

# Transforming Collaboration Data into Network Layers for Enhanced Analytics

Saharnaz E. Dilmaghani<sup>1</sup>, Apivadee Piyatumrong<sup>2</sup>, Pascal Bouvry<sup>1</sup>, Matthias R. Brust<sup>1</sup>

<sup>1</sup>Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg

<sup>2</sup>National Electronics and Computer Technology Center (NECTEC), Thailand

saharnaz.dilmaghani@uni.lu, apivadee.piy@nectec.or.th, pascal.bouvry@uni.lu, matthias.brust@uni.lu

**Abstract**—We consider the problem of automatically generating networks from data of collaborating researchers. The objective is to apply network analysis on the resulting network layers to reveal supplemental patterns and insights of the research collaborations. In this paper, we describe our data-to-networks method, which automatically generates a set of logical network layers from the relational input data using a linkage threshold. We, then, use a series of network metrics to analyze the impact of the linkage threshold on the individual network layers. Moreover, results from the network analysis also provide beneficial information to improve the network visualization. We demonstrate the feasibility and impact of our approach using real-world collaboration data. We discuss how the produced network layers can reveal insights and patterns to direct the data analytics more intelligently.

## I. INTRODUCTION

Network structures have drawn significant attention in big data due to the possibility to apply network theory and analysis to obtain extra insights from the data. Networks are ubiquitous [27] in research areas from biology and neuroscience (e.g., brain networks [6]) to modeling and analyzing galaxy distributions [17], and quantifying reputation in art [15]. These examples are use cases where network layers play an important role to represent and analyze the data.

Networks offer several advantages. They provide adaptability for dynamic structures benefiting from local principles on nodes [9]. Additionally, network analysis discloses valuable information for data visualization by defining an appropriate link definition. Networks often provide computationally efficient algorithms with lower complexity in comparison to a tabular structure [13]. Furthermore, data transformed into network structures can help providing evidence for missing information [30], [21] as well as predicting forthcoming events [26]. Besides, there are numerous algorithms that can be applied on networks such as *Louvain's* algorithm which is a community detection algorithm [7] and *Page Rank*, that identifies the most influential object within a network.

With these advantages of networks, we are confronted with the challenge on how to transform relational data into appropriate networks, which exhibit advantages for data analytics as well as data visualization. The challenge is twofold: It is not only on how to represent the elements of a network, but also the specific construction principles, since, for each dataset, there are numerous ways be transformed into a network representation [10]. Each network reveals a particular perspective

on the dataset. In this study, we investigate different linkage thresholds for the transformation of a collaboration dataset into networks.

In this paper, we propose a method that transforms collaboration data to network layers. Our approach favors scientific projects as nodes with links generated by using a specific linkage threshold. We apply our method on real-world data that describes collaboration of a renowned research institute. Our study uses different metrics to determine the influence of the network structure on the network properties. Additionally, we obtain results that provide information on better visualizing the produced networks and a possible interface to include privacy mechanisms into data analytics [28]. The remainder of this paper is structured as follows. Section II investigates related work. We describe our method to complement data with network layers in Section III. The experiment setup designed to analyze the proposed algorithm is represented in Section IV. Section V discusses the outcome and the opportunities for future work. Finally, Section VI concludes the paper.

## II. RELATED WORK

The challenge of converting data to networks is a well-known issue when it comes to geographical data [14], [16]. Graph studies on spatial data reveal valuable information on route networks, complex urban systems [22] and the relationship between different urban areas [31]. Nevertheless, the raw dataset on geographical information is not enough by its own to conduct proper graph studies. Karduni et al. [18] focuses on this challenge and has introduced an approach for geographical systems by defining a protocol describing the network properties to convert the spatial polyline data into a network.

In general, some of the studies have stressed the inference of links from relational data to design a network out of the relational data. Casiraghi et al. [11] developed a generalized hypergeometric ensembles approach to address the problem of inferring connections within relational data. The study represents a perspective of link prediction while applying predictive analysis. From a similar point of view, Xiang et al. [29] established a link-based latent variable model to infer the friendship relations within a social interaction. In addition, in another study [25] the international relations from a dataset consists of the news from different countries are extracted by a tensor factorization technique. Moreover, Akbas et al. [1]

proposed a social network generation by proposing a model based on various interactions (e.g., phone calls) considering smartphone data. Given a weight to each type of interaction, the authors define a link value as the combination of various interaction types. Akbas et al. followed up their study on network generation from interaction patterns by studies on how to infer social networks of animal groups [2], [3], [4]. Initiating from these studies, in particular from [1], we followed a similar approach in order to define the linkage threshold to generate the network layers [5], [12].

