Information Modelling and Knowledge Bases XXXIV
M. Tropmann-Frick et al. (Eds.)

doi:10.3233/FAIA220500

Analysis of the Shortest Path Method Application in Social Networks

Boštjan ŠUMAK[a,1] and Maja PUŠNIK[a]
[a] *University of Maribor, Faculty of Electrical Engineering and Computer Science*

Abstract. This paper analyzes the shortest path problem (SPP) in social networks, based on the investigation and implementation of different methods on a simulated example. The objectives of the paper include identification of the most commonly used methods for finding the shortest path in a social network as a strategic attempt to speed the search of network nodes, focusing on the application of the two most used SPP methods: the Dijkstra and Bellman-Ford algorithms. A comparative analysis is used as an investigation method for performance evaluation of different algorithms, based on their implementation and behavior, tested on a social network example. The research results indicate that the Dijkstra algorithm is faster, and therefore more suitable for searching the shortest connection in social networks.

Keywords. Social networks, Social network analysis, Shortest path problem, Dijkstra algorithm, Bellman-Ford algorithm, Optimization of search

1. Introduction

Social networks have become an indispensable concept in today's digital world, where devices and people connect online and the number of social network users is still rising. In 2010, there were roughly 1 billion users, rising to about three and half a billion in 2020. Although in some cases the users are virtual entities (users with multiple user accounts or generated user accounts managed by computers), it is estimated that one-third of the world's population is included into some kind of social networks [20], such as Facebook, YouTube, Instagram, Twitter, and others. All social network users are linked in some way to a shared site, categorized by interest groups such as the same geographic area (for example, VKontakte, intended primarily for Russian-speaking users, or QZone for Chinese-speaking users), user groups from the same professional field (example LinkedIn), user groups with similar hobbies or beliefs [22], and other groups. The importance of social networks is also visible in business environments [14], which, for example, also differ whether the user is a male or a female [24]. The listed social networks types present a vast and diverse problem field of the new-age requirement to link smaller, medium-sized and large numbers of users with similar interests, and the quick discovery of connections between them is our main research motivation. This paper addresses the social networks and their users, however, we focus on the optimal (shortest) path selection, highlighting the shortest path problem (SPP) methods [32] [37].

Since SPP methods solve the problem of finding the shortest path from point A to point B, using graphs is the most common way to illustrate the problem. They are used

[1] Corresponding Author.

in various fields, such as mapping (Google Maps), road connections, logistics processes, computer networks [3][7], in addition to the field of Social Media for social network analysis (SNA). SNA deals with the study of communication between people and groups of people in a business or a social network. Social networking is represented in mathematical theory as a graph, where graph nodes represent users, and communication between them is represented by connections. Knowing the shortest path between users on the network enables identification of (unknown) indirect connections. SNA includes metrics that use SPP methods to calculate the shortest path between people, giving importance to the individuals within the network, often exploited by social media providers for the recommendation system (RS). The RS is a technology that is used in different environments, suggesting the users (customers, visitors, readers, application users) an item (product, movie, event, and article) that might be of interest to them, also collecting information about users and their past searches. Users` interests are determined with the help of different algorithms A typical example of using an RS are advertisements on web pages that are displayed to a user, based on the user`s past searches, increasing the possibility that the user will become a potential buyer of the product, a visitor to the show, an article reader, etc. In social networks, the RS works by suggesting users who might share similar interests, enabling connection of previously unconnected users [23] [31], an important feature in today's global markets, which focuses increasingly on customer needs. Based on the existing body of research, SPP methods give reliable results on listed networks. A literature review is presented in the following sub-section.

1.1. Background/ Related literature

The larger body of research includes investigation of SPP in addition to social networks in various other areas, also focusing on travel time in the transportation network. The authors in [36] proposed a data-driven distributional robust shortest path model for finding optimal paths to minimize the mean-excess travel time. The authors in [17] addressed the multi-criteria shortest path problem and presented the interactive method of analyzing SPP by the reference point approach, pointing out the routing of hazardous materials as an example of an SPP application. The authors in [13] addressed Bi-objective SPP, finding (one-to-one) paths from a start node to an end node, while simultaneously minimizing two (conflicting) objective functions for large-scale road networks. There are also health-related areas of application, such as for predicting adolescent social networks to stop smoking in secondary schools with the authors in [16] researching the potential importance of adolescent friendships` selection, affecting the health domain. SPP is also applied in other problem-solving efforts, presented in [34], where the authors deal with a competitive Influence Maximization Problem, where two players make decisions sequentially; the first player (leader) wants to maximize the spread by activating an influential seed set, and the second player (follower) tries to minimize it by deactivating some of the activated nodes. Similar research, conducted by [27], approaches analytically to the problem of influence maximization in a social network, where two players compete by utilizing dynamic targeting strategies. A large-scale group decision-making model based on social network analysis, for example, is analyzed in [26]. The importance of network structure in social and economic systems is presented in the Evolution of social networks by [19] and a literature review was used to examine the application of the SPP methods in the Social Networks field, such as

Facebook, Twitter, and YouTube, which are used daily by a large number of users, often mixing the terms social media and social network.

For clarification, in the context of the World Wide Web, social media is a website and mobile application that allow users to communicate with other users in a digital way, thus, as a means of communication. Social media is a broader term, and covers many different media, such as images, videos, content, blogs, and others. In contrast, the term social network denotes the social structure of users who are united by common interests. A social network can exist between organizations, and can be set up on any social media, as well as in the real world, with the main purpose to connect users. Social media, therefore, allows users to establish communication links, and social networks take care of strengthening those links. As a result, social networking is a sub-category of social media [9] [29]. The importance of the two concepts is visible in the number of users of social media as well as networks, which is, as stated before, still growing, and according to [6] [28], reaching around 2 and a half billion users.

The massive number of social media users means that social media and related activities have become a lucrative business opportunity. Social media companies have become global corporations, and are among the most valuable companies with the highest annual income. Due to online advertising, the companies make huge profits, despite that the users register and use social media free of charge, providing their data. The data collected through SNA are used further for recommendations, advertising and targeted display of relevant products to the right groups of people on social media. [18] [15]. Ads are tailored to the individual user based on the RS, often controversial in terms of personal data protection. For example, companies like Google, Facebook, Twitter and others store personal information about their users (profiling) and provide them with tailored ads, including personalized content and suggesting other potentially interesting users, transforming the user into a product itself. By addressing customers who have a proven record of interest in a particular product or service, increases the percentage of advertising success significantly. Indisputably, the companies benefit from social media, despite some potentially negative effects like rapid dissatisfaction information spread [21].

The listed benefits could also be beneficial in smaller business companies and organizations, gaining a strategic advantage if users (employees) would be better connected, and their interests and needs would be understood better. Investigation of possible solutions, how to identify connected users and find the shortest path between them, is addressed in the following sections. The remainder of the paper is organized as follows. Section 2 reviews the methods and algorithms, and presents a comparison of the two selected algorithms and a simulated example of the chosen algorithms. Section 3 provides the results, followed by a discussion in section 4. The acknowledgements and bibliography follow.

