

Aggregating Edge Weights in Social Networks on the Web Extracted from Multiple Sources with Different Importance Degrees

Rasim M. Alguliev, Ramiz M. Aliguliyev, Fadai S. Ganjaliyev*

Institute of Information Technology, Azerbaijan National Academy of Sciences, Baku, Azerbaijan.
Email: rasim@science.az, r.alguliyev@gmail.com, *fadaig@yahoo.com

Received November 16th, 2011; revised January 30th, 2012; accepted February 7th, 2012

ABSTRACT

Information on a given set of entities can be derived from multiple sources on the Web. Social networks built from these sources, using these entities as nodes, will have different edge weight values, although the entities will be the same. If these sources are different, one will not normally trust each of them equally. One source will be considered more or less importance than the other. Completely ignoring sources with little importance may yield unexpected results. In this paper, we propose a method for aggregating weight values for social networks built from the Web using different sources. First, multiple social networks are built from different data sources. Then the received edge weights are aggregated, with the importance of a data source taken into account.

Keywords: OWA Operator; Orness; Data Source

1. Introduction

Presently there is a great deal of research being conducted on social network data extraction from the Web. Referral Web is considered the first such data-extraction system [1]. Y. Matsuo et al. developed a system called Polypho-net, which is used for extracting and analyzing social networks of academic figures [2]. The authors of [3] demonstrated how social network data can be extracted from email communication. There also exists a system that extracts social network information from a user's email inbox [4]. The authors of [5] proposed an approach to resolve the problem of searching for people who share similar interests by using a person's website content to represent that person. Some studies have extracted networks from friend-of-a-friend (FOAF) documents [6-8]. Academic researchers' network extraction on the Web is also addressed [9,10]. Social networks on the Web have also been extracted by retrieving relationships between entities automatically derived from multilingual news [11], from log files of online shared workspaces [12], and from multiple unstructured sources on the Web [13]. Extraction of social networks has also been conducted via Internet and Networked Sensing [14]. The authors of [15] developed a model that helps estimate the strength of relationships by analyzing user interactions in online social networks.

The Web provides a rich set of data on virtually everything. For any given topic, one may obtain information from a variety of different sources. For example, one can extract information about a famous scientist from the scientist's own website, from a social networking site, or from a digital library with scientific papers. Each data source will almost certainly provide different information about this scientist's research activities. In other words, we need to rank these data sources and clearly identify how much we value each of these sources.

A social network is defined as a set of entities and the relationships among them. Edge weight in a social network is calculated based on information extracted from different sources. For example, say we need to compute the strength of the relationship between two scientists. Two data sources are used: co-authored papers and collaboration on a social networking site. We know that information about scientists' research relationship from co-authored papers should be more reliable than from a social networking site. However, we cannot ignore information about their collaboration on the social networking site. Hence, we assign more rank to the co-authored papers but still consider the networking site.

Most of the research done on social networks on the Web involves working with one data source only. In this case, the network is single-relational, *i.e.*, the weight value is of one type because it is extracted from a single

*Corresponding author.

data source. If we have multiple sources we can compute multiple edge weights. These weights are combined into a single weight, creating a multi-relational edge weight, and the resultant social network will be multi-relational. We assume multiple social networks are given, each with a single but constructed from a different data source (single-relational). For each pair of nodes, we aggregate weight values to obtain a single resultant weight value, creating a multi-relational network. The aggregation method allows us to control the “weight piece” in the resultant weight based on the original weights. The next sections provide a detailed description of the method.

2. Aggregating Weights

Suppose multiple single-relational networks given. To aggregate edge weights for the networks, we used an approach that we proposed previously [16]. The single-relational social networks are merged into a single multi-relational one. The edge weight between any two entities in the resultant network is the sum of the corresponding weights of the same nodes in the individual networks. Moreover, each weight in the sum is multiplied by a coefficient. The sum of these coefficients equals the unit. In other words, a single weight is made using the sum of weights from multiple networks; this single weight will be an edge weight of the resultant network. Formula 1 shows the process:

$$w_{ij} = \sum_{k=1}^n \tilde{w}_k w_{ij}^k, \quad k=1, \dots, n \quad (1)$$

where n is the number of networks, w_{ij} is the weight of an edge between the actors i and j of the resultant heterogeneous network, $\tilde{w}_k \in [0, 1]$ is a coefficient that shows the degree of importance of the corresponding data source k (it shows how much we value and hence rank the source), and w_{ij}^k is an edge weight between actors i and j from source k . Moreover, $\sum_{k=1}^n \tilde{w}_k = 1$.

