organization links in social-affiliation networks are generally considered as dichotomous, i.e., an individual is either a member of an organization or not. This is not sufficient in our case, since

current and previous affiliations with an organization may have different influence on the willingness of individuals to engage in unethical behavior. An aspect of tie strength relevant for measuring these types of links is “relationship commitment.” Relationship commitment is defined as one “partner believing that an ongoing relationship with another is so important as to warrant maximum efforts at maintaining it; that is, the commitment party believes the relationship is worth working on to ensure that it endures indefinitely” (pg. 23) [23]. In this regard, recency and frequency of contact can be used to measure this aspect of tie strength.

For determining candidate-recommender link strength, we consider the following dimensions:

1. how long they have known each other,
2. the nature of the relationship, e.g., colleague, business partner, education, service provider,
3. recency of interactions, e.g., how recently did they work together, and
4. how supportive they are of each other, e.g., to what extent do they recommend each other.

These same dimensions are used to assign other person-person link weights. The weight of the link between a candidate and recommender i is given by:

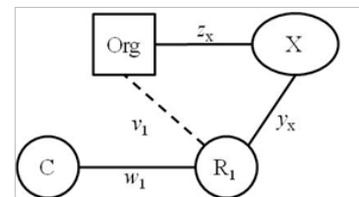
$$w_i = \sum_k \beta_k \cdot a_k(i) \quad (1)$$

where, $a_k(i)$ represent the dimensions stated above and weights β_k indicate the relative importance of one dimension over another. For person-organization links, we assign the maximum possible if the individual is currently a member of the organization. Otherwise, similar to [24] which considers email interactions, previous affiliations are weighted as a function of recency according to an exponential decay function:

$$v_i = \sum_{interactions} dur(j) \frac{(t_{now} - t(j)) / t_{now}}{\lambda} \quad (2)$$

where, $dur(j)$ is the duration and $t(j)$ is the end time of the j -th interaction, and half-life λ is a tunable parameter.

As stated earlier, there may also be indirect influence that others may have on the decision of an individual to engage in a collusive activity because of surveillance and reputation. For instance, consider the example shown in the following figure, in which node X is a person:



Here, in the worst case, R_1 could risk breaking ties with both Org and X if it chooses to be dishonest, since Org may communicate this experience to X . The likelihood of information flowing directly from Org to X is proportional to the strength of their relationship z_x . As such, in the worst case, the possibility of collusion between R_1 and C can be suggested if $w_1 > (v_1 + z_x y_x)$.

When there is more than one additional node present, we need to consider the global structure of the network to determine how fast information can spread. One measure we can use for this is “closeness centrality.” Individuals with high closeness centrality have minimum path distances to other nodes in the network and as such need only a small number of steps to interact and communicate with other individuals in the network [12]. In other words, high closeness centrality suggests high surveillance and extensive loss of reputation by acting unethically [6]. In an unweighted graph, closeness centrality is defined “as the inverse sum of shortest distances to all other nodes from a focal node” [25]. We use an extension of this measure proposed by Newman [26] for weighted graphs which relies on Dijkstra’s shortest path algorithm [27] and in which weights are inverted so they can be considered as costs (since weights in most weighted networks indicate tie strength and not the cost of the link, the tie weights need to be reversed [25]). More specifically, the length of the shortest path between two nodes is $d(i, j) = \min(\frac{1}{w_{i,h}} + \dots + \frac{1}{w_{h,j}})$, where h are intermediary nodes on paths between nodes i and j , and $w_{i,j}$ is the weight of the link between the two nodes. Using this, closeness centrality is defined as:

$$C_c(i) = [\sum_j^N d(i, j)]^{-1} \quad (3)$$

Having this measure, we define the cost incurred by recommender R_i lying to Org in the worst case as:

$$V(i) = v_i + C_c^r(Org) \cdot \sum_{j \in S(i)} w_{i,j} \quad (4)$$

where, v_i is the strength of the direct link between R_i and Org , $C_c^r(Org)$ is the closeness centrality of Org in the largest connected component of $RelNet$ when recommenders R_1, \dots, R_n and candidate C have been removed, $S(i)$ is the set of shared neighbors of R_i and Org in this network, and $w_{i,j}$ is the strength of the link between R_i and node j . $V(i)$ can be substituted for v_i in the analysis of potential collusive behavior.

