A Theoretical Approach for Discovery of Friends from Directed Social Graphs

Sehaj P. Singh
Department of Computer Science
University of Manitoba
Winnipeg, MB, Canada

Carson K. Leung(✉)
Department of Computer Science
University of Manitoba
Winnipeg, MB, Canada
kleung@cs.umanitoba.ca

***Abstract*—Since social networking has been popular in the current era of big data, numerous social networking sites (e.g., Instagram, Twitter) have generated huge volumes of social data at a rapid rate. Embedded into these data are valuable information and knowledge. This calls for social network analysis and mining. In this paper, we specifically aim to discover interesting relationships in directed social graphs via a theoretical approach. More specifically, we examine both graph theory and linear algebra approaches to discover interesting entities (e.g., popular followees, second-degree followees) from social networks represented in the form of big directional graphs.**

***Index Terms*—social network analysis, social network mining, data mining, big data science, big data analytics, graph theory, linear algebra, directed social graphs**

I. Introduction and Related Works

Empowered significantly by data, the modern era of technology is also known as a data-driven era of technology. "Data are new oil", which motivates organizations to develop better solutions to problems in numerous real-life applications and services [12, 13] such as biomedical fields [5], entertainment industry [7], and financial sectors [20], etc. Apart from generating promising results, data are effectively used in decision making processes. All these advancements have been possible due to the fact that huge volumes of valuable data have been generated and collected at rapid rates from different data-rich resources. More than 90% of the data have been generated in last four years [34]. Such a number clearly indicates that exploration and analytics of big data—through various tools such as data science [30], data mining [14] (e.g., periodic pattern mining [4, 19], uncertain mining [16, 23], frequent pattern mining [18, 19]) and machine learning [1, 39]—is going to be an essential key in shaping future technology.

A commonality among the aforementioned applications and services is that they all generally involve humans (i.e., social entities) in social networks. Social network data are the most common form of big data as today's world is connected via social-networking sites (e.g., Facebook, Instagram, LinkedIn, Twitter, etc.). The social network data primarily consist of two components: (a) social entities (i.e., users of the social network) and (b) relationships among these social entities (which can be *directional "following" relationships* between followers-followees, *undirectional mutual friendships*, etc.). For instance, the mutual friendship can be captured by Facebook, where two social entities are connected to each other if they add each other as friends. On the other hand, the "following" relationship can be captured by Twitter (or other popular social networking sites among teenagers like Instagram and Snapchat), where a user (i.e., follower) follows another user (i.e., followee).

The discovery of implicit, previously unknown, potentially useful relationship from such social network data is referred to as *social network analysis*. Previous research in social network analysis includes works on:

- detecting communities [35, 40],
- recommending friends [10, 11, 21, 33],
- propagating influence [6, 15],
- mining social patterns (e.g., interaction patterns [8], "following" patterns [28, 31]),
- visualizing social patterns [17, 29], and
- representing networks (e.g., as compressed bitmaps [22, 24], sparse networks [26, 27, 32], knowledge graphs [36], graph neural networks (GNN) [2]).

In addition, there have been related works on discovering friends various forms of social networks—such as precise networks [3, 25] and uncertain networks [9] both represented in *key-value stores*, as well as networks represented as *undirected graphs* [38] from which *mutual friendship* can be found. In recent years, social networks have also been represented by *graph databases* such as Neo4j [37] so that interesting information can be retrieved by using cypher queries. In contrast, in the current paper, we represent social networks as *directed graphs*—from which we discover the *"following" relationship* among social entities in the networks. In particular, we make good use of graph theory and linear algebra for the discovery. Hence, our *key contributions* of this paper is our theoretical approach for discovery of friends—in particular, discovery of interesting entities (e.g., popular followees, active followers, as well as their second- and higher-degree followees and followers) from social networks represented in the form of big directional graphs.

The remainder of this paper is organized as follows. The next section provides some background information. Section III presents our theoretical approach for analyzing and mining social networks represented in directed social graphs.

IEEE/ACM ASONAM 2020, December 7-10, 2020
978-1-7281-1056-1/20/$31.00 © 2020 IEEE

Evaluation results are shown in Section IV. Finally, Section V draws the conclusions.

