NetDriller-V3: A Powerful Social Network Analysis Tool

Salim Afra[1], Tansel Özyer[2], Jon Rokne[1], Reda Alhajj[1,3,4]

[1]Department of Computer Science, University of Calgary, Alberta, Canada
[2]Department of Computer Engineering, Ankara Medipol University, Ankara, Turkey
[3]Department of Computer Engineering, Istanbul Medipol University, Istanbul, Turkey
[4]Department of Health Informatics, University of Southern Denmark, Odense, Denmark

***Abstract*—The development in technology has led to the generation of huge amounts of data from various sources, including biological data, social networking data, etc. Accordingly, social network analysis has received considerable attention with the availability of more raw datasets which could be realized using a network structure. Most of the datasets can be represented as a social network which is a graph consisting of actors having relationships. Many tools exist for social network analysis inspired to extract knowledge from the networks. NetDriller has been developed as a social network extraction, manipulation and analysis tool to cover the lack that exists in other tools. It is capable of constructing social networks from raw data by employing a variety of data mining and machine learning techniques. In this paper, we describe an extend version of NetDriller, which has some new essential functions, including social network construction using data collection from Twitter, DBLP and IEEE. We also added (1) a new chart for viewing the network property and metrics, and (2) new graph manipulation techniques using GUI to keep the tool up to date with the huge volume of networks and the different types of raw data available on the web.**

***Keywords*—Social Network Analysis, Data Mining, Machine Learning, Network Construction, Link Prediction, Hierarchical Zooming**

I. Introduction

In the first and second versions of NetDriller [11], [2], we presented a social network analysis tool that helps researchers and analysts in creating and analyzing social networks. A social network is a graph where nodes represent actors and edges represent relationship(s) between actors. A wide range of raw data types such as transactions, scholar, social media, molecular data (genes, proteins, 0=miRNAs, diseases, etc.) and organization structure data can be converted into social networks with the help of this tool.

In the last version of NetDriller described in this paper, we have added social network construction and analysis techniques from both social media and scholar data. With the ongoing growth of social media, it is essential for graph analysis tools to be able to handle such data to study the relationships between people, tags and the influence of people in social media type of data. Given social media data, we focused on generating new types of networks that are helpful in identifying key aspects of the data. For this, we generated three types of social graphs that can be extracted from the initial data, namely, a set of key terms, a set of tweeters, and a set for two-mode networks between tweeters and key terms.

Social graph analysis measures were also added in this new version of NetDriller to be able to handle big graphs where analyzing individual nodes is insufficient. For this, we introduced hierarchical zooming which refers to the function of incrementally clustering the social network nodes into groups to facilitate for zooming in and out. This analysis measure is required especially in large networks where many nodes and links exist, making it difficult to extract useful information from the network when considered at large scale. Instead, it is more manageable to deal with a concise network where communities are merged to be treated like single nodes in a zoom out mode.. We also added new analysis measures to predict hidden relationships in the network using link prediction.

NetDriller has been actively developed by our research group since 2009 to cover the social network development and analysis that lacks on other tools, and to add new functionalities with the emergence of new types of data. The project has been under constant updates; new functionalities are always added to the tool to cover the needed techniques and measures in the social network analysis domain. We make use of our expertise in machine learning and data mining to provide additional capabilities in network construction and manipulation.

The rest of this paper is organized as follows. Section II provides a brief overview of similar tools. Section III presents the new functions that have been added to NetDriller to improve its coverage and power. Section IV is the conclusion.

II. Similar Tools

In this section, we briefly cover some of the well known SNA tools. Several tools have been developed for social network analysis both for commercial and research purposes, e.g., citeVanschoren2014,Tkaczyk2014,bastian2009gephi. Bastian, et al. developed Gephi as an open source software for exploring and manipulating networks [3]. Batagelj, and Mrvar introduced Pajekfor handling large network analysis [4].

