

## Central Actor Identification of Crime Group using Semantic Social Network Analysis

S P Tahalea<sup>\*1</sup>, S N Azhari<sup>2</sup>

<sup>1</sup>Department of Informatics Engineering, UPN “Veteran” Yogyakarta, Indonesia

<sup>2</sup>Department of Computer Science and Electronics, FMIPA UGM, Yogyakarta, Indonesia

E-mail: sylvert@upnyk.ac.id<sup>1</sup>, arisn@gmail.com<sup>2</sup>

**Masuk: 22 Juli 2019, direvisi: 11 Agustus 2019, diterima: 12 Agustus 2019**

**Abstrak.** Polisi sebagai penegak hukum yang merupakan pihak berwenang dalam hal perlindungan sosial diharapkan melakukan upaya pencegahan dan investigasi juga penyelesaian tindak kejahatan yang terjadi di masyarakat. Penelitian ini dapat membantu polisi untuk mengidentifikasi aktor utama lebih cepat dan mengarah untuk memecahkan tindak kejahatan. Penggunaan sentralitas menyeluruh (*overall centrality*) sangat membantu dalam menentukan pelaku utama berdasarkan perhitungan sentralitas lainnya. Tujuan dari penelitian ini adalah untuk mengidentifikasi pelaku utama dari tindak kejahatan yang dilakukan oleh beberapa orang. Analisis jaringan sosial semantik digunakan untuk melakukan identifikasi aktor utama dengan menggunakan lima pengukuran sentralitas, seperti sentralitas derajat, sentralitas perantara, sentralitas kedekatan, sentralitas eigenvektor, dan sentralitas menyeluruh. Adapun hubungan antara aktor, penelitian ini menggunakan hubungan sosial seperti persahabatan, kolega, keluarga, pacar atau kekasih, dan kenalan. Hubungan antara aktor yang diukur dengan menggunakan empat pengukuran sentralitas pertama kemudian diakumulasikan oleh sentralitas menyeluruh untuk menentukan aktor utama. Hasilnya menunjukkan 80,39% akurasi dari 102 kasus pidana yang dikumpulkan dengan setidaknya tiga pelaku yang terlibat dalam setiap kasus.

**Kata kunci:** Analisis jaringan sosial; *overall centrality*; kelompok kriminal; pelaku utama

**Abstract.** The Police as law enforcers who authorize in terms of social protection are expected to do both the prevention and investigation efforts also the settlement of criminal cases that occurred in the society. This research can help police to identify the main actor faster and leads to solving crime-cases. The use of overall centrality is very helpful in determining the main actors from other centrality measures. The purpose of this research is to identify the central actor of crimes done by several people. Semantic Social Network Analysis is used to perform central actor identification using five centrality measurements, such as degree centrality, betweenness centrality, closeness centrality, eigenvector centrality, and overall centrality. As for the relationship between actors, this research used social relation such as friendship, colleague, family, date or lover, and acquaintances. The relationship between actors is measured by first four centrality measures then accumulated by overall centrality to determine the main actor. The result showed 80.39% accuracy from 102 criminal cases collected with at least three actors involved in each case.

**Keywords:** Social Network Analysis; overall centrality; crime group; central actor

## 1. Introduction

Crime is a phenomenon in the society which is considered as a deviant act that disturbs the public. It can be done by several people, whether planned or not. This results in the emergence of small groups that committed criminal acts on a small scale to the formation of an organization that committed large-scale crimes. Criminal network is a new social relationship formed when criminals conduct crime together [1], [2]. This network can be a small group or large organization. A criminal group usually has only one main actor that has more power than any other actor in the group [3]. Main actor or central actor can be the leader, the strategist, or the most active member of the group.

Detecting central actor or main actor of a group has advantages depending on the needs but it is not an easy job [3]. The requirement to detect or identify a central actor from a group or a network is the complete network. Once the complete network is discovered, it can be visualized into a graph which makes the representation and cognitive tendencies of users improvised [4]. It is easy to draw a social graph for a small social group, but it can be difficult or even not feasible for large group [3]. A graph can be built using actors as node and relationship as the edge, then it can be analyzed using many methods. One of them is Social Network Analysis.

Social Network Analysis (SNA) has been long used to identify and determine actors role in a social group [5] in form of graphs [6]. The advantage is researcher can construct, visualize, and analyze the networks [7]. This makes SNA can be applied to any fields. In the medical field, SNA is used to analyze the social network influence in tobacco use to support protocol about tobacco in Australia [8]. SNA can also be applied to management when it is necessary. Jarman [9] said that SNA can be precedents to manage events in a city, where the social relationship can make a good impact on a festival economy. Meanwhile, [10] show that SNA was very valuable in identifying areas for improvement and can be utilized for effective decision-making. In the literary field, SNA is used to reveal relevant phenomena between research articles and can deepen the understanding of collecting and analyze social network data [11]. SNA also helps establish relationships between the most relevant authors in the citation of articles, which seems to not use any other methods such as bibliometrics and applied network analysis [12].