II. Background

Definition 1. *A* directed graph *or* digraph G *is an ordered pair* $G = (V, E)$ *where:*

- V*—also shown as* $V(G)$*—is a non-empty set of vertices (i.e.,* $V \neq \emptyset$*);*
- E*—also shown as* $E(G)$*—is a set of* directed edges/arcs*, where each edge is represented by an* ordered *pair of vertices, such that* $E \subseteq V \times V$*. Hence,* $|E| \leq |V| \times (|V| - 1)$*.*

Definition 2. *Vertex* $v_i \in V(G)$ *is said to be* adjacent *to vertex* $v_j \in V(G)$ *in a directed graph* G *if there exists an edge* $e = (v_i, v_j) \in E(G)$*, i.e.,* e *is leaving* v_i *and coming into* v_j*.*

Definition 3. *For a directed graph* $G = (V, E)$ *where* $V(G) = \{v_1, v_2, \ldots, v_n\}$*, the* adjacency matrix $A = [a_{i,j}]_{1 \leq i,j \leq n}$ *of* G *is a* $n \times n$ *matrix where:*

$$a_{i,j} = \begin{cases} 1 & \text{if } (v_i, v_j) \in E(G) \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Then, for a vertex $v_i \in V(G)$*:*

- *the* outer-neighborhood $N_G^+(v_i)$ *is given by*

$$N_G^+(v_i) = \{v_j \in V(G) \mid a_{i,j} = 1\}. \tag{2}$$

- *the* inner-neighborhood $N_G^-(v_i)$ *is given by*

$$N_G^-(v_i) = \{v_j \in V(G) \mid a_{j,i} = 1\}. \tag{3}$$

Definition 4. *Given a directed graph* $G = (V, E)$ *and a vertex* $v \in V(G)$*:*

- *the set of edges leaving* v *is denoted by* $out(v)$*, and the* outdegree *of* v *is given by* $d^+(v) = |out(v)|$*;*
- *the set of edges coming into* v *is denoted as* $in(v)$*, and the* indegree *of* v *is given by* $d^-(v) = |in(v)|$*.*

Lemma II.1. *For a directed graph* $G = (V, E)$ *with its adjacency matrix* $A = [a_{i,j}]_{n \times n}$ *where* $v_i \in V(G)$ *for* $1 \leq i \leq n$*:*

$$d^+(v_i) = |out(v_i)| = \sum_{j=1}^{n} a_{i,j} \tag{4}$$

$$d^-(v_i) = |in(v_i)| = \sum_{j=1}^{n} a_{j,i} \tag{5}$$

Definition 5. *A* directed walk $W = (v_0, v_1, \ldots v_k)$ *of length* k *in a directed graph* $G = (V, E)$ *is a sequence of vertices such that, for all* $i \in [1, k]$*,* $(v_{i-1}, v_i) \in E(G)$*.*

Lemma II.2. *Given an adjacency matrix* A *of directed graph* $G = (V, E)$ *where* $|V| = n$*, for any* $k \in \mathbb{Z}^+$*, the* (i, j)*-th entry of* $A^k = \left[a_{i,j}^{(k)}\right]_{1 \leq i,j \leq n}$ *gives the number of directed* v_i*-*v_j *walks of length* k *in* G*.*

Definition 6. *A* directed path *(or* path *for short) is a directed walk that does not visit the same vertex more than once.*

Definition 7. *A vertex* v *in a directed graph* $G = (V, E)$ *is a* second-degree followee *of another vertex* u *if* $(u, v) \notin E(G)$ *and there exists a directed path of length* 2 *from* u *to* v *(via a third vertex).*

III. Our Social Network Analysis and Mining on Directed Social Graphs

Let us analyze and mine a directed social graph G with its corresponding adjacency matrix A for interesting social entities such as popular followees and second-degree followees.