Considering the collaboration data, Newman [19] has established networks considering authors and their collaboration on scientific papers. The scientific collaboration networks have been also studied in a particular network structure, *hypergraphs*, by Ouvrard et. al. [20]. The authors emphasized on enhancing the visualization of these networks considering network properties.

### III. PROPOSED METHODOLOGY: DATA-TO-NETWORK LAYERS

We establish a method with the purpose to convert the relational data of research collaborations into network layers by describing a set of nodes and a linkage threshold to define the connections between nodes. Considering the thresholds we then generate different network layers and investigate each network with the network measurements. Using outcomes of the last phase we then visualize the networks for a better understanding of the relationship in our data.

For this study, we have access to a real-world dataset of collaborative projects within National Electronics and Computer Technology Center (NECTEC). The dataset consists of various projects that have been completed with certain members of the institute. Investigating on the relationships and extracting the collaboration patterns are the outcomes that the networks will provide in order to improve performance of the research projects by using the possible resources inside the organization. We, therefore, retrieved different features from the dataset to feed our method and generate the network layers.

In this study, our focus is particularly on the projects of the institute. Thus, we consider the elements of the network such that the projects are nodes of the network and the common members between the projects are represented as the links of the network. Each project is identified with an *id*, and a team of *m* members who contributed to the project, whereas the project members can participate in different projects at the same time. Contribution percentage is a feature extracted from the dataset that quantifies the contribution of a member within a project.

We calculate the linkage threshold by looking at certain features from the dataset: (1) the project members, and (2) the contribution percentage of each project member in a project. Let  $m_k^i, c_k^i$  be the *k*-th member of project *i* and the contribution percentage of the member in the project, respectively. Let  $m^i = (m_1^i, m_2^i, \dots)$  and  $c^i = (c_1^i, c_2^i, \dots)$  of project  $i \in \mathcal{ID}$ . Next, we define  $\mathcal{P}$  as the combination of features chosen to explain the explain linkage threshold, i.e.,  $\mathcal{P} = \bigcup_i (m^i, c^i)$ .

Then, the common members between two projects (e.g., *i, j*) is denoted as  $\mathcal{M}_{ij} = \{(i, j) \in \mathcal{ID} \mid m^i \cap m^j\}$ . Finally, considering the aforementioned parameters, the linkage threshold is defined as

$$\mathcal{T}_{ij} = \frac{1}{n_{ij}} \sum_{\forall k \in \mathcal{M}_{ij}} \frac{c_k^i + c_k^j}{2},$$

$$\mathcal{T} = \bigcup_{i, j \in \mathcal{ID}} (\mathcal{T}_{ij}), \quad (1)$$

where  $\mathcal{T}_{ij}$  is the linkage threshold within the projects (*i, j*),  $n_{ij}$  is the number of common members calculated in  $\mathcal{M}_{ij}$ , and  $\mathcal{T}$  is the linkage threshold for the whole dataset. Exploiting the linkage threshold, we propose Algorithm 1 to construct network layers with different linkage thresholds.

---

#### Algorithm 1 Data-to-Network Layers

---

**Input:**  $D$ , a dataset of research collaboration.

**Output:**  $\mathcal{G}$ , a vector of generated network layers.

---

```

1: procedure TRANSFORM-TO-NETWORK( $D$ )
2:    $nodesList \leftarrow \mathcal{ID}$ 
3:    $\mathcal{P} \leftarrow$  extract features from  $D$ 
4:    $trange \leftarrow$  generate a space vector from  $(\min(\mathcal{T}), \max(\mathcal{T}))$ 
5:   for each  $threshold$  in  $trange$  do
6:     for each pair of nodes in  $nodesList$  do
7:        $f \leftarrow$   $Aggregate(\mathcal{P}[nodes])$ 
8:       if  $f \geq threshold$  then
9:          $edgesList \leftarrow$  pair of nodes
10:       $Network\ G \leftarrow GenNet(nodesList, edgesList)$ 
11:      Insert  $G$  to  $\mathcal{G}$ 
12:   return  $\mathcal{G}$ 

```

---

*a) Description of the algorithm::* Assume  $\mathcal{D}$  is the relational dataset of collaborations. We need to define the nodes and the links extracting particular features from  $\mathcal{D}$  to generate the network layers. To define the list of nodes  $nodesList$ , we extract the *id* of the entities which in this study we used the identities of research projects (see line 2 in Algorithm 1). In order to describe the links, we first need to define a linkage threshold  $t$  considering the list of features  $\mathcal{P}$ . The linkage threshold of linear space vector consists of  $n$  points of  $threshold$  within the range of minimum and maximum values of  $\mathcal{P}$  (see line 3–4 in Algorithm 1). Additionally, each network from  $\mathcal{G}$  is measured by the set of network metrics.