Stated differently, the above formula represents an operator. Unknown variables are the weights of the single-relational networks, and the value of the operator is the weight assigned to the corresponding edge of the resultant network.

As an aggregation operator we use the *ordered weighted averaging operator* (OWA). OWA operators were first introduced by R. Yager [17] are defined as follows:

$$F : R^n \rightarrow R,$$

which has an associated weighting vector

$\tilde{W} = (\tilde{w}_1, \tilde{w}_2, \dots, \tilde{w}_n)$, such that $\tilde{w}_i \in [0, 1], 1 \leq i \leq n$ and $\sum_{k=1}^n \tilde{w}_k = 1$. Furthermore,

$$F(a_1, a_2, \dots, a_n) = \tilde{w}_1 b_1 + \tilde{w}_2 b_2 + \dots + \tilde{w}_n b_n, \quad (2)$$

where b_j is the largest element in the ordered bag $\langle a_1, a_2, \dots, a_n \rangle$. This ordered bag is called an ordered

collection vector.

A fundamental aspect of this operator is the re-ordering step. An aggregate a_i is not associated with a particular weight \tilde{w}_i but rather a weight is associated with a particular ordered position in the aggregate. Depending on the position in the sum, the same aggregate may have different weights.

Yager pointed out three important special cases of OWA aggregations:

- F^* : For this case, $\tilde{W} = \tilde{W}^* = (1, 0, \dots, 0)^T$ and $F^*(a_1, a_2, \dots, a_n) = \max\{a_1, a_2, \dots, a_n\}$;
- F_* : For this case, $\tilde{W} = \tilde{W}_* = (0, 0, \dots, 1)^T$ and $F_*(a_1, a_2, \dots, a_n) = \min\{a_1, a_2, \dots, a_n\}$;
- F_{mean} : For this case, $\tilde{W} = \tilde{W}_{mean} = (1/n, 1/n, \dots, 1/n)^T$ and $F_{mean}(a_1, a_2, \dots, a_n) = (a_1 + a_2 + \dots + a_n)/n$.

From the above it becomes clear that for any OWA operator F ,

$$F_*(a_1, a_2, \dots, a_n) \leq F(a_1, a_2, \dots, a_n) \leq F^*(a_1, a_2, \dots, a_n)$$

To classify OWA operators in regard to their location between “and” and “or,” a measure of *orness*, associated with any vector \tilde{W} , was introduced by Yager as follows:

$$orness(\tilde{W}) = \frac{1}{n-1} \sum_{i=1}^n (n-i) \tilde{w}_i.$$

It is easy to see that, for any \tilde{W} , the *orness* (\tilde{W}) is always in the interval $[0, 1]$. Furthermore, note that the nearer \tilde{W} is to an “or,” the closer its value is to one; while the nearer it is to an “and,” the closer it is to zero. Also note that for the vectors $\tilde{W}^* = (1, 0, \dots, 0)^T$, $\tilde{W}_* = (0, 0, \dots, 1)^T$ and $\tilde{W}_{mean} = (1/n, 1/n, \dots, 1/n)^T$, *orness* equals 1, 0, and 0.5, respectively.

Making analogues of formula (1) with formula (2), we find that b_j is the edge weight w_{ij}^k and the importance coefficient \tilde{w}_k in (1) is the corresponding weight vector value \tilde{w}_i in formula (2).

3. Finding Importance Coefficients

At this point, the edge weights of the resultant network could be calculated, except the values for the unknown coefficients in formula (2) have to be found. Several approaches have been proposed to identify weight vectors for OWA operators. We use the one presented in [18]. The authors present an analytical approach to determine the weight vector, achieved by transforming the following mathematical programming problem into a polynomial equation using the Lagrange multipliers method:

$$-\sum_{i=1}^n \tilde{w}_i \ln \tilde{w}_i \rightarrow \max,$$

$$\frac{1}{n-1} = \sum_{i=1}^n (n-i) \tilde{w}_i = \alpha, \quad 0 \leq \alpha \leq 1, \quad \sum_{i=1}^n \tilde{w}_i = 1, \quad 0 \leq \tilde{w}_i \leq 1, \quad i = 1, \dots, n.$$