Question 1. *Who are the most popular users (who have the largest number of followers) in a social graph?*

The most popular followees are those having the largest number of followers, i.e., those having the maximum indegree. By Lemma II.1, the most popular users in a directed social graph $G = (V, E)$, with $V(G) = \{v_1, v_2, \ldots, v_n\}$ can be computed from the set $\{v \in V(G) \mid d_G^-(v) = \Delta(G)\}$ where

$$\Delta(G) = \max_{v_i \in V(G)} \{d_G^-(v_i)\} = \max_{1 \leq i \leq n} \left\{ \sum_{j=1}^{n} a_{j,i} \right\} \tag{6}$$

Example 1. *For illustration, consider a directed social graph* $G = (V, E)$*, where* $V =$*{**A**art, **B**rechtje, **C**as, **D**anique, **E**vert, **F**amke, **G**errit, **H**annie, **I**gnaas, **J**acoba, **K**ees, **L**ara} with directed relationships as indicted below:*

- *Aart follows Brechtje, Danique and Evert.*
- *Brechtje follows Aart, Danique and Ignaas.*
- *Cas follows Brechtje, Danique, Ignaas and Kees.*
- *Danique follows Gerrit and Lara.*
- *Evert follows Aart, Cas, Danique and Famke.*
- *Famke follows Evert and Gerrit.*
- *Gerrit follows Hannie.*
- *Hannie follows Gerrit and Kees.*
- *Ignaas follows Jacoba and Lara.*
- *Jacoba follows Kees.*
- *Kees follows Jacoba.*
- *Lara follows Cas and Kees.*

The corresponding adjacency matrix A *is*

$$A = \begin{pmatrix} & A & B & C & D & E & F & G & H & I & J & K & L \\ A & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ B & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ C & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ D & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ E & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ F & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ G & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ H & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ I & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ J & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ K & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ L & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \tag{7}$$

For instance, $a_{2,4} \in A$ *with a 1 in the 2nd row (i.e., row B) and 4th column (i.e., column D) indicates that Brechtje follows Danique, and* $a_{4,2}$ *with a 0 in the 4th row and 2nd column indicates that Danique does not follow Brechtje. Recall from*

Definition 4 that the indegree of $v \in G$ is the number of edges going into v. Thus, the sum of entries in column i of the adjacency matrix A gives the indegree of vertex v_i:

$$\begin{pmatrix} d^-(A) \\ d^-(B) \\ d^-(C) \\ d^-(D) \\ d^-(E) \\ d^-(F) \\ d^-(G) \\ d^-(H) \\ d^-(I) \\ d^-(J) \\ d^-(K) \\ d^-(L) \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 2 \\ 4 \\ 2 \\ 1 \\ 3 \\ 1 \\ 2 \\ 2 \\ 4 \\ 2 \end{pmatrix} \tag{8}$$

The sums of entries in the third and eleventh columns are 4 each, which is the maximum. These columns correspond to Danique and Kees, i.e.,

$$\max_{v_i \in V(G)} \{d^-_G(v_i)\} = \max_{1 \le i \le n} \left\{ \sum_{j=1}^{n} a_{j,i} \right\} = \{\textit{Danique, Kees}\}.$$

Specifically, Danique is followed by 4 followers (namely, Aart, Brechtje, Cas and Evert). Similarly, Kees is also followed by 4 followers (namely, Cas, Hannie, Jacoba and Lara). Consequently, both Danique and Kees are the most popular users (i.e., followees) *in this example.*

Question 2. *Who are the least popular users (who have the smallest number of followers) in a social graph?*

The least popular followees are those having the smallest number of followers, i.e., those having the minimum indegree. By Lemma II.1, the least popular users in the directed social graph $G = (V, E)$ with $V(G) = \{v_1, v_2, \ldots, v_n\}$ can be computed from the set $\{v \in V(G) \mid d^-_G(v) = \delta(G)\}$ where

$$\delta(G) = \min_{v_i \in V(G)} \{d^-_G(v_i)\} = \min_{1 \le i \le n} \left\{ \sum_{j=1}^{n} a_{j,i} \right\} \tag{9}$$

Example 2. *Let us continue with Example 1. The sums of entries in the six and eight columns of the adjacency matrix A are 1 each, which is the minimum. These columns correspond to Famke and Hannie, i.e.,*

$$\min_{v_i \in V(G)} \{d^-_G(v_i)\} = \min_{1 \le i \le n} \left\{ \sum_{j=1}^{n} a_{j,i} \right\} = \{\textit{Famke, Hannie}\}.$$

Specifically, Famke is followed by only 1 follower (namely, Evert). Similarly, Hannie is also followed by only 1 follower (namely, Gerrit). Thus, both Famke and Hannie are the least popular users (i.e., followees) *in this example.*

From these two examples, the *indegree* of the vertex is observed to act as the central measure for "popularity" in a directed social graph. A common feature in social networking sites (e.g., Instagram, Snapchat, Twitter) is recommending users (e.g., popular followees). In most cases, this recommendation is based on the concept *second-degree followee* where a user C is recommended to another user A when A is following a user B who follows C. Naturally progressing from above two questions, one would like to conduct social network analysis and mining on the second-degree followees.