We use the before mentioned collaboration data as input for the presented Algorithm 1. Then, we measure the network metrics and utilize them for a better visualization of data. The resulting generated network layers are visualized in Fig. 1. In this specific example, we consider 6 thresholds for defining the edges starting from 0 (the minimum collaboration observed in the dataset) to 100 by using the connected components to enhance the visualization of data.

*b) Complexity analysis:* The complexity of Algorithm 1 depends on the two main parts of the algorithm: (1) Comparing each pair of nodes to find those that serve the determined condition for threshold is the most expensive one with the

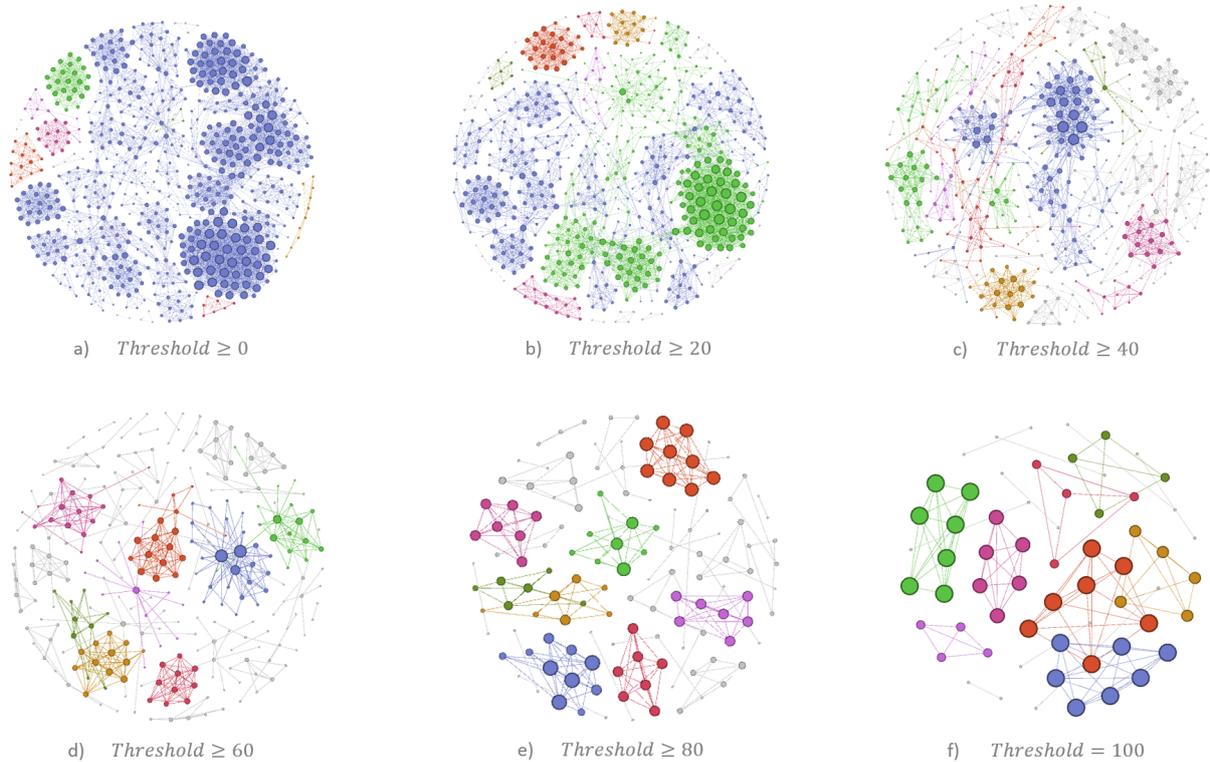


Fig. 1: Visualization of the generated networks. In each network, the size of the nodes represents the degree of a node and the color illustrates the components. Such that blue shows components with the highest number of nodes, whereas gray represents the smallest components of a network. Moreover, green and red describe components which have a number of nodes within the range of previous cases.

complexity of  $O(n^2)$ . (2) The complexity of the network generation is linear such that for  $n$  nodes and  $m$  edges the complexity is  $O(n + m)$ .

#### IV. EXPERIMENT SETUP

##### A. Dataset

We benefit from a particular collaboration data derived from the *National Electronics and Computer Technology Center (NECTEC)* that presents different projects and collaborations in the area of R&D. The dataset is stored in a relational database consisting of research projects conducted between July 2013 and July 2018. Each project may consist of different deliverables: *intellectual property (IP)*, *papers*, or *prototypes* as well as may comprise different members from different teams of the same institute. The dataset is the knowledge management about the project where the key information is to know (1) the type of the project, (2) project contributors and contributions.