A relationship between actors or nodes in SNA is an important factor. A research about knowledge transfer showed that SNA could be used to analyze knowledge transfer in a project based on employees ties between each other, the stronger the ties, the more knowledge can be transferred [13]. In another research showed that relationship appeared as an important factor to analyze disaster-safety industry in Korea [14].

The use of SNA was later combined with a semantic framework for analyzing network in any field of study. Based on the success of SNA in analyzing social networks, Kassiri [15] later proposed social network ontology to make a unified semantic model online social networks and it results in attracting a lot of interest from many fields. Bisio [16] proposed a new method ConceptNet, a semantic network that runs on the Twitter social network to show the chain of Twitter traffic. Not only in Twitter but Fu [17] also used semantic distance to build and analyze relevancy between nodes and contents in Weibo Social Network. The result showed that semantic framework built helped understand the semantic meaning of texts. Raji and Surendran [18] then present Resource Description Framework (RDF) approach on social network analysis in semantic web areas.

In crimes study, SNA used to extract criminal data such as criminal role, harm score, travel distance [19]. The advantages of using SNA in crimes study are its visualization [4], finding link between crimes, profiling offenders, identifying criminal networks, and predicting criminal suspects [20]. The graph built can be analyzed and shows a lot of information. Rashed [4] utilizes this advantage to make a visualization of narcotics circulation in Chicago. SNA can also be used to identify the leader or the most influencing actor in a criminal group [5], [21]. The criminal network must be constructed to identify the most influencing actor with node represent criminal actors and edge represent the relationship or communications between actors [21]. Tayebi et. al. [22] used five centrality measurement such as degree, betweenness, closeness, eigenvector, and page rank centrality in their research to locate central actor for co-offender but still it only showed one actor for each centrality measures, meanwhile it helped identify actors engagement in large scale network. Whereas

for the criminal actors, there are differences between crime group like co-offenders or gamblers group with crime organization based on their need. People within crime organization are fully organized while people in crime group sometimes are spontaneous and unorganized like social group [23] that can be analysed using social network analysis [24].

The purpose of this research is to help police identify the central actor of a criminal group using social relationships as edge and criminals as actors. Computational method proposed in this study can help identify central actor faster than it used to be by the police. This method can save time from doing interviews with all the criminals, not to say when some of criminal still missing or hidden. The snowball sampling used in this research can build large-scale network using information gathered from current interview result.

Social relationships used in this research are friendship, colleague, family, date or lover, and acquaintances. Relationship Centrality measures used in this research are degree centrality, betweenness centrality, closeness centrality, eigenvector centrality, and overall centrality. Overall centrality is used to accumulate another centrality measure so that the actor that has the highest centrality value from all centrality measures can be identified as the central actor of a criminal group. Data used in this research supported by Yogyakarta City Police.

## 2. Methodology

This research used five centrality measures that are degree centrality, betweenness centrality, closeness centrality, eigenvector centrality, and overall centrality. Each centrality measure was used to identify the central actor from many perspectives, except for overall centrality that was used to accumulate score from other centrality and shown the result as the central actor [25]. The data was collected using snowball sampling technique for each case.

### 2.1. Degree Centrality

Degree centrality can be measured from the number of connections owned by actors. More connections mean more central [3], [6], [26]. Degree centrality of nodes in a network can be measured using Equation (1).

$$C_D(i) = \sum \frac{d_i}{n-1} \quad (1)$$

$C_D(i)$  is degree centrality of actor  $i$ ,  $d$  is the number of relationships owned by actor  $i$ , and  $n$  is the number of actors in the network.

### 2.2. Betweenness Centrality

Betweenness centrality is a centrality measure that considers the central actor from being a bridge that connecting other actors [26]. It can be measured by using the number of shortest paths which passes through the node as shown in Equation (2).

$$C_B(i) = \sum \frac{g_{j,k}(i)}{g_{j,k}} \quad (2)$$

$C_B(i)$  is betweenness centrality of actor  $i$ ,  $g_{j,k}$  is the number of shortest-paths from actor  $j$  to actor  $k$ , and  $g_{j,k}(i)$  is the number of shortest paths from actor  $j$  to actor  $k$  through actor  $i$ .

### 2.3. Closeness Centrality

Closeness centrality shows how close an actor to all other actors in the network [5]–[7], [11], [21], [26]. Actors that have high closeness centrality score are usually easier or quicker to reach other actors in the network [12], [18], [26]. It can be measured using Equation (3).

$$C_C(i) = \frac{n-1}{\sum D_{ij}} \quad (3)$$

$C_C(i)$  is the closeness centrality of actor  $i$ ,  $D_{ij}$  is the shortest path from actor  $i$  to other actors  $j$  in the network, and  $n$  is the number of actors exist in the network.