Lemma III.1. *For a directed graph $G = (V, E)$ with an adjacency matrix $A_{n\times n}$ and $V(G) = \{v_1, v_2, \ldots, v_n\}$, the number of second-degree followees $\text{secdeg}_G(v_i)$ of v_i $(1 \le i \le n)$ can be computed by:*

$$\text{secdeg}_G(v_i) = \begin{cases} \varrho_A(i) - 1 - \rho_A(i) & \textit{if } d^+_G(v_i) \neq 0,\ \exists\ v_i\textit{-}v_i \textit{ path} \\ \varrho_A(i) - \rho_A(i) & \textit{if } d^+_G(v_i) \neq 0,\ \nexists\ v_i\textit{-}v_i \textit{ path} \\ 0 & \textit{otherwise} \end{cases} \tag{10}$$

where

- *$\varrho_A(i)$ is the number of non-zero entries in the i-th row of the square matrix A^2 (which is the square of matrix A), and*
- *$\rho_A(i)$ is the number of corresponding entries that are not zero in the i-th rows of both matrices A and A^2.*

Question 3. *Which users have the largest numbers of second-degree followees in a directed social graph $G = (V, E)$ with $V(G) = \{v_1, v_2, \ldots, v_n\}$?*

Let $A_{n\times n}$ be the adjacency matrix corresponding to G. Then, $A^2 = [a^2_{i,j}]_{n\times n}$ is its square. By our Lemma III.1, users who have the maximum number of second-degree followees in G can be computed from the set

$$\{v_j \in V(G) \mid \text{secdeg}_G(v_j) = \max_{1 \le i \le n} \{\text{secdeg}_G(v_i)\}\}$$

To compute $\text{secdeg}_G(v_i)$ and $\rho_A(i)$, we form a matrix $R = [r_{i,j}]_{n\times n}$ such that

- diagonal entries $r_{i,i} \in R$ with directed v_i-v_i walks of length 2 are marked with the symbol E:

$$r_{i,i} = \begin{cases} E & \text{if } a^2_{i,i} > 0 \\ 0 & \text{otherwise (i.e., } a^2_{i,i} = 0) \end{cases} \tag{11}$$

  Note that, although the value of $a^2_{i,i}$ indicates the number of v_i-v_i walks of length 2, we only need to know the corresponding Boolean value (i.e., whether or not there exists v_i-v_i walks of length 2) in the computation of $\text{secdeg}_G(v_i)$. Specifically, Eq. (10) can be simplified as:

$$\text{secdeg}_G(v_i) = \begin{cases} \varrho_A(i) - 1 - \rho_A(i) & \text{if } r_{i,i} = E \\ \varrho_A(i) - 0 - \rho_A(i) & \text{if } r_{i,i} = 0 \end{cases} \tag{12}$$

- non-diagonal entries $r_{i,j} \in R$ (where $i \neq j$) conveying corresponding non-zero entries in both A and A^2 are marked by the symbol X:

$$r_{i,j} = \begin{cases} X & \text{if } a_{i,j}{\neq}0 \ \&\ a^2_{i,j}{\neq}0 \text{ (where } i{\neq}j) \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

  In other words, non-diagonal entries of R can be considered as an intersection of the corresponding entries A and A^2, i.e., $A \cap A^2$:

$$r_{i,j} = a_{i,j} \cap a^2_{i,j} \quad (\text{where } i{\neq}j) \tag{14}$$

Then, $\rho_A(i)$ in Eq. (10) or Eq. (12) can be computed by counting the number of X in $r_{i,j}$ (i.e., in the i-th row).