The dataset of combined team tables have almost 8k records which is the information of more than 2.3k projects. Among them 630 are related to IP, 1717 to papers, and 539 to prototypes. Overall, the institute has more than 1000 members who are contributing on different projects with certain features (e.g., contribution percentage) which have been defined within the organization to evaluate the contributions. One of the main features we have used is the contribution percentage. The total

percentage assigned to each project is 100% that is divided between the project members according to their contribution on the project. Furthermore, IC-score is another feature that is developed by the institute and it illustrates the value of each project (e.g., prototype) based on its status (e.g., lab, industrial). To obtain the IC-score for each member, the total IC-score value of each project is divided by the contribution percentage of each member. The details regarding the values of both features (contribution percentage and IC-score) have been further discussed further in Section V.

##### B. Network Metrics

A network (or graph)  $G = (V, E)$  consists of a set of vertices  $V$  which are connected by the edges from set  $E$ . There exists different parameters (e.g., centrality measures) to analyze and study the networks. We choose centrality measures to analyze the generated network layers, which help to find the most important vertices within a network. Besides the centrality measures, we also consider other metrics such as network density and connected components to analyze the properties of the network. The following is a brief description of each metric.

*a) Closeness Centrality:* defines the closeness of a node to other nodes by measuring the average shortest path from that node to the all other vertices within the network. Hence, the more central a node is, the closer it is to all other nodes

[23] calculated as  $C_C(v) = \sum_y \frac{1}{d(v,u)}$ , where  $d(v,u)$  is the distance between vertices  $v$  and  $u$ .

b) *Betweenness Centrality*: indicates the number of times a node acts as a bridge along the shortest path between two other nodes. For a given node the number of shortest paths that passes through the node implies the betweenness centrality of the node. Nodes with high betweenness may have significant influence in a network due to their control over the flow of information passing between others through them. In a network  $G = (V, E)$  betweenness centrality for node  $v$  is [8]:  $C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$  where  $\sigma_{st}$  total number of shortest paths from node  $s$  to node  $t$  and  $\sigma_{st}(v)$  is the number of those paths that pass through  $v$ .

c) *Degree Centrality*: identifies the number of direct links which are connected to a vertex within the network. The importance of the nodes with higher degree is due to the immediate risk of these node while some information is flowing through the network. The degree of a node  $v$  is represented as,  $C_D(v) = deg(v)$ .

d) *Clustering Coefficient*: presents the likelihood of nodes in a network that tend to cluster together. The value of clustering coefficient lies between 0 and 1. When a network is clique which means that every two distinct vertices are adjacent, the value is 1, however, in a star network in which a node's neighbours are not connected to each other at all, clustering coefficient is 0. For an unweighted network, the clustering of a node  $v$  is the fraction of possible triangles through that node that exist,  $CC(v) = \frac{2T(v)}{deg(v)(deg(v)-1)}$  where  $T(v)$  is the number of triangles through node  $v$  and  $deg(v)$  is the degree of  $v$  [24].

e) *Network Density*: is the ratio of potential links to existing links in a network. The range of this metric varies from 0 for a network with no links (sparse network) and 1 for networks with all possible links (dense network).  $nd = \frac{2m}{n(n-1)}$ , where  $n$  is the number of nodes and  $m$  is the number of edges in network  $G$ .

f) *Connected Components*: are sub-networks in which there are at least two vertices connected to each other through a path. In other words, two vertices are in the same sub-network if there is a path between them in the network. We use  $n_{comp}$  as a notation to address the connected components in this the paper.

It is to be observed that the chosen metrics are (1) *local* when they are only considering a node itself and the information of its neighbour to calculate, centrality measures are of this category, or (2) *global* when they calculate a parameter considering the whole knowledge of network properties, nodes and edges, such as network density. Exploiting these metrics on networks generated from collaborative data, the purpose is to analyze the networks from different perspectives. Centrality measures are indicators to define the important node within a network. For instance, closeness centrality defines whether a project has a higher value for the institute such as delivering different outcomes (IP, papers, and prototypes). Moreover, there might be an argument regarding the topic of the project such that it is covering fundamental topics which other projects

need to collaborate with. Degree centrality, represents the members' collaboration of a project with other projects. On the other hand, metrics like network density and connected components illustrates the general overview of a network. Network density represents how much a network is away from being a fully collaborative network in which all projects are connected together.