2. Methods

The paper's main motivation is using the SPP method in social networks, and identifying which node is the most connected to other nodes and which path between them is the shortest. A search for the shortest pathfinding methods was conducted, used in the existing field of Social Network Analysis. The most commonly used methods are the DIJKSTRA ALGORITHM, the BELLMAN FORD ALGORITHM, the FLOYD-

WARSHALL ALGORITHM, the JOHNSON ALGORITHM, the A* ALGORITHM and some others:

The Dijkstra algorithm uses the breadth search method to solve the shortest single-point path problem. It does not support negative weights at the nodes, but it performs fast. The Dijkstra algorithm (DA) has several variants, and in some cases can also be used to find the shortest path for all pairs. The Dijkstra algorithm works by creating a tree of shortest paths from the starting node to all other nodes [7] [4].

The Bellman-Ford algorithm (BFA) is used to find the shortest path from one starting node to all the other nodes in the directed and weighted graph, with the links having negative weights. If the graph is not oriented, it must be transformed into directed in such a way that two directed links are created from each non-directed link. It is based on the concept that the shortest path has at most N - 1 connections, since the shortest path between two nodes never has a cycle [3] [8].

The Floyd-Warshall algorithm solves the problem of finding the shortest path for all pairs of nodes in the graph, using a dynamic programming technique. Node connections can have positive or negative weights. The Floyd-Warshall algorithm breaks down the problem into several smaller parts, and combines the results of the smaller parts to solve the big problem. It operates according to the principle of the following statement: For each of the three nodes in the graph, if there is a shortest path A to C that is longer than the sum of the shortest path A to B and B to C, then the shortest path is A to C equal to this sum. The algorithm works well on graphs with many links and cycle graphs. [5] [8].

The Johnson algorithm is used to find the shortest path for all pairs of nodes in the graph, where the connections can be negative or positive. It works by first executing the BFA, and using the results to re-weight the graph by eliminating any negative weights. Then, the DA for each graph node is executed, to obtain the shortest path for all pairs of nodes. The implication is that, for each node the DA runs, because the Johnson algorithm works well for graphs with few links, where it has a shorter execution time compared to the Floyd-Warshall algorithm. [7].

The A * algorithm is an extension of the DA with some features of the breadth method. The A * algorithm creates a tree of shortest paths from the starting node to all other nodes. It differs from the DA in that it uses for each node a function that gives an estimate of the total cost (a cost is a weight) of the path through that node. The function, therefore, consists of the instantaneous cost of reaching the node and of predicting the cost from that node to the destination node. Because of this, the A * algorithm is a heuristic method, which means that it produces results that are the result of predictions and are not necessarily correct, but, consequently, the A * algorithm is faster than the DA [4].

According to the literature review, the DA is the most famous and fastest method, [33] and also the most commonly used in the analysis of social networks. Since the two most common identified methods were the BFA and the DA, they will be explored in more detail within this paper. Both algorithms base on the principle of finding the shortest path from the starting node to all other nodes in the graph, with a significant difference: The DA does not allow negative weights on the links between nodes, unlike the BFA.

2.1. Algorithm comparison

Based on the literature review, the two most common identified shortest path methods are DA and BFA; therefore, a closer comparison was made with the help of an experiment. Within the experiment, the speed of operation of the DA and BFA was tested in practice, on graphs. Both represent a method for finding the shortest path in weighted graphs, and both are able to find the shortest path from the selected node to all other nodes in the graph easily. SPP in social networks is connected closely to the graph theory, where the problem of finding the shortest path is known as finding the shortest connection between two points on a graph. There are several types of graphs. The first feature that determines the type of graph is the orientation of its links. Connections between nodes can be one-way or two-way. If the links are one-way, the graph is called a directed graph, however, if the links are bidirectional (in both directions), the graph is called a non-directional graph. If the graph's links are one-way and two-way at the same time, each two-way link is replaced by two one-way links (each in one direction), which is equivalent to a two-way link, and such a graph is then fully oriented [7]. The second feature of the graphs is related to link weights. A graph that has no weights on the links is called a non-weighted graph; otherwise, when there are weights on the links, it is called a weighted graph. The values of the weights can also be negative, which is not supported by all the shortest path algorithms. The third feature of graphs that influences the choice of the shortest path search algorithm is cyclicality. A cycle is defined as any path through a graph that visits the same node more than once. So if there is any cycle path in the graph, the graph is cyclical, otherwise, it is acyclic. So, there are the following 6 types of graphs, with two in a pair mutually exclusive: non-directional / directional graph, weighted/unweighted graph, cyclic/acyclic graph.

The DA works according to the principle of breadth search, which means that first all the output connections of the current node are released, then we take the nearest neighboring node and release all its output connections, and continue until all the nodes are processed. With the BFA, on the other hand, all output connections of all nodes are identified and released, repeated $| V - 1 |$ times, where V is the number of nodes of the graph. In the BFA, each connection is loosened several times (for a graph with 3 or more nodes), while it is always used only once by the DA, which is reflected in the greater time complexity of the BFA for graphs with many nodes. In contrast, increasing the number of connections at the same number of nodes reduces the running time of the BFA, which can be attributed primarily to the use of a data structure as a priority type for link implementation. Based on this, the implementation of the BFA, using the priority connection type, works most optimally on graphs with many connections. Alternatively, the implementation of the DA, using the "folder" data structure for links, has the effect of increasing the execution speed linearly by increasing the number of links. The BFA, therefore, performs iterations over all nodes and connections, so its time complexity is measured as the number of nodes multiplied by the number of connections. Instead, the time complexity of the DA depends on the variant of the algorithm. There are some specialized variants of the algorithm that are optimized for specific graph shapes. A general variant of the DA, in which graph nodes are stored in the form of a list of adjacent nodes with a data structure folder or priority type, in the worst case, has a time complexity $O = | E | + | V | \log | V |$, where V is the number of nodes and E is the number of connections. Table 1 shows a comparative analysis of the DA and BFA based on several properties (Implementing nodes, Implementing connections, Type of algorithm, The use of heuristics, Search result, Speed, Weighted graph, Negative weights, Optimal performance, Time complexity, Implementation phases, Modus operandi), revealing several joint characteristics, however, also several differences.

Although different, both algorithms are appropriate for finding the shortest path within a network. According to the comparison results (Table 1) the DA is faster than the BFA, the reason being the simpler implementation. The DA does not allow the entry of negative link weights, which means that there is no possibility of a negative cycle. As a result, the implementation of the DA can omit the phase of verifying the existence of a negative cycle and is, therefore, faster. Application of them on a simulated network is presented in the following sub-section

Table 1. Comparative analysis of the Dijkstra and Bellman-Ford algorithms

Property	Dijkstra algorithm	Bellman-Ford algorithm
Implementing nodes	The set	The list
Implementing connections	Folder	Priority species
Type of algorithm	One starting point	One starting point
The use of heuristics	No	No
Search result	Precise	Precise
Speed	Quick	Less fast
Weighted graph	Enables	Enables
Negative weights	It does not	Enables
Optimal performance	Graph with few links	Graph with many links
Time complexity	O = |E| + |V|log|V|	O = |V| * |E|
Implementation phases	2 phases: • initialization, • evaluation.	3 phases: • initialization, • evaluation, • checking.
Modus operandi	Releasing the links into the width	Releasing all connections for |V – 1| times