Borgatti, et al. [5] described Ucinet as a network analysis tool which works under windows. Camacho, et al. [6] discussed the four dimensions of social network analysis. They presented: an overview of research methods, applications, and

IEEE/ACM ASONAM 2022, November 10-13, 2022

software tools. Csardi, et al. [8] described the igraph software package for complex network research. Handcock, et al. [10] proposed Statnet as a software tool for the representation, visualization, analysis and simulation of network data. Michail, et al. [12] introduced JGraphTa as a Java library for graph data structures and algorithms.

Several other network analysis tools were discussed in the previous NetDriller paper [2]. NetMiner [1] is a commercial software which visualizes and analyzes networks to extract relevant information. The tool contains a data collector extension which collects data from social media contents (Facebook, YouTube, Twitter or Instagram); they also provide the ability to collect scholar data using their biblio data collector. We have added to NetDriller another source of data, mainly collecting data from the web (Google).

NodeXL [14] is a free, open-source SNA tool that can be added as a template for Microsoft Excel. The tool also provides a way to collect social media posts and exporting them to an excel sheet. The social network generated from the tool can then be used to apply Analysis techniques, such as centrality measures, groups identification and graph theory techniques. However, it is not possible to manually edit the structure of the network and add nodes on the spot as described in the sequel. Other popular social network analysis tools include Ucinet [5] and Gephi [3]. They offer construction and analysis of social graphs, such as calculating centrality measures, community detection, and shortest paths.

III. Functionalities Added

This section describes the new main functionalities added to the current version of NetDriller. These new functions include adding new techniques capable of collecting data from various sources to construct a social network, and some new measures for graph manipulation.

A. Data Collection

The first step of any social network analysis is to first find the right tool capable of appropriately analyzing the dataset. In this version of NetDriller, we have implemented a social network data extractor from a variety of sources (including Twitter, DBLP, IEEE, Web) to help users generate their own customized dataset which can then be imported to NetDriller for graph construction and analysis. This is a very helpful functionality as users don't have to use a separate tool or have to deal with generating their own dataset. Instead, users will have higher confidence in the whole process by having a single tool for data extraction and analysis; hence no gap between the two tasks will exist.

1) Twitter: It is very common to generate a social network from social media data. This leads to analyze the importance of people involved in the social media discussion, or how information is getting propagated in a network. Twitter has been heavily studied by researchers and has been proven to be very helpful source for predicting events and gathering people general opinion on topics. Based on this, Twitter data has been used to predict United State's presidential elections using sentiment analysis [15].

[1]NetMinerhttp://www.netminer.com Accessed on 13/10/2022

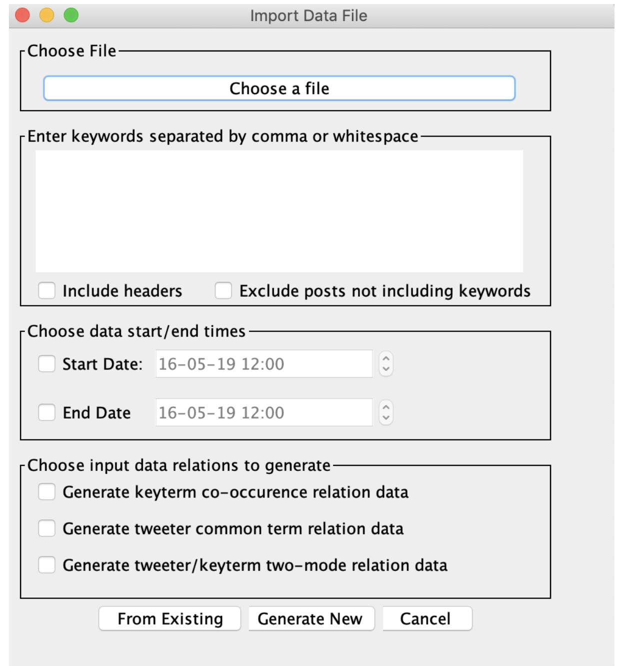

Fig. 1: Twitter GUI data collection

A menu item has been added to NetDriller to crawl Twitter data as shown in Figure 1. In this GUI, the user can either import a twitter dataset first to generate the different types of social network options listed at the bottom of the view, or enter keywords to be used for crawling only specific Twitter data. We use twitter4j **??** java library to call Twitter API and search for the relevant tweets based on user's keywords. Using the same interface, a user can generate three different kinds of social network datasets:

- a dataset of key terms and the relationship between them based on their co-occurrence in the same tweets
- a dataset of tweeters (persons who posted the tweets) and relationships between them based on common terms in their tweets.
- a dataset for two-mode network between tweeters and key terms to show terms written by each tweeter.