IV. ESTIMATING LINK STRENGTHS

With the above framework in place for identifying potential collusive behavior, we first have to determine the strength of the links in $RelNet$. The method must be independent of the information provided by the recommenders as we do not know whether to trust them. The question is whether a social network such as LinkedIn contains enough information to estimate the strengths, hence indicate the possibility of collusion. Note that we are not looking for irrefutable proof from these networks, simply enough information to warrant a deeper look at the references. In the following we consider each of the dimensions for determining strengths to see to what extent LinkedIn provides an estimate.

How long have they known each other? LinkedIn does not explicitly provide this information. Never the less, we can estimate this by finding the earliest time that the recommender and candidate worked at the same organization or attended the same educational institution, as determined by their stated employment history and educational background.

What is the nature of their relationship? An analysis of each of their employment histories, in particular what role they each played when at the same organization or educational institution at the same times, can be used to determine whether they were peers or not.

How recently did they work together? Again, an analysis of their employment histories and educational backgrounds will provide us with this information.

How supportive are they of each other? To determine how supportive they are, we can analyze the recommendations provided in LinkedIn. In particular, does their exist a symmetric recommendation relationship between the recommender and candidate? If so, this is an indication of the strength of their links. In the situation where there are two recommenders and a candidate, if we find that the three are Strongly Connected, then it may provide further evidence of the strength of their links

The analysis encountered the following challenges. 1) Crawling the entire recommendation network is not feasible, as such we resort to using samples of the graph. Crawling only a subset of a graph by ending a breadth-first search (BFS) early, called the snowball method [30, 31], is known to produce a biased sample of nodes. In particular, partial BFS crawls are likely to overestimate node degree, and underestimate the level of symmetry and the power-law coefficient [32]. 2) The recommendation graph in LinkedIn can only be crawled by following links in the backward direction (i.e., we cannot easily determine the set of nodes which point out of a given node). Using only backward links does not necessarily crawl an entire connected component; instead, it explores the connected component that reaches the set of seed users. This limitation is typical for studies that crawl directed online networks [32]. 3) LinkedIn allows users to declare their profiles and recommendations as private. We were unable to determine any information about such nodes.

In the remainder of this section we report results for two rounds of data collection, namely, the network of recommendations obtained by starting with:

- 100 mechanical engineers in Toronto as seeds (*ME-Net*).
- 100 software engineers in Toronto as seeds (*SE-Net*).

Of the initial seeds, 45 in the *ME-Net* and 73 in the *SE-Net* had at least one recommendation. We used these seeds to crawl the network up to level 4 (with the seeds as level 0). Table I reports high-level statistics of the two crawls.

TABLE I. HIGH-LEVEL STATISTICS OF THE TWO CRAWLS

	ME-Net	SE-Net
Number of seeds	45	73
Number of nodes	8300	14931
Number of links	9398	18539
Network density	0.000136	0.000083
Fraction of symmetric links	7.82%	11.51%

Having these two networks, we extracted strongly connected components (SCC) of size greater than one. These are subgraphs that indicate recommendation patterns that

produce results similar to those produced by a colluding group. ME-Net has 65 strongly connected components (SCC) consisting of 8.43% of the nodes. SE-Net, on the other hand, has 127 SCCs consisting of 12.73% of the nodes. Table II reports the number of SCC of different size found in each crawl.

From Table II, we can see that about 40% of the components in both crawls consist of only two nodes (45% in SE-Net). The largest SCC in ME-Net is composed of 195 nodes and 506 edges whereas the largest SCC in SE-Net is composed of 1198 nodes and 3193 edges. Maximum in-degree is 73 and 34, and maximum out-degree is 62 and 31 in the ME-Net and SE-Net respectively. Considering components of size greater than two, network density in the ME-Net ranges from 0.013 to 1 with an average of 0.426, and in the SE-Net it ranges from 0.002 to 1, with an average of 0.420.