## V. RESULTS OF THE NETWORK ANALYSIS

### A. Data Analysis

We perform a preliminary analysis on the data set in order to conduct on the linkage threshed. We exploit histograms to plot the frequency of the score and contribution percentage. Fig. 2 presents the histograms of IC-score and contribution percentage for each member in the collaboration data regarding all projects. The histograms represent the number of members with a certain value of IC-score (or contribution percentage) in the dataset. Moreover, mean, standard deviation, and variance are calculated for IC-score which are 3.16, 4.24, and 1.79, respectively. For contribution percentage, the mean, standard deviation and variance are obtained as 23.30, 22.80 and 5.20, respectively. IC-score concentrates on lower values better than contribution percentage, nevertheless, contribution percentage represent the dataset better as the coverage range is broader. Thus, we construct the linkage threshold defined in Equation (III) considering the contribution percentage.

### B. Network Analysis

We applied the proposed Algorithm 1 on our collaboration dataset. As a result, we obtained a vector of network layers, each represents a certain linkage threshold. For each network we calculated the set of network metrics which has been introduced in Section IV.

We chose 6 linkage thresholds which are 0, 20, 40, 60, 80, and 100. For instance, with the linkage threshold equal to 20 two projects in the network are connected if the average contribution percentage of the common members between those projects is equal or greater than 20%. Thus, those two projects are neighbours in the network. While increasing the threshold, the number of nodes that could not satisfy the condition increases dramatically. Thus, the number of isolated nodes increases which impact the outcomes of network metrics. In order to analyze the network regardless of the influence of these nodes, the network metrics are applied after removing the isolated nodes. We measured Betweenness Centrality  $C_B(v)$ , Degree Centrality  $C_D(v)$ , Closeness Centrality  $C_C(v)$ , Clustering Coefficient  $CC(v)$ , Network Density  $nd$ , and Connected Components  $n_{comp}$  for each network. Additionally, for the *local* metrics, we calculated the average of nodes for the whole network.

We first applied our algorithm considering only IP projects data, and then on the combination of all projects (i.e., IP, paper, and prototype). Fig. 3 describes the metrics on networks that are constructed particularly on IP dataset and Fig. 4 provides the results for a similar setup while considering the combination of all projects.

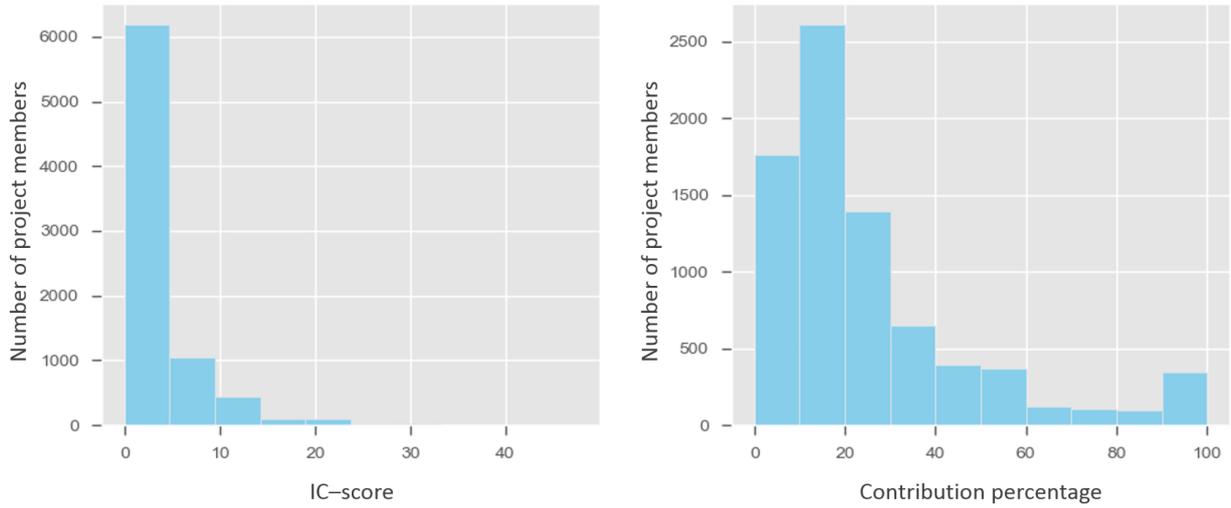


Fig. 2: Histograms of IC-score and contribution percentage in the collaboration dataset.