2.2. Social network metrics and test simulation

Social network analysis provides both mathematical and visual analysis of human relationships. In the field of Social Network Analysis, centrality is a concept used to determine the relative importance of a node in a given network. Centrality, therefore, addresses the question of who is the most important or central person in the network. The following metrics are used most commonly to measure centrality [35] [1] [2] [12]:

- Degree Centrality - The simplest metric for centrality is the degree of centrality. This metric calculates the number of connections of each node in a given network. The more connections the node has, the greater the degree of centrality. This metric is crucial for identifying important nodes, as it identifies nodes that hold a lot of information quickly. Nodes receive and transmit information, so the input (only for inbound connections) and output (only for outbound connections) node centrality can be calculated separately, which is useful in transaction analysis. [1] [2]
- Betweenness Centrality - Intermediate centrality is a metric that identifies nodes that often occur on the shortest path of communication between two nodes. This means that a node has a high value of intermediate centrality if it often acts as a facilitator in exchanging a large amount of information. It is measured by how many times the node lies on the shortest path among the other nodes. In social networks, nodes with high intermediate centrality are often on the outskirts of two densely populated groups. [1] [2] [12].
- Closeness Centrality - Proximity centrality is a metric for centrality that indicates the distance of a node to all other nodes. This value is obtained by calculating the reach of a node in the network. The node with the highest

centrality proximity value reaches all nodes in the network faster than any other node. Therefore, it has the shortest path to all other nodes. Nodes with high proximity to centrality are well connected throughout the network and have a good overview of what's happening on the network. Removing such a node would cause communication problems. [35] [1].

Various tools for analyzing social networks online simplify retrieving data from social networks, creating social network graphs from given data, and calculating metrics. The purpose of all these tools is to simplify the processing of social network data, since the user does not need to implement algorithms for finding the shortest path. Some of the tools used to analyze social networks are Gephi [25], Cuttlefish [11], Spider [11], Cytoscape [30], NodeXL [25], Social Network Visualizer [25] and NetMiner [25] [11]. Based on analysis of the listed tools, NodeXL software tool was used, with which we obtained test data from social media and conducted an action survey/social network simulation. NodeXL is a template that can be used with Microsoft Excel to capture data from various files and web sources. The tool was chosen since it allows capturing data directly from social media (Twitter) (Figure 2). By capturing tweets from Twitter, we obtained test data to simulate social networking. As shown in Figure 1, we obtained tweets containing the desired keyword "Slovenia" and limited the number of tweets to 2000.

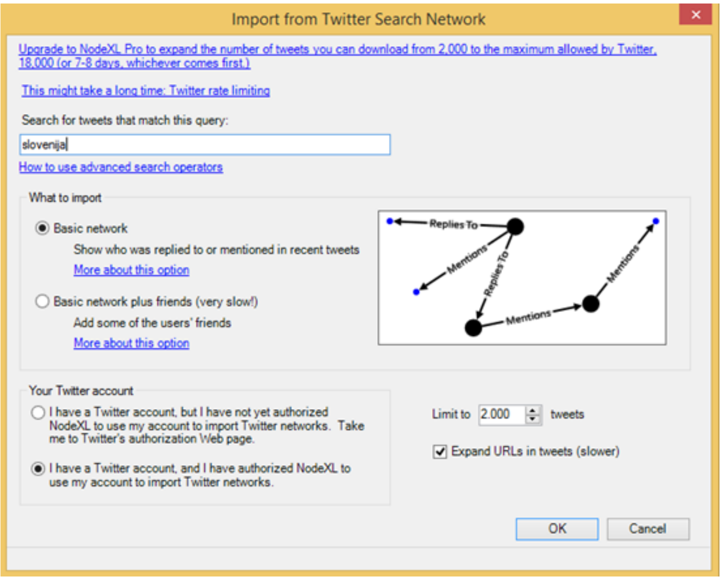
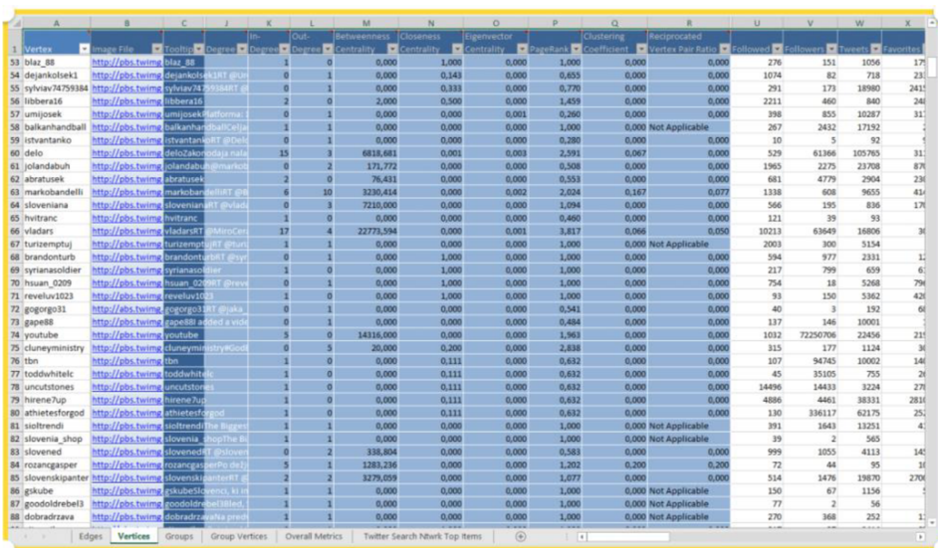


Figure 1. NodeXL for Twitter network import.

A Node Table (Figure 2) and a Link Table (Figure 3) were created in Excel. Both Tables have been transformed into the preferred format. Next, we performed a link analysis that showed that there were duplicates between the links. We combined the duplicate links with the tool by increasing the link weights by one for each duplicate link (the last column in Figure 3 shows the link weights). This resulted in a final number of nodes of 848 and a final number of connections of 2156. The metrics for social network analysis were calculated based on the data. The tool inserted a column showing the metrics for the social network analysis (the light blue columns in Figure 2). The following metrics were calculated: entry-level of centrality, exit level of centrality, intermediate centrality, proximity to centrality, own centrality, page rank, clustering coefficient and reciprocated vertex pair ratio.