Example 3. *Let us continue with Example 1. We compute A^2 as the square of the adjacency matrix A:*

$$A^2 = \begin{pmatrix} 2 & 0 & 1 & 2 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 2 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 2 & 0 & 2 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 2 & 0 & 2 & 2 & 0 & 2 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix} \quad (15)$$

We observe from the matrix that, for all $v_i \in V(G)$, $d_G^+(v_i) \neq 0$ and thus $\text{secdeg}_G(v_i) = \varrho_A(i) - 1 - \rho_A(i)$ or $\varrho_A(i) - \rho_A(i)$. For instance, let us consider Aart:

- *The first row $[a^2_{1,j}]$ of A^2 is*

$$[a^2_{1,j}] = \begin{pmatrix} 2 & 0 & 1 & 2 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \quad (16)$$

  which reveals that $\varrho_A(1) = 7$ non-zero entries.
- *The diagonal entry $a^2_{1,1} = 2$ reveals that there exist 2 directed walks of length 2 from Aart to Aart (namely, Aart-Brechtje-Aart and Aart-Evert-Aart). To simplify, as $a^2_{1,1} \neq 0$, there exist directed walks of length 2 from Aart to Aart.*
- *Then, the first row $[a_{1,j}]$ of A from Eq. (7) is*

$$[a_{1,j}] = \begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad (17)$$

  which reveals that $\rho_A(1) = 1$ entry not having zero in the 1st row of both A and A^2 (specifically, $a_{1,4} \in A$ and $a^2_{1,4} \in A^2$ are not zero).
- *Consequently,* $\text{secdeg}_G(Aart) = \varrho_A(1) - 1 - \rho_A(1) = 7 - 1 - 1 = 5$.

Alternatively, with A^2 and Lemma III.1, we calculate the number of second-degree followees for each of the vertices of G as follows. To compute $\text{secdeg}_G(v_i)$ and $\rho_A(i)$, we form matrix $R = [r_{i,j}]$:

$$R = \begin{pmatrix} E & 0 & 0 & X & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & E & 0 & X & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & X & 0 & 0 & 0 & 0 & X & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & X & E & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & E & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & E & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & E & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & E & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & E & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X & 0 \end{pmatrix} \quad (18)$$

Here, diagonal entries $r_{i,i} \in R$ marked with E are those with directed v_i-v_i walks of length 2 (e.g., $r_{1,1}$ is marked by E because there exist directed Aart-Aart walks of length 2; $r_{3,3}$ is not marked by E because there does not exist any directed Cas-Cas walk of length 2). Non-diagonal entries $r_{i,j} \in R$ marked with X are those with non-zero corresponding entries in both A and A^2 (e.g., $r_{1,4}$ is marked by X because $a_{1,4} = 1 \neq 0$ and $a^2_{1,4} = 2 \neq 0$; $r_{1,2} = 0$ because $a^2_{1,2} = 0$; $r_{1,3} = 0$ because $a_{1,3} = 0$; $r_{1,8} = 0$ because $a_{1,8} = a^2_{1,8} = 0$).

With A^2 in Eq. (15) and R in Eq. (18), we compute the second-degree followees of vertices in G by Eq. (10):

$$\begin{pmatrix} \text{secdeg}(A) \\ \text{secdeg}(B) \\ \text{secdeg}(C) \\ \text{secdeg}(D) \\ \text{secdeg}(E) \\ \text{secdeg}(F) \\ \text{secdeg}(G) \\ \text{secdeg}(H) \\ \text{secdeg}(I) \\ \text{secdeg}(J) \\ \text{secdeg}(K) \\ \text{secdeg}(L) \end{pmatrix} = \begin{pmatrix} 7 \\ 6 \\ 6 \\ 3 \\ 7 \\ 5 \\ 2 \\ 2 \\ 2 \\ 1 \\ 1 \\ 5 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ 2 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 5 \\ 4 \\ 4 \\ 3 \\ 5 \\ 4 \\ 1 \\ 1 \\ 2 \\ 0 \\ 0 \\ 4 \end{pmatrix} \quad (19)$$

Here, the fifth row $[a^2_{5,j}]$ of A^2 reveals that $\varrho_A(5) = 7$ non-zero entries. From R in Eq. (18), the diagonal entry $r_{5,5} = E$ reveals that there exist directed walks of length 2 from Evert to Evert. Thus, as per Eq. (12), $\text{secdeg}_G(v_5) = \varrho_A(5) - 1 - \rho_A(5) = 6 - \rho_A(5)$ *where $\rho_A(5) = 1$ because there is only 1 X in the fifth row $[r_{5,j}]$ of R. Hence,* $\text{secdeg}_G(v_5) = 5$.