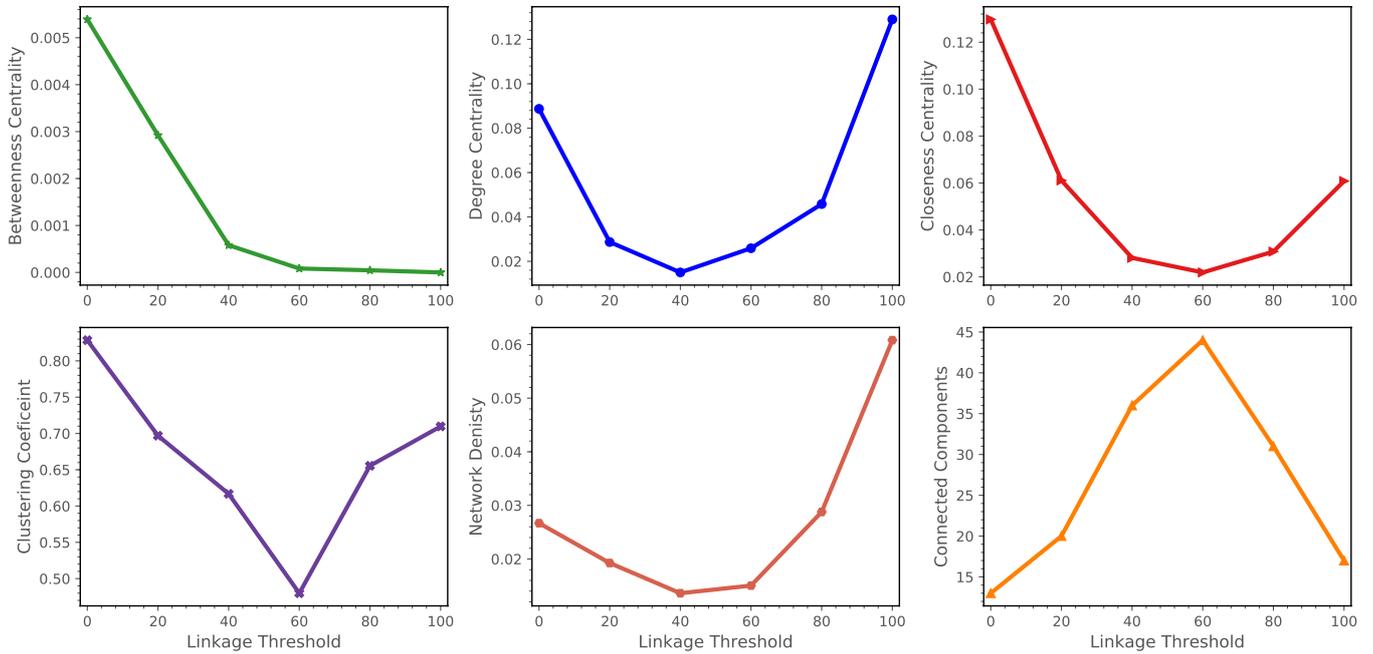


Fig. 3: The effect of the set of network metrics,  $C_C(v)$ ,  $C_B(v)$ ,  $C_D(v)$ ,  $CC(v)$ ,  $nd$ ,  $n_{comp}$ , on the generated network layers from IP projects. The result in each plot represents the average for all the nodes within a network layer. Thus, the x-axis illustrates the *linkage threshold* in which the corresponding network has been created and the metric is measured.

According to Fig. 3 starting from 0 as the linkage threshold increases the values of betweenness, degree and closeness centrality, clustering coefficient, network density decrease whereas the number of connected components increases until the linkage threshold is 60. However, the pattern of the results for all metrics have dramatically changed within the range of 40 to 60. Besides the outcomes from Fig. 3, that represents the result only for the IP projects, Fig. 4 provides the results of applying network metrics including all types of projects (IP, papers, and prototype). The results of both figures (Fig.

3 and 4) are representing the similar patterns. In other words, the topology of networks are not any different from one project type to another. Furthermore, although the linkage threshold equal to 0 provides detailed information of projects, the linkage threshold equal to 100 represents a particular perspective of the dataset which describes the main leaders contribution in different projects.

### C. Optimization and Future Work

Our methodology to construct network layers from collaboration data reveals several optimization criteria. Optimizing the