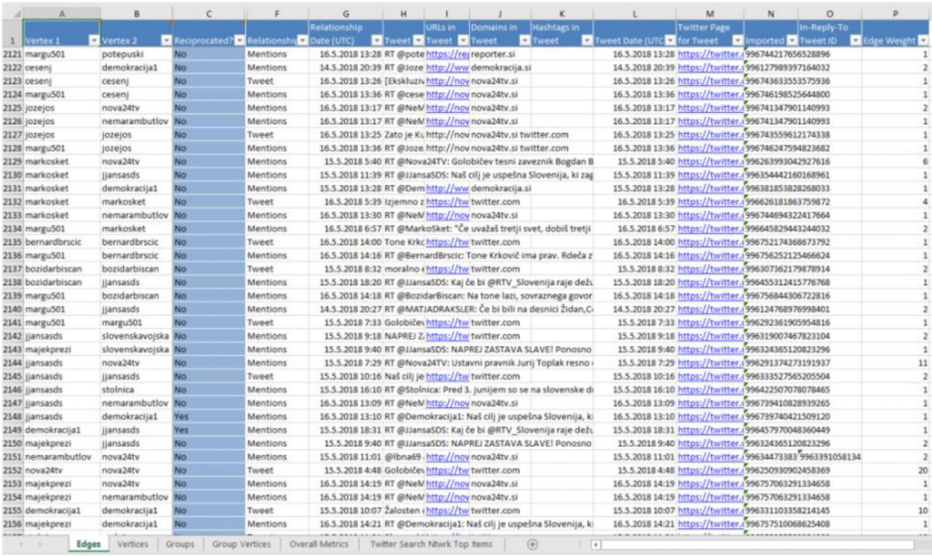
The Table of nodes, therefore, contains a list of all Twitter social media user profiles that are in any way related to posts that contain the selected keyword. The node name in the Table is the same as the social media user profile name. On the other hand, the Link Table contains a column (the fourth column in Figure 3) that describes the type of link. For social media posts, the tool separates the following types of links: “Tweet”, “Mentions”, and “Replies to”. A "Twit" link is created when a user posts a tweet to a social media profile, and the name of the user profile is entered as the start and end nodes of the link. These are, therefore, links that have a beginning and an end in the same node which we call self-loops. A "Mention" type link occurs when a user mentions another user's profile or tweet, with the start node being the name of the user that mentions and the end node the name of the user whose profile or tweet is mentioned. A "Reply to" link occurs when a user replies to another user's tweet, with the start node being the name of the responding user, and the end node the name of the user who posted the tweet. A connection to the same start and end node can occur several times, so the weight of the connection represents the number of occurrences of such connections.



	Vertex	Image File	Tooltip	Degree	In-Degree	Out-Degree	Betweenness Centrality	Closeness Centrality	Eigenvector Centrality	PageRank	Clustering Coefficient	Reciprocated Vertex Pair Ratio	Followed	Followers	Tweets	Favorites
53	blaz_88	http://pbs.twimg	blaz_88		1	0	0,000	1,000	0,000	1,000	0,000	0,000	276	151	1056	17
54	dejankolsek1	http://pbs.twimg	dejankolsek1RT @Ur		0	1	0,000	0,143	0,000	0,655	0,000	0,000	1074	82	718	23
55	sylviav74759384	http://pbs.twimg	sylviav74759384RT @		0	1	0,000	0,333	0,000	0,770	0,000	0,000	291	173	18980	241
56	libbera16	http://pbs.twimg	libbera16		2	0	2,000	0,500	0,000	1,459	0,000	0,000	2211	460	840	24
57	umijosek	http://pbs.twimg	umijosekPlatforma:		0	1	0,000	0,000	0,001	0,260	0,000	0,000	398	855	10287	31
58	balkanhandball	http://pbs.twimg	balkanhandballCelja		1	1	0,000	0,000	0,000	1,000	0,000	Not Applicable	267	2432	17192	
59	istvantanko	http://pbs.twimg	istvantankoRT @Del		0	1	0,000	0,000	0,000	0,280	0,000	0,000	10	5	92	
60	delo	http://pbs.twimg	deloZakonodaja nala		15	3	6818,681	0,001	0,003	2,591	0,067	0,000	529	61366	105765	31
61	jolandabuh	http://pbs.twimg	jolandabuh@markob		0	2	171,772	0,000	0,000	0,508	0,000	0,000	1965	2275	23708	87
62	abratusek	http://pbs.twimg	abratusek		2	0	76,431	0,000	0,000	0,553	0,000	0,000	681	4779	2904	23
63	markobandelli	http://pbs.twimg	markobandelliRT @B		6	10	3230,414	0,000	0,002	2,024	0,167	0,077	1338	608	9655	41
64	sloveniana	http://pbs.twimg	slovenianaRT @vlad		0	3	7210,000	0,000	0,000	1,094	0,000	0,000	566	195	836	17
65	hvitranc	http://pbs.twimg	hvitranc		1	0	0,000	0,000	0,000	0,460	0,000	0,000	121	39	93	
66	vladars	http://pbs.twimg	vladarsRT @MiroCer		17	4	22773,594	0,000	0,001	3,817	0,066	0,050	10213	63649	16806	3
67	turizemptuj	http://pbs.twimg	turizemptujRT @turi		1	1	0,000	0,000	0,000	1,000	0,000	Not Applicable	2003	300	5154	
68	brandonturb	http://pbs.twimg	brandonturbRT @syr		0	1	0,000	1,000	0,000	1,000	0,000	0,000	594	977	2331	1
69	syrianasoldier	http://pbs.twimg	syrianasoldier		1	0	0,000	1,000	0,000	1,000	0,000	0,000	217	799	659	6
70	hsuan_0209	http://pbs.twimg	hsuan_0209RT @reve		0	1	0,000	1,000	0,000	1,000	0,000	0,000	754	18	5268	79
71	reveluv1023	http://pbs.twimg	reveluv1023		1	0	0,000	1,000	0,000	1,000	0,000	0,000	93	150	5362	42
72	gogorgo31	http://abs.twimg	gogorgo31RT @jaka_		0	1	0,000	0,000	0,000	0,541	0,000	0,000	40	3	192	6
73	gape88	http://pbs.twimg	gape88I added a vide		0	1	0,000	0,000	0,000	0,484	0,000	0,000	137	146	10001	
74	youtube	http://pbs.twimg	youtube		5	0	14316,000	0,000	0,000	1,963	0,000	0,000	1032	72250706	22456	21
75	cluneyministry	http://pbs.twimg	cluneyministry#God		0	5	20,000	0,200	0,000	2,838	0,000	0,000	315	177	1124	3
76	tbn	http://pbs.twimg	tbn		1	0	0,000	0,111	0,000	0,632	0,000	0,000	107	94745	10002	14
77	toddwhitelc	http://pbs.twimg	toddwhitelc		1	0	0,000	0,111	0,000	0,632	0,000	0,000	45	35105	755	2
78	uncutstones	http://pbs.twimg	uncutstones		1	0	0,000	0,111	0,000	0,632	0,000	0,000	14496	14433	3224	27
79	hirene7up	http://pbs.twimg	hirene7up		1	0	0,000	0,111	0,000	0,632	0,000	0,000	4886	4461	38331	281
80	athietesforgod	http://pbs.twimg	athietesforgod		1	0	0,000	0,111	0,000	0,632	0,000	0,000	130	336117	62175	25
81	sioltrendi	http://pbs.twimg	sioltrendiThe Bigges		1	1	0,000	0,000	0,000	1,000	0,000	Not Applicable	391	1643	13251	4
82	slovenia_shop	http://pbs.twimg	slovenia_shopThe Bi		1	1	0,000	0,000	0,000	1,000	0,000	Not Applicable	39	2	565	
83	slovened	http://pbs.twimg	slovenedRT @sloven		0	2	338,804	0,000	0,000	0,583	0,000	0,000	999	1055	4113	14
84	rozancgasper	http://pbs.twimg	rozancgasperPo dež		5	1	1283,236	0,000	0,000	1,202	0,200	0,200	72	44	95	1
85	slovenskipanter	http://pbs.twimg	slovenskipanterRT @		2	2	3279,059	0,000	0,000	1,077	0,000	0,000	514	1476	19870	270
86	gskube	http://pbs.twimg	gskubeSlovenci, ki in		1	1	0,000	0,000	0,000	1,000	0,000	Not Applicable	150	67	1156	
87	goodoldrebel3	http://pbs.twimg	goodoldrebel3Bled,		1	1	0,000	0,000	0,000	1,000	0,000	Not Applicable	77	2	56	
88	dobradrzava	http://pbs.twimg	dobradrzavaNa pred		1	1	0,000	0,000	0,000	1,000	0,000	Not Applicable	270	368	252	1

Figure 2. Table of Nodes retrieved from tweets for the keyword "Slovenia".