V. EVALUATION

In the ME-Net, 682 links (7.82%), and in the SE-Net, 1913 links (11.51%) were symmetric. Symmetry can make it harder to identify potential collusive sets just by considering the structure of the network and identifying strongly connected components. Consider, for example, chains and k -stars. With such subgraphs, we may be able to conclude that the nodes are possibly engaging in pair-wise collusive behavior. However, in order to make the conclusion that the entire subgraph is potentially collusive, we must also consider the nature of the relationship between each node. With k -stars in our dataset, we found that the central node was either a project manager or a senior member who had managed the other nodes, or was a contract worker. With chains, it was usually the case that the individual had changed jobs and the neighboring nodes possibly did not even know each other. Another possible cause of a chain could be the hierarchical structure within an organization, although we did not observe this in our dataset. As such, we may want to remove from a SCC the nodes that are connected to only one other node via a symmetric link and label them as candidate colluders.

TABLE II. NUMBER OF STRONGLY CONNECTED COMPONENT OF DIFFERENT SIZE

Component Size	ME-Net	SE-Net
2	26	57
3	11	19
4	5	8
5	4	11
6	3	4
7	3	4
8	4	2
9	1	2
10	2	4
> 10	6	16

The existence of dense strongly connected components in the LinkedIn dataset suggests higher link strengths among the nodes. But when combined with other dimensions (e.g., the

nature of their relationship), we can further analyze the relationship between recommenders and candidates. LinkedIn's API returns the label of a recommendation link as one of *colleague*, *education*, *service-provider*, or *business-partner* (which limits our knowledge of the nature of the relationship). Of the symmetric links in ME-Net, 48.97% were colleagues, 18.77% were business-partners, and 10.41% were service-provider/business-partners. These numbers were 83.17%, 4.70%, and 4.29% in the SE-Net. One might expect that if the recommender is a colleague, there is a higher probability of collusion than in other relationships.

VI. CONCLUSION

In this paper we proposed a model for detecting the possibility of collusion. By representing the strength of the relationship between recommenders and candidates, and between recommenders and organizations, we postulated conditions under which recommenders may possibly collude to provide false information about the candidate. We then proposed dimensions along which the strength of relationships may be measured and investigated the availability of these dimensions within LinkedIn.

Our analysis of LinkedIn shows that it contains significant amounts of information to support the determination of link strengths, but that some of the information is not available via the API and has to be manually extracted from the web pages. In particular, better access to recommender and candidate roles in the organization combined with recommendations would greatly enhance the determination of weights.

Finally, the ultimate evaluation of this approach awaits the analysis of a large number of hires whose performance did not match their references.

REFERENCES

- [1] Y. Naveh, Richter, Y., Altshuler, Y., Gresh, D.L., Connors, D.P., "Workforce Optimization: Identification and Assignment of Professional Workers using Constraint Programming," *IBM Journal of Research and Development - Business optimization*, vol. 51, pp. 263-279, 2007.
- [2] S. G.W., "A Critical Review of the Science and Practice of Competency Modeling," *Human Resource Development Review*, vol. 12, pp. 86-107, 2013.
- [3] S. Liyanage, "Competency Management and Systems," in *Competencies in Organizational E-Learning: Concepts and Tools*, ed: Idea Group Inc, 2007, pp. 10-40.
- [4] M. Fazel-Zarandi, Fox, M.S., "Constructing Expert Profiles over Time for Skills Management and Expert Finding," in *Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies (i-KNOW 2011)*, Graz, Austria, 2011.
- [5] B. Qureshi, Min, G., Kouvatso, D., "Collusion Detection and Prevention with FIRE+ Trust and Reputation Model," in *Proceedings of the 10th International Conference on Computer and Information Technology (CIT'2010)*, Bradford, United Kingdom, 2010, pp. 2548-2555.
- [6] D. J. Brass, Butterfield, K.D., Skaggs, B.C., "Relationships and Unethical Behavior: A Social Network Perspective," *Academy of Management Review*, vol. 23, pp. 14-31, 1998.
- [7] Q. Lian, Zhang, Z., Yang, M., Zhao, B.Y., Dai, Y., Li, X., "An Empirical Study of Collusion Behavior in the Maze P2P File-Sharing