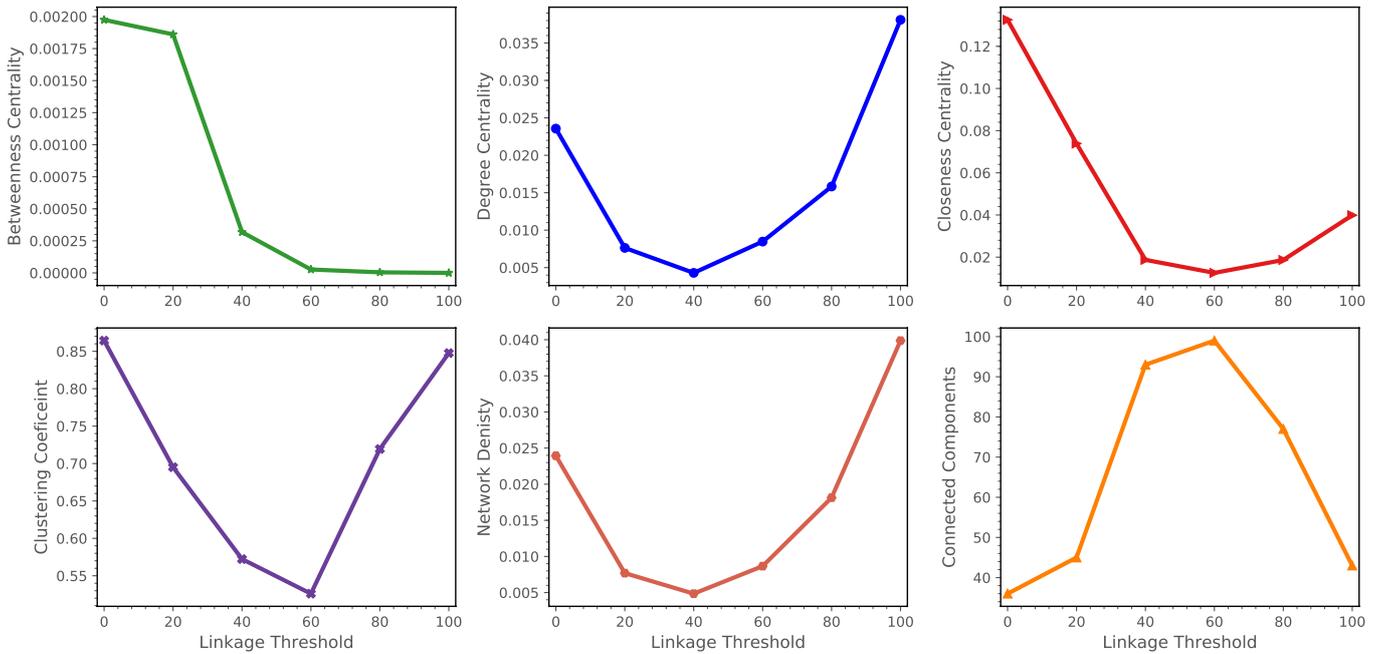


Fig. 4: The effect of the set of network metrics on the generated network layers from IP, paper, and prototype projects.

number of network layers while still containing the maximum on distinct information for enhanced analytics is a challenging task. Moreover, the linkage threshold we have defined in this paper can be generalized to a utility function to be performed on any given collaboration dataset. In addition, deciding on an optimal linkage threshold based on predefined criteria and conditions could further improve the performance, but also widen the applicability, of our algorithm. Additionally, we will consider different network representations for the same data in future work. We also plan on using more real-world collaboration data from distinct sources to further generalize our approach.

## VI. CONCLUSION

The approach outlined in this paper infers possible collaboration networks of researchers within projects of an organization. Our method uses a linkage threshold to automatically generate these network layers from the relational input data. We conducted a network analysis on the produced networks using metrics such as clustering coefficient, closeness and betweenness centrality, and illustrate their impact on the different network layers. We, then, utilize the results of the metrics as an important input to visualize the generated graph in each configuration. We conclude that the linkage threshold has a crucial impact on the network properties and must be chosen with caution. Additionally, the influence of the linkage threshold on the results of the metrics indicates that the network representation can be optimized.

**Acknowledgement.** This work is partially funded by

the joint research programme UL/SnT-ILNAS on Digital Trust for Smart-ICT.

## REFERENCES

- [1] M. Akbas, M. Brust, and D. Turgut. Social network generation and role determination based on smartphone data. In *IEEE International Conference on Computer Communications (INFOCOM) Student Workshop*, 2012.
- [2] M. I. Akbas, M. R. Brust, C. H. C. Ribeiro, and D. Turgut. Deployment and mobility for animal social life monitoring based on preferential attachment. In *IEEE Conference on Local Computer Networks*, pages 484–491, Oct 2011.
- [3] M. I. Akbas, M. R. Brust, C. H. C. Ribeiro, and D. Turgut. facebook - animal social life monitoring with wireless sensor and actor networks. In *IEEE Global Telecommunications Conference - GLOBECOM*, pages 1–5, Dec 2011.
- [4] M. I. Akbas, M. R. Brust, D. Turgut, and C. H. Ribeiro. A preferential attachment model for primate social networks. *Computer Networks*, 76:207 – 226, 2015.
- [5] A. Andronache, M. R. Brust, and S. Rothkugel. Hycast-podcast discovery in mobile networks. In *Proceedings of the 3rd ACM workshop on Wireless multimedia networking and performance modeling*, pages 27–34. ACM, 2007.
- [6] D. S. Bassett, P. Zurn, and J. I. Gold. On the nature and use of models in network neuroscience. *Nature Reviews Neuroscience*, page 1, 2018.
- [7] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.
- [8] U. Brandes. A faster algorithm for betweenness centrality. *Journal of mathematical sociology*, 25(2):163–177, 2001.
- [9] M. R. Brust, H. Frey, and S. Rothkugel. Adaptive multi-hop clustering in mobile networks. In *Int. Conf. on mobile technology, applications, and systems*, pages 132–138. ACM, 2007.
- [10] C. T. Butts. Revisiting the foundations of network analysis. *science*, 325(5939):414–416, 2009.
- [11] G. Casiraghi, V. Nanumyan, I. Scholtes, and F. Schweitzer. From relational data to graphs: Inferring significant links using generalized hypergeometric ensembles. In *International Conference on Social Informatics*, pages 111–120. Springer, 2017.