In the final phase of the study, the NodeXL tool was used to analyze and plot the entire graph (Figure 4). By analyzing the graph, we obtained information about the type of graph, the number of connections and nodes, the number of loops, and the density of the graph. Also, the NodeXL tool does user and tweet analysis, providing a list of users with the most tweets, mentions and replies, and the top 10 results in each of the following categories: (1) The most frequently mentioned URL in tweets, (2) The most frequently mentioned domain in tweets, (3) The most frequently mentioned hashtag in your tweet, (4) The most frequently mentioned word in tweets, (5) The most frequently mentioned word pair in tweets.



	A	B	C	F	G	H	I	J	K	L	M	N	O	P
1	Vertex 1	Vertex 2	Reciprocated?	Relationshi	Relationship Date (UTC)	Tweet	URLs in Tweet	Domains in Tweet	Hashtags in Tweet	Tweet Date (UTC	Twitter Page for Tweet	Imported	In-Reply-To Tweet ID	Edge Weight
2121	margu501	potepuski	No	Mentions	16.5.2018 13:28	RT @pote	https://re	reporter.si		16.5.2018 13:28	https://twitter.c	996744217656528896		1
2122	cesenj	demokracija1	No	Mentions	14.5.2018 20:39	RT @Joze	http://ww	demokracija.si		14.5.2018 20:39	https://twitter.c	996127989397164032		2
2123	cesenj	cesenj	No	Tweet	16.5.2018 13:26	[Ekskluziv	http://nov	nova24tv.si		16.5.2018 13:26	https://twitter.c	996743633553575936		1
2124	margu501	cesenj	No	Mentions	16.5.2018 13:36	RT @cese	http://nov	nova24tv.si		16.5.2018 13:36	https://twitter.c	996746198525644800		1
2125	jozejos	nova24tv	No	Mentions	16.5.2018 13:17	RT @NeM	http://nov	nova24tv.si		16.5.2018 13:17	https://twitter.c	996741347901140993		2
2126	jozejos	nemarambutlov	No	Mentions	16.5.2018 13:17	RT @NeM	http://nov	nova24tv.si		16.5.2018 13:17	https://twitter.c	996741347901140993		1
2127	jozejos	jozejos	No	Tweet	16.5.2018 13:25	Zato je Ku	http://nov	nova24tv.si twitter.com		16.5.2018 13:25	https://twitter.c	996743559612174338		1
2128	margu501	jozejos	No	Mentions	16.5.2018 13:36	RT @Joze	http://nov	nova24tv.si twitter.com		16.5.2018 13:36	https://twitter.c	996746247594823682		1
2129	markosket	nova24tv	No	Mentions	15.5.2018 5:40	RT @Nova24TV: Golobičev tesni zaveznik Bogdan B				15.5.2018 5:40	https://twitter.c	996263993042927616		6
2130	markosket	jjansasds	No	Mentions	15.5.2018 11:39	RT @JJansaSDS: Naš cilj je uspešna Slovenija, ki zag				15.5.2018 11:39	https://twitter.c	996354442160168961		1
2131	markosket	demokracija1	No	Mentions	15.5.2018 13:28	RT @Dem	http://ww	demokracija.si		15.5.2018 13:28	https://twitter.c	996381853828268033		1
2132	markosket	markosket	No	Tweet	16.5.2018 5:39	Izjemno z	https://tw	twitter.com		16.5.2018 5:39	https://twitter.c	996626181863759872		4
2133	markosket	nemarambutlov	No	Mentions	16.5.2018 13:30	RT @NeM	http://nov	nova24tv.si		16.5.2018 13:30	https://twitter.c	996744694322417664		1
2134	margu501	markosket	No	Mentions	16.5.2018 6:57	RT @MarkoSket: "Če uvažaš tretji svet, dobiš tretji				16.5.2018 6:57	https://twitter.c	996645829443244032		2
2135	bernardbrscic	bernardbrscic	No	Tweet	16.5.2018 14:00	Tone Krkc	https://tw	twitter.com		16.5.2018 14:00	https://twitter.c	996752174368673792		1
2136	margu501	bernardbrscic	No	Mentions	16.5.2018 14:16	RT @BernardBrscic: Tone Krkovič ima prav. Rdeča z				16.5.2018 14:16	https://twitter.c	996756252125466624		1
2137	bozidarbiscan	bozidarbiscan	No	Tweet	15.5.2018 8:32	moralno (	https://tw	twitter.com		15.5.2018 8:32	https://twitter.c	996307362179878914		2
2138	bozidarbiscan	jjansasds	No	Mentions	15.5.2018 18:20	RT @JJansaSDS: Kaj če bi @RTV_Slovenija raje dežu				15.5.2018 18:20	https://twitter.c	996455312415776768		1
2139	margu501	bozidarbiscan	No	Mentions	16.5.2018 14:18	RT @BozidarBiscan: Na tone lazi, sovraznega govor				16.5.2018 14:18	https://twitter.c	996756844306722816		1
2140	margu501	jjansasds	No	Mentions	14.5.2018 20:27	RT @MATJADRAKSLER: Če bi bili na desnici Židan,C				14.5.2018 20:27	https://twitter.c	996124768976998401		2
2141	margu501	margu501	No	Tweet	15.5.2018 7:33	Golobičev	https://tw	twitter.com		15.5.2018 7:33	https://twitter.c	996292361905954816		1
2142	jjansasds	slovenskavojska	No	Mentions	15.5.2018 9:18	NAPREJ Z.	https://tw	twitter.com		15.5.2018 9:18	https://twitter.c	996319007467823104		2
2143	majekprezi	slovenskavojska	No	Mentions	15.5.2018 9:40	RT @JJansaSDS: NAPREJ ZASTAVA SLAVE! Ponosno				15.5.2018 9:40	https://twitter.c	996324365120823296		1
2144	jjansasds	nova24tv	No	Mentions	15.5.2018 7:29	RT @Nova24TV: Ustavni pravnik Jurij Toplak resno				15.5.2018 7:29	https://twitter.c	996291374273191937		11
2145	jjansasds	jjansasds	No	Tweet	15.5.2018 10:16	Naš cilj je	https://tw	twitter.com		15.5.2018 10:16	https://twitter.c	996333527565205504		2
2146	jjansasds	stolnica	No	Mentions	15.5.2018 16:10	RT @Stolnica: Pred 3. junijem so se na slovenske d				15.5.2018 16:10	https://twitter.c	996422507078078465		1
2147	jjansasds	nemarambutlov	No	Mentions	16.5.2018 13:09	RT @NeM	http://nov	nova24tv.si		16.5.2018 13:09	https://twitter.c	996739410828939265		1
2148	jjansasds	demokracija1	Yes	Mentions	16.5.2018 13:10	RT @Demokracija1: Naš cilj je uspešna Slovenija, ki				16.5.2018 13:10	https://twitter.c	996739740421509120		1
2149	demokracija1	jjansasds	Yes	Mentions	15.5.2018 18:31	RT @JJansaSDS: Kaj če bi @RTV_Slovenija raje dežu				15.5.2018 18:31	https://twitter.c	996457970048360449		1
2150	majekprezi	jjansasds	No	Mentions	15.5.2018 9:40	RT @JJansaSDS: NAPREJ ZASTAVA SLAVE! Ponosno				15.5.2018 9:40	https://twitter.c	996324365120823296		2
2151	nemarambutlov	nova24tv	No	Mentions	15.5.2018 11:01	@lbna69	http://nov	nova24tv.si		15.5.2018 11:01	https://twitter.c	99634473383	9963391058134	2
2152	nova24tv	nova24tv	No	Tweet	15.5.2018 4:48	Golobičev	https://tw	twitter.com		15.5.2018 4:48	https://twitter.c	996250930902458369		20
2153	majekprezi	nova24tv	No	Mentions	16.5.2018 14:19	RT @NeM	http://nov	nova24tv.si		16.5.2018 14:19	https://twitter.c	996757063291334658		1
2154	majekprezi	nemarambutlov	No	Mentions	16.5.2018 14:19	RT @NeM	http://nov	nova24tv.si		16.5.2018 14:19	https://twitter.c	996757063291334658		1
2155	demokracija1	demokracija1	No	Tweet	15.5.2018 10:07	Žalosten (	https://tw	twitter.com		15.5.2018 10:07	https://twitter.c	996331103358214145		10
2156	majekprezi	demokracija1	No	Mentions	16.5.2018 14:21	RT @Demokracija1: Naš cilj je uspešna Slovenija, ki				16.5.2018 14:21	https://twitter.c	996757510068625408		1