Similar computation is applied to other rows representing other social entities. As a result, Eq. (19) reveals that $\text{secdeg}_G(Aart) = 5$, $\text{secdeg}_G(Evert) = 5$, *and other 10 users have second-degrees less than 5. Specifically, Aart has 5 second-degree followees (namely, Cas, Famke, Gerrit, Ignaas and Lara), whereas Evert also has 5 second-degree followees (namely, Brechtje, Gerrit, Ignaas, Kees and Lara). Consequently,* both Aart and Evert are users who have largest numbers of second-degree followees *in the illustrative social network.*

IV. Evaluation

To evaluate our presented graph theory and linear algebra approach for discovery of "following" relationships from directed social graphs, we used the datasets from Stanford Large Network Dataset Collection for the Stanford Network Analysis Project (SNAP)[1]:

- ego-Gplus dataset, which contains 13,673,453 directed edges among 107,614 nodes in social circles from Google+; and
- ego-Twitter dataset, which contains 1,768,149 directed edges among 81,306 nodes in social circles from Twitter.

Results show that, when the number of nodes increases, the corresponding number of directed edges increases and thus runtime increases for both datasets. This shows the scalability of our approach.

V. Conclusions

In this paper, we described our graph theory and linear algebra approach to discover friends (e.g., popular followees, second-degree followees)—through the "following" relationships—from social networks represented as directed

[1] https://snap.stanford.edu/data/

social graphs. Evaluation results on two real-life datasets show the practicality and scalability of our theoretical approach. As ongoing and future work, we explore alternative approaches to discover interesting relationships from directed (and undirected) social graphs.

ACKNOWLEDGMENT

This work is partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), as well as the University of Manitoba.