Figure 2 shows the output of the network of the tweeter network based on crawling politician names (e.g., Trump, Putin, Obama, Macron). We can see that the network forms groups of four heavy linked clusters based on each politician name where the nodes share an edge if they mention the same keyword (politician). The edges between the different clusters mean that a user has tweeted both keywords in the same tweet. After importing the network, and viewing it as shown in Figure 2, the user can then proceed to apply the existing social network analysis measures provided by NetDriller.

2) Scholar Data: The new version of NetDriller provides the functionality to crawl scholar data based on a scholar name or specific keywords. Figure 3 shows the GUI used to generate scholar data. The user can enter the keywords and then select the source to crawl the scholar data from (DBLP and IEEE).

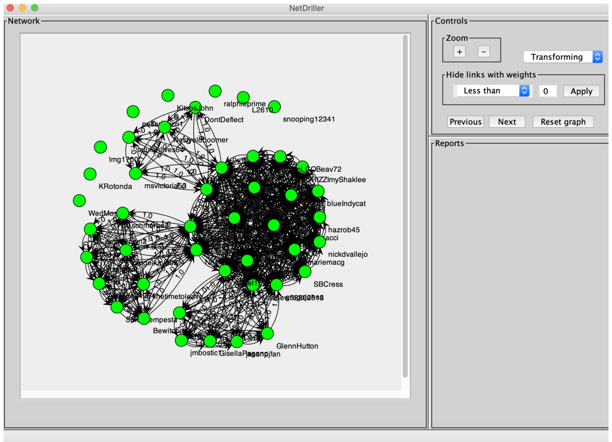

Fig. 2: Twitter data collection social network output

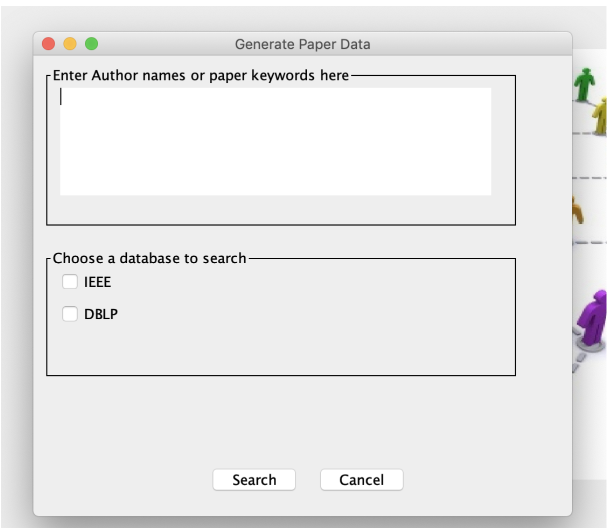

Fig. 3: Scholar GUI data collection

After the user crawls the scholar data, they can then proceed to generate three different types of social networks based on their preferences:

- a dataset of authors with relationship based on the number of coauthored papers.
- a dataset of keywords from titles of papers and relationship between keywords is based on their co-occurrence in the same title.
- a dataset for weighted two-mode network between authors and keywords.

3) Web: We designed a web crawler which works as follows. After importing a labeled network into Netdriller and ranking its nodes using the centrality measures, nodes from the network can be selected to be used in web crawling. Every selected node starts a new search tree which is processed by the web crawl function. The crawl function takes some root node and a depth value. When a new Crawler object is instantiated, one should provide a maximum value for the search depth, which is stored in the max depth attribute. This value corresponds to the maximum height of the search tree rooted at the provided node (given by the string start). Thus, given some node, crawl function will re-curse for at most max depth layers. On each layer of the search, at most three new nodes are added, and each of these nodes is connected with the provided root node. For instance, suppose the root node is Alice, and max depth = 2. Then the system will find three names most closely-related to Alice and add them to the network as shown in Figure 4. Assume these names are Bob, Clarence, and Deborah. Then Bob, Clarence, and Deborah will be added to the network graph, and will be connected to Alice with weights given by their frequency values found in documents related to Alice. Then, the crawler will be recursively called with roots Bob, Clarence, and Deborah, and the three most closely-related names to each of these nodes will be added to the graph and connected to their respective roots (for a total of nine new nodes for this layer). The search will then terminate since the maximum depth value has been reached (with respect to the original root node, Alice). This search tree is shown without edge weights below.