Edges | Vertices | Groups | Group Vertices | Overall Metrics | Twitter Search Ntwrk Top Items

Figure 3. Table of Links from the keyword »Slovenia.

The next step included a test of the social network simulation provided with the NodeXL tool on the implementation of the Dijkstra and Bellman-Ford algorithms. For this purpose, we exported the data on the nodes and connections in the form of an Excel sheet and imported it into Gephi software. We imported the Connection Table and selected the required columns, the Node Table was created automatically concerning all start and end nodes in the Connection Table. Figure 5 shows a Table of Links ready for export (columns need specific names and order, so they were reformatted accordingly). The Node and Link Tables were then exported in the form of properly structured CSV files that the Java parser could read.

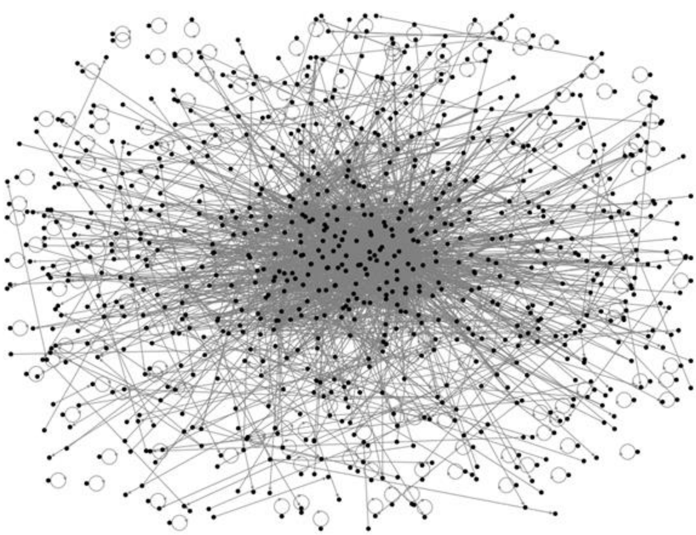
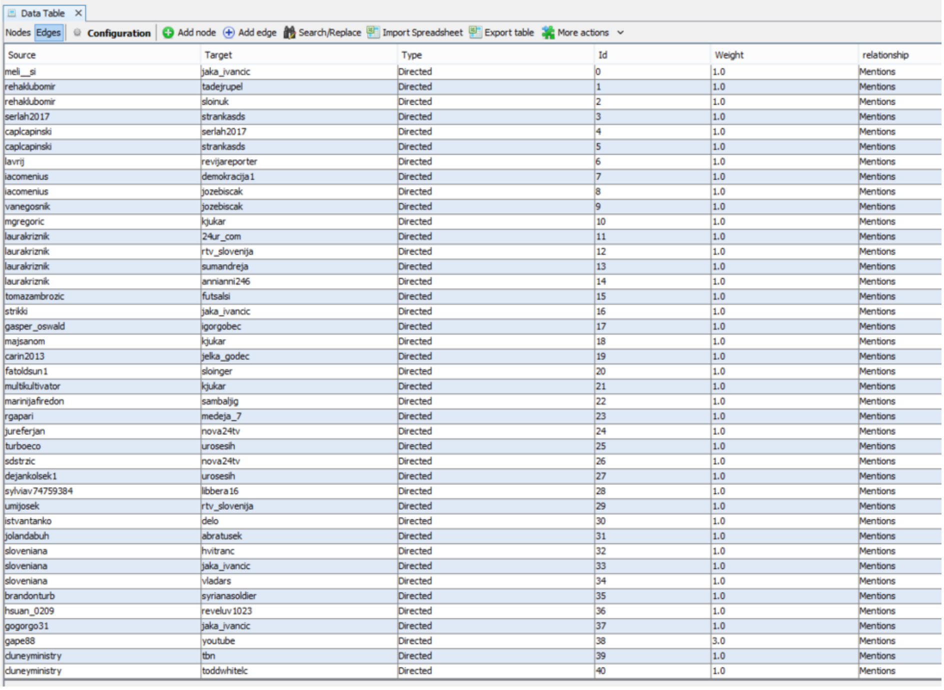


Figure 4. Outline of a social network graph using NodeXL.