Solving this, the following formulas are used to obtain the optimal values of the weight vectors:

$$\tilde{w}_1 \left[(n-1)\alpha + 1 - n\tilde{w}_1 \right]^n = \left((n-1)\alpha \right)^{n-1} \left[((n-1)\alpha - n)\tilde{w}_1 + 1 \right]$$

and

$$\tilde{w}_n = \frac{((n-1)\alpha - n)\tilde{w}_1 + 1}{(n-1)\alpha + 1 - n\tilde{w}_1}.$$

Note that in the case of $n = 3$, we see that $\tilde{w}_2 = \sqrt{\tilde{w}_1 \tilde{w}_3}$.

4. Experiments

Note that the purpose of this paper is not to compute edge strength but to aggregate already-computed weights. Rather than considering the methods for building a social network from a given data source, we assume that the necessary networks have already been constructed. This can be accomplished in multiple ways, for example, using a general search engine. To verify the outcome of the presented method experimentally we suppose that three social networks are constructed from different sources for given five actors, as shown in **Figures 1-3**. In each source, we possess information about one relationship only.

Table 1 shows the edge weights that we assigned to each of the networks.

Suppose now that $\alpha = 0.2$. In this case, we obtain a fourth-degree equation for finding the value of \tilde{w}_1 . Solving this equation, we receive four roots, of which three are negative and one is positive. As a value for \tilde{w}_1 , we pick

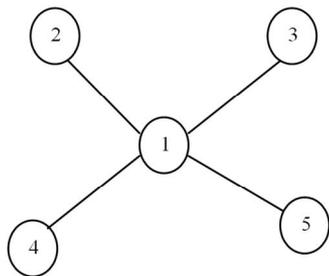


Figure 1. Social network from the first data source.

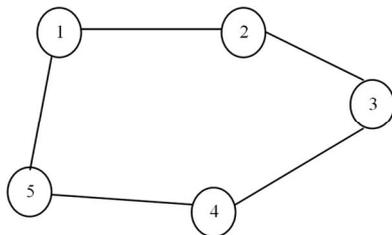


Figure 2. Social network from the second data source.

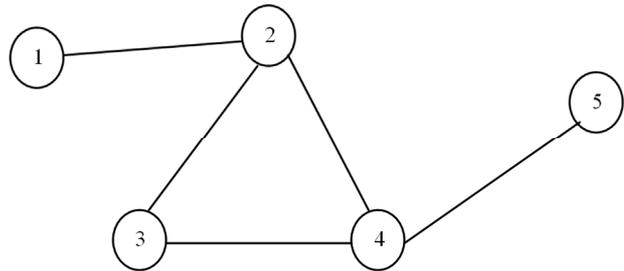


Figure 3. Social network from the third data source.

Table 1. Weight values for the social networks derived from the three data sources.

Pair	Network		
	1	2	3
w_{12}	0.21	0.50	0.22
w_{13}	0.10	0.00	0.00
w_{14}	0.30	0.00	0.00
w_{15}	0.30	0.80	0.00
w_{23}	0.00	0.70	0.30
w_{24}	0.00	0.00	0.10
w_{25}	0.00	0.00	0.00
w_{34}	0.00	0.40	0.80
w_{35}	0.00	0.00	0.00
w_{45}	0.00	0.92	0.91

up the positive root of 0.08. From the last equation for \tilde{w}_n , we find that $\tilde{w}_3 = 0.68$. Finally, from $\tilde{w}_2 = \sqrt{\tilde{w}_1 \tilde{w}_3}$, we determine that $\tilde{w}_2 = 0.23$. At this point, the corresponding edge weights for the resultant network can be found. For example, for w_{12} the ordered collection vector is $B = (0.5, 0.22, 0.21)$. Hence,

$$w_{12} = F(w_{12}^1, w_{12}^2, w_{12}^3) = [\tilde{w}_1, \tilde{w}_2, \tilde{w}_3] B = [0.08, 0.23, 0.68] \begin{bmatrix} 0.5 \\ 0.22 \\ 0.21 \end{bmatrix} = 0.23$$

Proceeding in this way, we can calculate all of the edge weights for the resultant network. For comparison purposes, we find these values for both $\alpha = 0.5$ and $\alpha = 0.8$.