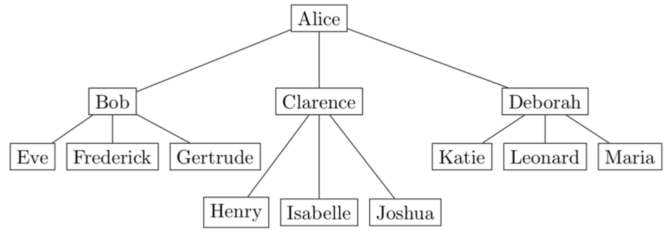

Fig. 4: Web Crawler

We collect the required information from the web using Google's CustomSearch API [2]. The term provided (root) is passed to this API, and links are extracted from the returned structure. Each link is processed and entity recognition is applied to extract names from the text. Whenever a name is seen, its frequency is updated and it is reflected on the weight of the edge.

B. Metric Visualization

After importing/constructing a social network from the user data, the network is shown in the main panel of the tool in terms of the nodes, labels, edges and corresponding weights. Then, users can choose to apply SNA techniques on the network to identify the most important nodes using the "Metrics" menu, including Betweenness, Closeness, Degree, Eigenvector centrality, etc. The results of these centrality measures are shown on the right panel. In this version, we added a bar chart as a different view for the users to show their results as shown in Figure 5 which helps in viewing the distribution of the results.

C. Node Property

Previously, networks were added to NetDriller either labelled or unlabelled. Further, node general properties such as

[2]CustomSearch API: https://developers.google.com/custom-search/v1/overview

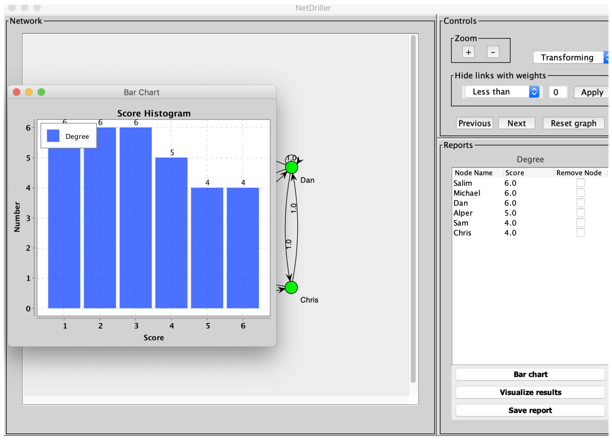

Fig. 5: SNA metric analysis

description, or node specific properties such as age, height, etc. were not provided in the previous versions of NetDriller. In this version, we added the functionality to import node properties file related to the imported network so that users can view node properties by clicking on nodes or even search for nodes by their properties.

D. Locate Nodes

When using large labelled networks, finding nodes visually can be really difficult, especially when the network is dense. Therefore, we implemented the functionality to locate nodes in this version of NetDriller. It is possible to highlight nodes in the network based on node name or description. Figure 6 shows the menu used for users to enter their search criteria, while Figure 7 shows nodes which match the specified search criteria. If the user clicks on the "View" button, they will be redirected to the social graph where the clicked node is highlighted.