Data Table
Nodes Edges Configuration Add node Add edge Search/Replace Import Spreadsheet Export table More actions

Source	Target	Type	Id	Weight	relationship
meli__si	jaka_ivancic	Directed	0	1.0	Mentions
rehaklubomir	tadejrupel	Directed	1	1.0	Mentions
rehaklubomir	sloinuk	Directed	2	1.0	Mentions
serlah2017	strankasds	Directed	3	1.0	Mentions
caplcapinski	serlah2017	Directed	4	1.0	Mentions
caplcapinski	strankasds	Directed	5	1.0	Mentions
lavrij	revijareporter	Directed	6	1.0	Mentions
iacomenius	demokracija1	Directed	7	1.0	Mentions
iacomenius	jozebiscak	Directed	8	1.0	Mentions
vanegosnik	jozebiscak	Directed	9	1.0	Mentions
mgregoric	kjukar	Directed	10	1.0	Mentions
laurakriznik	24ur_com	Directed	11	1.0	Mentions
laurakriznik	rtv_slovenija	Directed	12	1.0	Mentions
laurakriznik	sumandreja	Directed	13	1.0	Mentions
laurakriznik	annianni246	Directed	14	1.0	Mentions
tomazambrozic	futsalsi	Directed	15	1.0	Mentions
strikki	jaka_ivancic	Directed	16	1.0	Mentions
gasper_oswald	igorgobec	Directed	17	1.0	Mentions
majsanom	kjukar	Directed	18	1.0	Mentions
carin2013	jelka_godec	Directed	19	1.0	Mentions
fatoldsun1	sloinger	Directed	20	1.0	Mentions
multikultivator	kjukar	Directed	21	1.0	Mentions
marinijafiredon	sambaljig	Directed	22	1.0	Mentions
rgapari	medeja_7	Directed	23	1.0	Mentions
jureferjan	nova24tv	Directed	24	1.0	Mentions
turboeco	urosesih	Directed	25	1.0	Mentions
sdstrzic	nova24tv	Directed	26	1.0	Mentions
dejankolsek1	urosesih	Directed	27	1.0	Mentions
sylviav74759384	libbera16	Directed	28	1.0	Mentions
umijosek	rtv_slovenija	Directed	29	1.0	Mentions
istvantanko	delo	Directed	30	1.0	Mentions
jolandabuh	abratusek	Directed	31	1.0	Mentions
sloveniana	hvitranc	Directed	32	1.0	Mentions
sloveniana	jaka_ivancic	Directed	33	1.0	Mentions
sloveniana	vladars	Directed	34	1.0	Mentions
brandonturb	syrianasoldier	Directed	35	1.0	Mentions
hsuan_0209	reveluv1023	Directed	36	1.0	Mentions
gogorgo31	jaka_ivancic	Directed	37	1.0	Mentions
gape88	youtube	Directed	38	3.0	Mentions
cluneyministry	tbn	Directed	39	1.0	Mentions
cluneyministry	toddwhitelc	Directed	40	1.0	Mentions

Figure 5. Link Table in Gephi software.

Figure 6 presents the implementation of the CSV file with a Node and Link Table, and converting them into a node structure and a list of links in the data structure. We added both lists to the Graph object to obtain social network data in a format suitable for entering the input parameters of the shortest path search algorithms. The test included performing the Dijkstra and BFA for the first node in the list and measuring the execution time of each algorithm. By performing the test over all the nodes of the graph repeatedly, we received the following results: In 20 trials the JUnit test was performed on all graph nodes and the run times were recorded. The results were analyzed statistically using the SPSS software tool, and it was found that a relatively small social network (848 nodes and 2156 connections) already slowed down the implementation of the BFA. The mean value presents the average time of finding the shortest path among users. The BFA takes about 22 seconds to calculate the shortest path between all graph nodes, while the DA calculates the shortest path for all graph nodes to an average of about 50 ms. The DA is, therefore, a better option for finding the shortest path among users on social networks.

When reviewing the results of the implementation of both algorithms, both the DA implementation and the BFA implementation are not capable of processing the loops that were created in the social network simulation correctly by posting tweets (posting a tweet creates a connection to the beginning and end at the same node). Also, both implementations are not capable of processing multiple equally directed connections correctly between two nodes.

```
@Test
public void testRunSimulation()
{
    CSVParser parser = new CSVParser();

    List<Node> nodes = parser.parseNodesFileWithLabels("./src/resources/slovenijaNodes.csv");

    Graph graph = new Graph(nodes.size());
    for(Node node : nodes)
    {
        graph.addNode(node);
    }

    List<Edge> edges = parser.parseEdgesFileWithLabels("./src/resources/slovenijaEdges.csv", graph);

    for(Edge edge : edges)
    {
        graph.addEdge(edge.getSource(), edge.getDestination(), edge.getWeight());
    }



    long elapsedTime = 0;

    System.out.println("Bellman-Ford algoritem:");
    Date startDate = new Date();
    Algorithm bellmanFord = new BellmanFord();
    for(Node node : nodes)
    {
        bellmanFord.calculateShortestPathFromSource(graph, node);
    }
    Date endDate = new Date();
    elapsedTime = endDate.getTime() - startDate.getTime();
    System.out.println("Bellman-Ford elapsed time: " + elapsedTime );

    System.out.println("Dijkstra algoritem:");
    startDate = new Date();
    Algorithm dijkstra = new Dijkstra();
    for(Node node : nodes)
    {
        dijkstra.calculateShortestPathFromSource(graph, node);
    }
    endDate = new Date();
    elapsedTime = endDate.getTime() - startDate.getTime();
    System.out.println("Dijkstra elapsed time: " + elapsedTime );

}
```

Figure 6 Test of the shortest path search algorithms` performance on social network data.

We have found that, for finding the shortest path, these types of links are unnecessary and misleading. In the case of multiple connections between two points, the shortest path will always follow the link that has the lowest weight, so all other duplicate links can be removed. We also do not need loops to connect to the same node, since, in the case of a positive loop weight, the shortest path will never go through the loop. If the loop has a weight equal to 0, we have an infinite number of shortest paths, which causes the algorithm never to terminate, so, even in this case, it is necessary to remove such a link. The negative weights on the Dijkstra loop do not allow the algorithm, and the BFA reports the existence of a negative cycle, which makes it impossible to calculate the shortest paths, so, also in this case, it is necessary to remove such a link. Only by removing these types of links, did the shortest path finding algorithms work correctly.

Statistics

		Bellman-Ford (ms)	Dijkstra (ms)
N	Valid	20	20
	Missing	0	0
Mean		22309,90	50,25
Median		22353,50	50,00
Mode		21672[a]	46
Std. Deviation		340,686	6,112
Variance		116067,147	37,355
Range		1493	22
Minimum		21672	42
Maximum		23165	64

a. Multiple modes exist. The smallest value is shown

Figure 7 Statistics of the results of action research in the SPSS software tool.

3. Results

This paper addresses three main topics: (1) Which are the most common methods for finding the shortest path, (2) Analyzing and comparing the two most common ones and (3) Applying the SPP methods on an example of a social network. By reviewing the literature, we checked which methods are used to find the shortest path and determined the usage frequency of each SPM, and defined which of them are used in social network analysis. We found the most commonly used shortest path search and chose the Dijkstra and Bellman-Ford algorithms for more detailed examination. By employing a comparative analysis we determined the properties of each algorithm. The findings of the comparison are shown in Table 1. The advantage of the DA is its speed of operation and ease of implementation. However, the advantage of the BFA is the ability to enter negative link weights and optimize performance on graphs with many links. The disadvantage of the DA is not allowing the insertion of negative link weights, while the main disadvantage of the BFA is its less rapid operation and relative implementation complexity. There are different implementations for both algorithms depending on the data structure, in addition to using, storing and processing data.