For each value of α , we derive a separate network. In each of these cases, edge weight will be an aggregate of corresponding edge weights from the single-relational networks, and thus the resultant network will be multi-relational. A previous study demonstrated a method for building a multi-relational network when multiple single-relational networks are given [10]. We used the same network data from [10] in this study to compare the results. **Table 2** compares the results of this study to those of [10] (the rightmost column).

Table 2. Resultant network edge weights.

w	α			\hat{w}
	0.2	0.5	0.8	
w_{12}	0.230	0.465	0.378	0.390
w_{13}	0.008	0.033	0.054	0.015
w_{14}	0.024	0.099	0.162	0.044
w_{15}	0.133	0.363	0.552	0.560
w_{23}	0.125	0.330	0.408	0.510
w_{24}	0.008	0.033	0.054	0.021
w_{25}	0.000	0.000	0.000	0.000
w_{34}	0.156	0.396	0.552	0.430
w_{35}	0.000	0.000	0.000	0.000
w_{45}	0.283	0.604	0.771	0.770

An interesting observation is that formula (1) guarantees that we will never lose information about the existence of a relationship. For both methods, the resultant weight value of an edge equals zero if there is no relationship between the edge nodes in all of the sources. The resultant weight value, \hat{w} from the table is closer to w when $\alpha = 0.8$ compared to the other cases.

There is no point in computing weight values for the resultant network for $\alpha = 0$ and $\alpha = 1$. In the first case, the weight values of the resultant network will match the values of the network in **Figure 3**. In the second case $\alpha = 1$ will coincide with **Figure 1**. When $\alpha = 0.5$, each value of the single-relational network is equal, *i.e.* $\tilde{w}_1 = \tilde{w}_2 = \tilde{w}_3 = 0.33$. In this case, the weight values of the original networks do not necessarily matter because each of the weights will be multiplied by the same coefficient.

In other cases, a comparative analysis needs to be conducted. In these cases, it is normal for an edge weight of the resultant network to be less than that of any of the original networks. This occurs because, for different cases of *orness*, we assign more or less rank (\tilde{w}_i) to a data source. It is the value of *orness* that determines the importance of each data source. By manipulating the value of *orness*, different versions of the resultant multi-relational network are obtained. Each of them, to some degree, will be what the decision-maker expects. Choice of *orness* depends entirely on the decision-maker. From **Table 2** it can be seen that, although single-relational networks are the same in both methods, the resultant multi-relational network edge weights may be different. There is no direct relationship between the methods. The only relationship is the objective: to merge multiple networks into a single one.

5. Conclusion

The relationship strength for a given set of entities can be

extracted from various sources on the Web. Due to the large volume of information available, data received will rarely be the same. To evaluate these data, there needs to be a way to rank these sources. Omitting information from any of the sources may lead to incorrect results when computing overall relationship strength between entities. In this study, we have provided a new way to aggregate edge weights received from multiple networks; these networks, in turn, are built from different sources. Using the concept of OWA operators, we have demonstrated how aggregation can be achieved for social networks. In addition, we compared our results to an alternative method. Currently, we do not possess software that would allow us to calculate weight values for networks with a large number of nodes. Thus, we are unable to apply the above formulas to huge networks. In the future, we hope to use real data in these formulas to demonstrate real-world applications.