REFERENCES

[1] U. Arora, et al., "Multitask learning for blackmarket tweet detection," in IEEE/ACM ASONAM 2019, pp. 127-130.
[2] A. Benamira, et al., "Semi-supervised learning and graph neural networks for fake news detection," in IEEE/ACM ASONAM 2019, pp. 568-569.
[3] P. Braun, et al., "Knowledge discovery from social graph data," Procedia Computer Science 96, 2016, pp. 682-691.
[4] A.K. Chanda, et al., "A new framework for mining weighted periodic patterns in time series databases," ESWA 79, 2017, pp. 207-224.
[5] J. De Guia, et al., "DeepGx: deep learning using gene expression for cancer classification," in IEEE/ACM ASONAM 2019, pp. 913-920.
[6] C. Doyle, et al., "Mining personal media thresholds for opinion dynamics and social influence," in IEEE/ACM ASONAM 2018, pp. 1258-1265.
[7] C. Fan, et al., "Social network mining for recommendation of friends based on music interests," in IEEE/ACM ASONAM 2018, pp. 833-840.
[8] A. Fariha, et al., "Mining frequent patterns from human interactions in meetings using directed acyclic graphs," in PAKDD 2013, Part I, pp. 38-49.
[9] C.S.H. Hoi, et al., "Supporting social information discovery from big uncertain social key-value data via graph-like metaphors," in ICCC 2018, pp. 102-116.
[10] F. Jiang, et al., "Big data mining of social networks for friend recommendation," in IEEE/ACM ASONAM 2016, pp. 921-922.
[11] F. Jiang, et al., "Finding popular friends in social networks," in CGC 2012, pp. 501-508.
[12] F. Jiang, C.K. Leung, "A data analytic algorithm for managing, querying, and processing uncertain big data in cloud environments," Algorithms 8(4), 2015, pp. 1175-1194.
[13] A. Kobusinska, et al., "Emerging trends, issues and challenges in Internet of Things, big data and cloud computing," FGCS 87, 2018, pp. 416-419.
[14] C.K. Leung, "Big data analysis and mining," in Encyclopedia of Information Science and Technology, 4e, 2018, pp. 338-348.
[15] C.K. Leung, "Mathematical model for propagation of influence in a social network," in Encyclopedia of Social Network Analysis and Mining, 2e, 2018, pp. 1261-1269.
[16] C.K. Leung, "Uncertain frequent pattern mining," in Frequent Pattern Mining, 2014, pp. 417-453.
[17] C.K. Leung, C.L. Carmichael, "Exploring social networks: a frequent pattern visualization approach," in IEEE SocialCom 2010, pp. 419-424.
[18] C.K. Leung, C.L. Carmichael, "FpVAT: a visual analytic tool for supporting frequent pattern mining," ACM SIGKDD Explorations 11(2), 2009, pp. 39-48.
[19] C.K. Leung, C.L. Carmichael, "FpViz: a visualizer for frequent pattern mining," in ACM KDD-VAKD 2009, pp. 30-39.
[20] C.K. Leung, et al., "A machine learning approach for stock price prediction," in IDEAS 2014, pp. 274-277.
[21] C.K. Leung, et al., "Big data analytics of social network data: Who cares most about you on Facebook?" Highlighting the Importance of Big Data Management and Analysis for Various Applications, 2018, pp. 1-15.
[22] C.K. Leung, et al., "Efficient and flexible compression of very sparse networks of big data," in Big Data and Social Media Analytics - Trending Applications, 2021
[23] C.K. Leung, et al., "Fast algorithms for frequent itemset mining from uncertain data," in IEEE ICDM 2014, pp. 893-898.
[24] C.K. Leung, et al., "Flexible compression of big data," in IEEE/ACM ASONAM 2019, pp. 741-748.
[25] C.K. Leung, et al., "Knowledge discovery from big social key-value data," in IEEE CIT 2016, pp. 484-491.
[26] C.K. Leung, et al., "Mining 'following' patterns from big but sparsely distributed social network data," in IEEE/ACM ASONAM 2018, pp. 916-919.
[27] C.K. Leung, et al., "Mining 'following' patterns from big sparse social networks," in IEEE/ACM ASONAM 2016, pp. 923-930.
[28] C.K. Leung, et al., "Parallel social network mining for interesting 'following' patterns," CCPE 28(15), 2016, pp. 3994-4012.
[29] C.K. Leung, et al., "Visual analytics of social networks: mining and visualizing co-authorship networks," in HCII-FAC 2011, pp. 335-345.
[30] C.K. Leung, F. Jiang, "A data science solution for mining interesting patterns from uncertain big data," in IEEE BDCloud 2014, pp. 235-242.
[31] C.K. Leung, F. Jiang, "Big data analytics of social networks for the discovery of 'following' patterns," in DaWaK 2015, pp. 123-135.
[32] C.K. Leung, F. Jiang, "Efficient mining of 'following' patterns from very big but sparse social networks," in IEEE/ACM ASONAM 2017, pp. 1025-1032.
[33] M. Mai, et al., "Big data analytics of Twitter data and its application for physician assistants: who is talking about your profession in Twitter?" in Data Management and Analysis, 2020, pp. 17-32.
[34] B. Marr, "How much data do we create every day? The mind-blowing stats everyone should read," Forbes, May 21, 2018.
[35] K. Ozturk, et al., "An evolutionary approach for detecting communities in social networks," in IEEE/ACM ASONAM 2017, pp. 966-973.
[36] A. Pingle, et al., "RelExt: relation extraction using deep learning approaches for cybersecurity knowledge graph improvement," in IEEE/ACM ASONAM 2019, pp. 879-886.
[37] M. Sestak, et al., "Applying k-vertex cardinality constraints on a Neo4j graph database," FGCS 115, 2021, pp. 459-474.
[38] S.P. Singh, et al., "A theoretical approach to discover mutual friendships from social graph networks," in iiWAS 2019, pp. 212-221.
[39] K. Tu, et al., "gl2vec: learning feature representation using graphlets for directed networks," in IEEE/ACM ASONAM 2019, pp. 216-221.
[40] N. Zarayeneh, A. Kalyanaraman, "A fast and efficient incremental approach toward dynamic community detection," in IEEE/ACM ASONAM 2019, pp. 9-16.