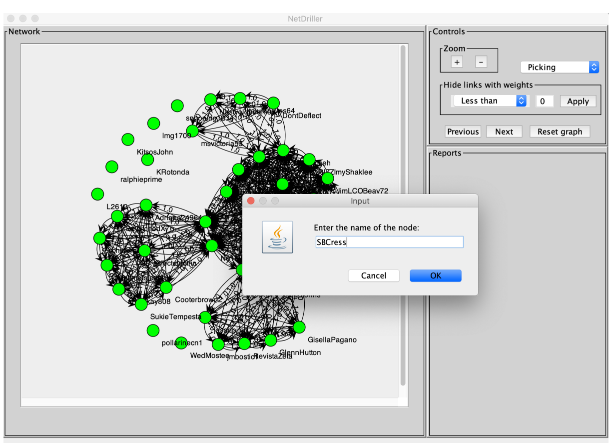

Fig. 6: Search for nodes using node name or description

E. Graph Interaction

One of the major changes we did in this version, is to make NetDriller networks dynamic and more interactive with the user instead of just showing a network on the main panel

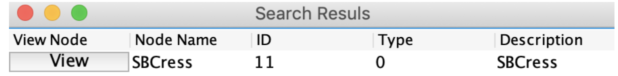

Fig. 7: Node Located in network based on user search

with different layout views. To achieve this, we implemented several functions to add/edit nodes and edges in the network and to facilitate the ability to move and connect nodes around. The functions added are:

- Make nodes in the graph interactive so that users can move the nodes around with the network getting adjusted depending on how the user moves the nodes.
- Highlight nodes when clicked and show the description of the node in a box.
- Provide a function to hide node names. Users can then click on a node to view its name.
- Rename/Delete/Insert nodes by right clicking on the nodes in the graph.
- Make a save functionality so that after a user edits a graph using the previous methods, he/she can save the graph to a csv file that can be later imported as a NetDriller network.

Figure 8 shows the menu that pops up when the user right clicks on the main panel where the network resides. A user can add a new node or hide label names. Figure 9 shows the menu when the user clicks on a node where they can rename/delete/add edges to the selected node. Also functionality based on clicking on the edge is provided to delete or modify the weight of an edge.