REFERENCES

- [1] H. Kautz, B. Selman and M. Shah, "The Hidden Web," *AI Magazine*, Vol. 18, No. 2, 1997, pp. 27-36.
- [2] Y. Matsuo, J. Mori and M. Hamasaki, "Polyphonet: An Advanced Social Network Extraction System from the Web," *Proceedings of the 15th International Conference on World Wide Web*, Edinburgh, 22-26 May 2006, pp. 397-406. [doi:10.1145/1135777.1135837](https://doi.org/10.1145/1135777.1135837)
- [3] R. Rome, G. Creamer, S. Hershkop and S. Stolfo, "Automated Social Hierarchy Detection through Email Network Analysis," *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*, California, 12-15 August, 2007, pp. 109-117.
- [4] A. Culotta, R. Bekkerman and A. McCallum, "Extracting Social Networks and Contact Information from Email and the Web," *Proceedings on the 1st Conference on Email and Anti-Spam*, California, 30-31 July, 2004.
- [5] Q. Li and Y. B. Wu, "People Search: Searching People Sharing Similar Interests from the Web," *Journal of the American Society for Information Science and Technology*, Vol. 59, No. 1, 2008, pp. 111-125. [doi:10.1002/asi.20736](https://doi.org/10.1002/asi.20736)
- [6] T. Finin, L. Ding and L. Zou, "Social Networking on the Semantic Web," *The Learning Organization*, Vol. 12, No. 5, 2005, pp. 418-435. [doi:10.1108/09696470510611384](https://doi.org/10.1108/09696470510611384)
- [7] J. Goldbeck and M. Rothstein, "Linking Social Networks on the Web with FOAF," *Proceedings of the 23rd National Conference on Artificial Intelligence*, Chicago, July 13-17, 2008, pp. 1138-1143.
- [8] B. Aleman-Meza, M. Nagarajan, C. Ramakrishnan, *et al.*, "Semantic Analytics on Social Networks: Experiences in Addressing the Problem of Conflict of Interest Detection," *Proceedings of the 15th International World Wide Web Conference*, Edinburgh, 22-26 May 2006, pp. 407-416. [doi:10.1145/1135777.1135838](https://doi.org/10.1145/1135777.1135838)

- [9] J. Tang, D. Zang and L. Yao, "Social Network Extraction of Academic Researchers," *Proceedings of the 7th IEEE International Conference on Data Mining*, Omaha, 28-31 October 2007, pp. 292-301.
- [10] R. Alguliev, R. Aliguliyev and F. Ganjaliyev, "Extracting a Heterogeneous Social Network of Academic Researchers on the Web Based on Information Retrieved from Multiple Sources," *American Journal of Operations Research*, Vol. 1, No. 2, 2011; pp. 33-38.
[doi:10.4236/ajor.2011.12005](https://doi.org/10.4236/ajor.2011.12005)
- [11] B. Poliqean, H. Tanev and M. Atkinson, "Extracting and Learning Social Networks out of Multilingual News," *Proceedings of the Social Networks and Application Tools Workshop (SocNet-08)*, Skalica, 19-21 September 2008, pp. 13-16.
- [12] P. Nasirifard, V. Peristeras, C. Hayes and S. Decker, "Extracting and Utilizing Social Networks from Log Files of Shared Workspaces," *10th IFIP Working Conference on Virtual Enterprises*, Thessaloniki, 7-9 October 2009, pp. 643-650.
- [13] G. Geleijnse and J. Korst, "Creating a Dead Poets Society: Extracting a Social Network of Historical Persons from the Web," *Proceedings of the 6th International and 2nd Asian Conference on Asian Semantic Web Conference*, Busan, 11-15 November 2007, pp. 155-168.
- [14] T. Nishimura and Y. Matsuo, "A Method of Social Network Extraction via Internet and Networked Sensing," *Proceedings of the 3rd International Conference on Networked Sensing Systems*, Chicago, 31 May-2 June 2006, pp. 123-127.
- [15] R. Xiang, J. Neville and M. Rogati, "Modeling Relationship Strength in Online Social Networks," *Proceedings of the 19th International World Wide Web Conference*, New York, 26-30 April 2010, pp. 981-990.
[doi:10.1145/1772690.1772790](https://doi.org/10.1145/1772690.1772790)
- [16] F. Ganjaliyev, "Building a Heterogeneous Social Network of Academic Researchers," *Proceedings of the 3rd International Conference of Problems of Cybernetics and Informatics*, Baku, 6-8 September 2010, pp. 179-182.
- [17] R. R. Yager, "On Ordered Weighted Averaging Aggregation Operators in Multicriteria Decision Making," *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 18, No. 1, 1988, pp. 183-190. [doi:10.1109/21.87068](https://doi.org/10.1109/21.87068)
- [18] R. Fuller and P. Majlende, "An Analytic Approach for Obtaining Maximal Entropy OWA Operator Weights," *Fuzzy Sets and Systems*, Vol. 124, No. 1, 2001, pp. 53-57.
[doi:10.1016/S0165-0114\(01\)00007-0](https://doi.org/10.1016/S0165-0114(01)00007-0)