The practical use of both algorithms was evaluated based on the comparison of the chosen algorithms. Regarding the Dijkstra algorithm, we found that there are several different ways of implementing the algorithm, which differ from each other in the choice of the data structure for operation with a Node Table and a Table of Connections. We also found that the DA runs very fast, and has the most optimal performance for graphs with few links. Regarding the Bellman-Ford algorithm, its use in practice is similar to the DA. There are several implementation methods for the BFA as well, that vary depending on the data structure selected for the node and connection operations. The experiment results show that the BFA runs slower than the DA, which is the result of an additional implementation phase that checks for the existence of a negative cycle. Also, because of multiple bypassing all links, the BFA has the most optimal performance for graphs with many links.

4. Conclusions

The paper addresses the question of whether the aforementioned short-path search algorithm can be applied to the problem of social networking. We conducted an action survey to test the implementation of DA and the BFA on the real Twitter social network. We found that, in the case of the Twitter social network, there are loop links to the same node, and multiple links between two nodes that are not capable of being processed properly by the algorithms, and therefore return the wrong result of finding the shortest path on the social network. For such a result, all links need to be removed, which leads to the conclusion that Dijkstra and Bellman-Ford can only be applied partially to a specific example of a social network. In this paper, a literature review was performed to find existing SPP methods, focusing mainly, but not exclusively, on the SNA domain. The search focus was limited to SPP methods, not considering methods and algorithms which are used primarily for other purposes and are, potentially, also appropriate to find links within the Social Networks field.

Finding the shortest path is a time-consuming task, however, an important task in today's fast-changing economy with many social networks, which are integrated with all aspects of our lives, in all possible domains, from social to professional. To find the most influential users within a network or connect the two suitable nodes in the most optimal way possible provides a strategic advantage of businesses, as well as saves and uses resources smartly.

The addressed problem of finding the shortest path is solvable with many methods and approaches. However, we chose to investigate the solutions from the field of Operational Research, addressing the most used shortest path methods. Since there are several possible solutions, and, in this paper, two were included, this presents one of the main limitations of this paper. The results show that, among the two chosen ones, the Dijkstra algorithm is the faster algorithm, and, thus, suitable for smaller social networks.

Acknowledgement

The authors acknowledge the financial support from the Slovenian Research Agency (Research Core Funding No. P2-0057).

References

[1] A. Disney. (2013). Social network analysis measures.
[2] A. Disney. (2014). Social network analysis: Centrality measures.
[3] Abdelnabi, O. K. A. (2018). Shortest Path Algorithms.
[4] Abiy, T., Pang, H., Williams, C., Khlm, J., & Ross, E. (2020). Dijkstra's Shortest Path Algorithm.
[5] Aini, A., & Salehipour, A. (2012). Speeding up the Floyd–Warshall algorithm for the cycled shortest path problem. Applied Mathematics Letters, 25(1), 1–5. https://doi.org/10.1016/j.aml.2011.06.008
[6] B. Hale. (2015). The History of Social Media: Social Networking Evolution.
[7] Chumbley, A. (2018). Shortest Path Algorithms.
[8] Chumbley, A., Moore, K., Yang, J., Khlm, J., & Ross, E. (2020). Floyd-Warshall Algorithm.
[9] Cohn, M. (2011). Social Media vs Social Networking.
[10] Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). Introduction to algorithms. MIT Press.
[11] D. Desale. (2015). Top 30 Social Network Analysis and Visualization Tools.
[12] Du, D. (2019). Social Network Analysis: Centrality Measures.

[13] Duque, D., Lozano, L., & Medaglia, A. L. (2015). An exact method for the biobjective shortest path problem for large-scale road networks. European Journal of Operational Research, 242(3), 788–797. https://doi.org/10.1016/j.ejor.2014.11.003
[14] Engelbrecht, A., Gerlach, J. P., Benlian, A., & Buxmann, P. (2019). How employees gain meta-knowledge using enterprise social networks: A validation and extension of communication visibility theory. The Journal of Strategic Information Systems, 28(3), 292–309. https://doi.org/10.1016/j.jsis.2019.04.001
[15] Farasat, A., & Nikolaev, A. G. (2016). Social structure optimization in team formation. Computers & Operations Research, 74, 127–142. https://doi.org/10.1016/j.cor.2016.04.028
[16] Fetta, A., Harper, P., Knight, V., & Williams, J. (2018). Predicting adolescent social networks to stop smoking in secondary schools. European Journal of Operational Research, 265(1), 263–276. https://doi.org/10.1016/j.ejor.2017.07.039
[17] Granat, J., & Guerriero, F. (2003). The interactive analysis of the multicriteria shortest path problem by the reference point method. European Journal of Operational Research, 151(1), 103–118. https://doi.org/10.1016/S0377-2217(02)00594-5
[18] Halper, F. (2010). Social Network Analysis: What is it and why should we care?
[19] Hellmann, T., & Staudigl, M. (2014). Evolution of social networks. European Journal of Operational Research, 234(3), 583–596. https://doi.org/10.1016/j.ejor.2013.08.022
[20] J. Clement. (2020). Number of global social network users 2017-2025 Published by J. Clement, Jul 15, 2020.
[21] Jessica Tyler. (2018). The 10 most valuable retail brands in the world.
[22] Kallas, P. (2020). Top 15 Most Popular Social Networking Sites and Apps [2020].
[23] Kordík, P. (2016). Recommender systems explained.
[24] Krasnova, H., Veltri, N. F., Eling, N., & Buxmann, P. (2017). Why men and women continue to use social networking sites: The role of gender differences. The Journal of Strategic Information Systems, 26(4), 261–284. https://doi.org/10.1016/j.jsis.2017.01.004
[25] Kumar, V. (2018). 22 Free Social Network Analysis Tools.
[26] Liu, B., Zhou, Q., Ding, R.-X., Palomares, I., & Herrera, F. (2019). Large-scale group decision making model based on social network analysis: Trust relationship-based conflict detection and elimination. European Journal of Operational Research, 275(2), 737–754. https://doi.org/10.1016/j.ejor.2018.11.075
[27] Mandel, A., & Venel, X. (2020). Dynamic competition over social networks. European Journal of Operational Research, 280(2), 597–608. https://doi.org/10.1016/j.ejor.2019.07.028
[28] McFadden, C. (2020). A Chronological History of Social Media.
[29] Moss, H. (2019). Difference Between Social Media and Social Networking.
[30] Nackerud, S. (2018). Why Engage in (Social) Network Analysis.
[31] Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. Social Networks, 32(3), 245–251. https://doi.org/10.1016/j.socnet.2010.03.006
[32] School, L. U. M. (2018). What is Operational Research?
[33] Sedeño-noda, A., & Colebrook, M. (2019). A biobjective Dijkstra algorithm. European Journal of Operational Research, 276(1), 106–118. https://doi.org/10.1016/j.ejor.2019.01.007
[34] Tanınmış, K., Aras, N., & Altınel, I. K. (2019). Influence maximization with deactivation in social networks. European Journal of Operational Research, 278(1), 105–119. https://doi.org/10.1016/j.ejor.2019.04.010
[35] V. Krebs. (2020). Social Network Analysis: An Introduction.
[36] Wang, Z., You, K., Song, S., & Zhang, Y. (2020). Wasserstein distributionally robust shortest path problem. European Journal of Operational Research, 284(1), 31–43. https://doi.org/10.1016/j.ejor.2020.01.009
[37] Woodruff, J. (2019). Importance of Operations Research in Decision-Making.