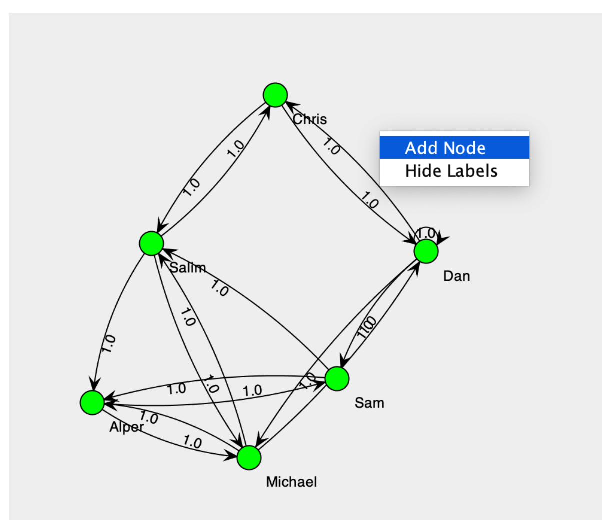

Fig. 8: Add new nodes to the network

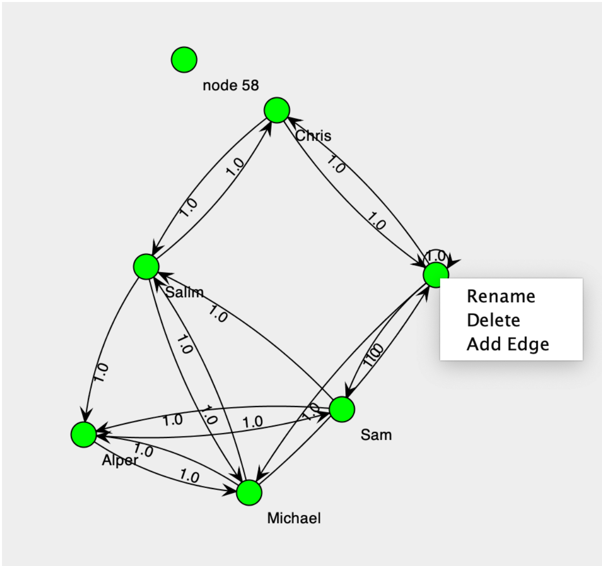

Fig. 9: Manipulate existing nodes in the social network

IV. Conclusion

We presented in this paper the third version of NetDriller for social network construction and analysis. NetDriller is a powerful tool to apply analysis on social networks composed with views of social relationships consisting of nodes and links. We added several important social network construction and analysis techniques to NetDriller to work with current up-to-date datasets. Users now can populate their own dataset using NetDriller from Twitter or from scholar databases such as DBLP and IEEE. With the upcoming rise of social media, it is essential for social network analysis tools to populate and generate different networks from social media. We do this using one and two mode networks from terms and tweeters. We also added new visualizing ability for the network metrics for analysis purposes. Relating nodes with specific attributes for the network has also been added with the ability to locate nodes by their properties. This is useful for large networks. Lastly, we introduced several graph interaction mechanisms to manipulate the social graph and make it interactive.

References

[1] S. Afra, A. Aksaç, T. Õzyer, and R. Alhajj, "Link prediction by network analysis," in *Prediction and Inference from Social Networks and Social Media*. Lecture Notes on Social Networks Series, Springer, 2017, pp. 97–114.

[2] S. Afra, T. Özyer, and J. Rokne, "Netdriller version 2: A powerful social network analysis tool," in *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2018, pp. 1475–1480.

[3] M. Bastian, S. Heymann, M. Jacomy *et al.*, "Gephi: an open source software for exploring and manipulating networks." *Icwsm*, vol. 8, pp. 361–362, 2009.

[4] V. Batagelj, and A. Mrvar, "Pajek-program for large network analysis", Connections, Vol. 21, No. 2, pp.47–57, 1998.

[5] S. P. Borgatti, M. G. Everett, and L. C. Freeman, "Ucinet for windows: Software for social network analysis," 2002.

[6] D. Camacho, A. Panizo-Lledot, G. Bello-Orgaz, A. Gonzalez-Pardo, and E. Cambria, "The four dimensions of social network analysis: an overview of research methods, applications, and software tools", Information Fusion, November, Vol. 63, pp.88–120, 2020.

[7] J. Chen, Y. Tang, J. Li, C. Mao, and J. Xiao, "Community-based scholar recommendation modeling in academic social network sites," in *Proceedings of the International Conference on Web Information Systems Engineering*. Springer, 2013, pp. 325–334.

[8] G. Csardi, T. Nepusz, et al. "The igraph software package for complex network research", InterJournal, Complex Systems, Vol. 1695, No. 5, pp.1–9, 2006.

[9] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10–18, 2009.

[10] M.S., Handcock, D.R., Hunter, C.T., Butts, S.M. Goodreau, and M. Morris, "Statnet: software tools for the representation, visualization, analysis and simulation of network data", Journal of Statistical Software, Vol. 24, No. 1, p.1548, 2008.

[11] N. Koochakzadeh, A. Sarraf, K. Kianmehr, J. Rokne, and R. Alhajj, "Netdriller: A powerful social network analysis tool," in *Proceedings of the IEEE 11th International Conference on Data Mining, Workshops (ICDMW)*. IEEE, 2011, pp. 1235–1238.

[12] D. Michail, J. Kinable, B. Naveh, and J.V. Sichi, "JGraphTa Java library for graph data structures and algorithms", ACM Transactions on Mathematical Software, Vol. 46, No. 2, pp.1–29, 2020.

[13] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, "Cytoscape: A software environment for integrated models of biomolecular interaction networks," *Genome Research*, vol. 13, pp. 2498–2504, 2009.

[14] M. Smith, N. Milic-Frayling, B. Shneiderman, E. Mendes Rodrigues, J. Leskovec, and C. Dunne, "Nodexl: a free and open network overview, discovery and exploration add-in for excel 2007/2010," 2010.

[15] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe, "Predicting elections with twitter: What 140 characters reveal about political sentiment," in *Fourth international AAAI conference on weblogs and social media*, 2010.

[16] P. Wang, B. Xu, Y. Wu, and X. Zhou, "Link prediction in social networks: the state-of-the-art," *Science China Information Sciences*, vol. 58, no. 1, pp. 1–38, 2015.

[17] S. Wasserman and K. Faust, *Social network analysis: Methods and applications*. Cambridge university press, 1994, vol. 8.