- System," in *Proceedings of the 27th International Conference on Distributed Computing Systems (ICDCS'07)*, Toronto, Canada, 2007, pp. 56-69.
- [8] P. T. Chua, Ramaya, R., Chan, K., Ang, C., Goh, T.K.M., "The Validity of Peer Ratings under Developmental versus Evaluative Contexts," in *Proceedings of the 47th International Military Testing Association Annual Conference*, Singapore, 2005.
- [9] G. Toegel, Conger, J., "360-Degree Assessment: Time for Reinvention," *Academy of Management Learning and Education*, vol. 2, pp. 297-311, 2003.
- [10] J. P. Johnson, "Collaboration, Peer Review and Open Software," *Information Economics and Policy*, pp. 477-497, 2006.
- [11] D. J. Brass, Galaskiewicz, J., Greve, H. R., Tsai, W., "Taking Stock of Networks and Organizations: A Multilevel Perspective," *Academy of Management Journal*, vol. 46, pp. 795-817, 2004.
- [12] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*. New York, NY: Cambridge University Press, 1995.
- [13] B. Wellman, "The Community Question Re-evaluated," *Comparative Urban and Community Research*, vol. 1, pp. 81-107, 1988.
- [14] D. Easley, Kleinberg, J., *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. New York, NY, USA: Cambridge University Press, 2010.
- [15] M. S. Granovetter, "The Strength of Weak Ties," *The American Journal of Sociology*, vol. 78, pp. 1360-1380, 1973.
- [16] B. Wellman, Wortley, S., "Different Strokes from Different Folks: Community Ties and Social Support," *The American Journal of Sociology*, vol. 96, pp. 558-588, 1990.
- [17] N. Lin, Ensel, W. M., Vaughn, J.C., "Social Resources and Strength of Ties: Structural Factors in Occupational Status Attainment," *American Sociological Review*, vol. 46, pp. 393-405, 1981.
- [18] E. Gilbert, Karahalios, K., "Predicting Tie Strength With Social Media," in *Proceedings of the 27th International Conference on Human Factors in Computing Systems (CHI'09)*, New York, NY, USA, 2009, pp. 211-220.
- [19] I. Kahanda, Neville, J., "Using Transactional Information to Predict Link Strength in Online Social Networks," in *Proceedings of the Third International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2009, pp. 74-81.
- [20] R. Xiang, Neville, J., Rogati, M., "Modeling Relationship Strength in Online Social Networks," in *Proceedings of the 19th International Conference on World wide Web*, Raleigh, North Carolina, USA, 2010, pp. 981-990.
- [21] M. McPherson, Smith-Lovin, L., Cook, J., "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology*, vol. 27, pp. 415-444, 2001.
- [22] P. R. Monge and N. S. Contractor, *Theories of communication networks*. Oxford ; New York: Oxford University Press, 2003.
- [23] R. M. Morgan, Hunt, S. D., "The commitment-trust theory of relationship marketing," *Journal of Marketing*, vol. 58, pp. 20-38, 1994.
- [24] M. Roth, Ben-David, A., Deutscher, D., Flysher, G., Horn, I., Leichtberg, A., Leiser, N., Matias, Y., Merom, R., "Suggesting Friends Using the Implicit Social Graph," in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2010, pp. 233-242.
- [25] T. Opsahl, Agneessens, F., Skvoretz, J., "Node Centrality in Weighted Networks: Generalizing Degree and Shortest Paths," *Social Networks*, vol. 32, pp. 245-251, 2010.
- [26] M. E. J. Newman, "Scientific Collaboration Networks: II. Shortest Paths, Weighted Networks, and Centrality," *Physical Review E*, vol. 64, 2001.
- [27] E. W. Dijkstra, "A Note on Two Problems in Connexion with Graphs," *Numerische Mathematik*, vol. 1, pp. 269-271, 1959.
- [28] E. Nuutila, Soisalon-Soinen, E., "On Finding the Strongly Connected Components in a Directed Graph," *Information Processing Letters*, vol. 49, pp. 9-14, 1994.
- [29] R. E. Tarjan, "Depth First Search and Linear Graph Algorithms," *SIAM Journal on Computing*, vol. 1, pp. 146-160, 1972.
- [30] B. H. Erickson, Nosanchuk, T.A., "Applied Network Sampling," *Social Networks*, vol. 5, pp. 367-382, 1983.
- [31] L. A. Goodman, "Snowball Sampling," *Annals of Mathematical Statistics*, vol. 32, pp. 148-170, 1961.
- [32] A. Mislove, Marcon, M., Gummadi, K. P., Druschel, P., Bhattacharjee, B., "Measurement and Analysis of Online Social Networks," in *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, San Diego, California, USA, 2007, pp. 29-42.