- [12] J. Chen, M. R. Brust, A. R. Kiremire, and V. V. Phoha. Modeling privacy settings of an online social network from a game-theoretical perspective. In *9th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing*, pages 213–220. IEEE, 2013.
- [13] J. G. Davis, J. K. Panford, and J. B. Hayfron-Acquah. Big and connected data analysis with graph and relational databases using collaborative filtering technique. *Int. Journal of Computer Science and Information Security (IJCSIS)*, 15(12), 2017.
- [14] S. Derrible and C. Kennedy. Applications of graph theory and network science to transit network design. *Transport reviews*, 31(4):495–519, 2011.
- [15] S. P. Fraiberger, R. Sinatra, M. Resch, C. Riedl, and A.-L. Barabási. Quantifying reputation and success in art. *Science*, 2018.
- [16] M. T. Gastner and M. E. Newman. The spatial structure of networks. *The European Physical Journal B*, 49(2):247–252, 2006.
- [17] S. Hong, B. C. Coutinho, A. Dey, A.-L. Barabási, M. Vogelsberger, L. Hernquist, and K. Gebhardt. Discriminating topology in galaxy distributions using network analysis. *Monthly Notices of the Royal Astronomical Society*, 459(3):2690–2700, 2016.
- [18] A. Karduni, A. Kermanshah, and S. Derrible. A protocol to convert spatial polyline data to network formats and applications to world urban road networks. *Scientific data*, 3:160046, 2016.
- [19] M. E. Newman. Scientific collaboration networks. i. network construction and fundamental results. *Physical review E*, 64(1):016131, 2001.
- [20] X. Ouvrard, J.-M. L. Goff, and S. Marchand-Maillet. Networks of collaborations: Hypergraph modeling and visualisation. *arXiv preprint arXiv:1707.00115*, 2017.
- [21] L. Pan, T. Zhou, L. Lü, and C.-K. Hu. Predicting missing links and identifying spurious links via likelihood analysis. *Scientific reports*, 6:22955, 2016.
- [22] F. Peiravian, A. Kermanshah, and S. Derrible. Spatial data analysis of complex urban systems. In *IEEE Int. Conf. on Big Data*, pages 54–59. IEEE, 2014.
- [23] G. Sabidussi. The centrality index of a graph. *Psychometrika*, 31(4):581–603, 1966.
- [24] J. Saramäki, M. Kivela, J.-P. Onnela, K. Kaski, and J. Kertesz. Generalizations of the clustering coefficient to weighted complex networks. *Physical Review E*, 75(2):027105, 2007.
- [25] A. Schein, J. Paisley, D. M. Blei, and H. Wallach. Bayesian poisson tensor factorization for inferring multilateral relations from sparse dyadic event counts. In *ACM Int. Conf. on Knowledge Discovery and Data Mining*, 2015.
- [26] Z. Sha, Y. Huang, J. S. Fu, M. Wang, Y. Fu, N. Contractor, and W. Chen. A network-based approach to modeling and predicting product coconsideration relations. *Complexity*, 2018, 2018.
- [27] S. Shirinivas, S. Vetrivel, and N. Elango. Applications of graph theory in computer science an overview. *International Journal of Engineering Science and Technology*, 2(9):4610–4621, 2010.
- [28] F.-J. Wu, M. R. Brust, Y.-A. Chen, and T. Luo. The privacy exposure problem in mobile location-based services. In *2016 IEEE Global Communications Conference (GLOBECOM)*, pages 1–7. IEEE, 2016.
- [29] R. Xiang, J. Neville, and M. Rogati. Modeling relationship strength in online social networks. In *Proceedings of the 19th international conference on World wide web*, pages 981–990. ACM, 2010.
- [30] J. Yang and X.-D. Zhang. Predicting missing links in complex networks based on common neighbors and distance. *Scientific reports*, 6:38208, 2016.
- [31] C. Zhong, S. M. Arisona, X. Huang, M. Batty, and G. Schmitt. Detecting the dynamics of urban structure through spatial network analysis. *Int. Journal of Geographical Information Science*, 28(11):2178–2199, 2014.