#### 2.4. Eigenvector Centrality

Eigenvector centrality determines the connection of a node [7]. It generalizes the degree centrality and considers the importance of a node based on other connected nodes [3], [6], [14], [26]. It can be measured using Equation (4).

$$C_e(i) = \frac{1}{\lambda} \sum_{j \in G} a_{i,j} \quad (4)$$

$C_e(i)$  is the eigenvector centrality of node  $i$ ,  $\lambda$  is a constant,  $a_{i,j}$  is the adjacency matrix from node  $i$  to any other nodes in network  $G$ .

#### 2.5. Overall Centrality

Overall centrality is a centrality measure that indicates the overall role of an actor inside a network [25]. It can be measured based on results from any other centrality measures as shown in Equation (5).

$$O_i = v_i CD_i + v_i CB_i + v_i CC_i + v_i CE_i \quad (5)$$

#### 2.6. Snowball Sampling

Snowball sampling is a sampling technique to gather data or information from a specific group of people [27] and mostly used to study hidden population [28]. Snowball sampling is used in this research to reveal every actor involved in each case either directly or indirectly because it is very effective when the population are hidden or hardly accessible [27]. This technique also allows to make statements about the network structure and actors relationships property [29]. Vogt in [30] states that snowball sampling can be defined as a technique to find other subjects from the information given by the first subject. In simple words, a large-scale network can be built by using information from an actor.

Crime group network in this research is built from information gathered from the first actor for each case, then linked to other actors involved with him. Other information gathered will indicate their relationship and engagement inside or outside a case.

### 3. Result and Discussion

#### 3.1. Relationship Weight

Weighting is based on the results of the interview with the police. There are several relationships that can be used in this study as shown in Table 1. The bigger the number, the closer the relationship between actors, compared to smaller weight. Closer relationship can give impact in someone decision making to either commit crime or not.

**Table 1.** Relationship Weight between Actors

Weight	Distance	Relationship
1	Very Close	Family
1	Very Close	Spouse
2	Close	Colleague
2	Close	Close/Best Friend
2	Close	Date / Lover
3	Fairly Close	Neighbour
3	Fairly Close	Friend
4	Not Close	Acknowledge

### 3.2. Ontology

Ontology built in this study uses 2 classes with 7 datatype properties and 11 object properties as shown in Table 2, Table 3, and Table 4 respectively. Class for crime case instances is kasusKejahatan and class for criminals is Person. One group crime can be done by two or more actors, and one actor can commit on one or more group crime.

**Table 2. Classes**

No.	Class	Restriction
1.	kasusKejahatan	-
2.	Person	-

Both classes have their own datatype properties to define properties value of an individual. Each class has its own datatype properties. Class kasusKejahatan has datatype properties such as jenisKasus for types of crime committed, idKasus for case numbers, Pasal for articles in law that clarify about that crime, tanggalKejadian for date and time that crime happened. Whereas class Person has ID to define identity/license number of actors, namaLengkap for actors name, and alias for their aliases.

**Table 3. Datatype Properties**

No.	Datatype property	Domain	Range
1.	jenisKasus	kasusKejahatan	xsd:string
2.	idKasus	kasusKejahatan	xsd:string
3.	Pasal	kasusKejahatan	xsd:string
4.	tanggalKejadian	kasusKejahatan	xsd:dateTime
5.	ID	Person	xsd:string
6.	namaLengkap	Person	xsd:string
7.	Alias	Person	xsd:string

Relationship between individuals is defined by object properties. This ontology consists of keluarga for family relationship, kenalan for acknowledge relationship, pacar for girlfriend or boyfriend relation, rekanKerja for co-worker relation, suamiIstri for spouse (husband or wife) relation, teman for friendship relation, temanDekat for best friend or closest friendship relation, tetangga for neighbourhood relation, terlibat and melibatkan for bound between criminal case and criminals, pelakuUtama for central actor of crime group case.

**Table 4. Object Properties**

No.	Object Property	Domain	Range	Relational Characteristic
1.	keluarga	Person	Person	symmetric
2.	kenalan	Person	Person	symmetric
3.	pacar	Person	Person	symmetric
4.	rekanKerja	Person	Person	symmetric
5.	suamiIstri	Person	Person	symmetric
6.	teman	Person	Person	symmetric
7.	temanDekat	Person	Person	symmetric
8.	tetangga	Person	Person	symmetric
9.	terlibat	Person	kasusKejahatan	-
10.	melibatkan	kasusKejahatan	Person	-
11.	pelakuUtama	kasusKejahatan	Person	functional

### 3.3. Central Actor Identification

Central actor identification is done by measuring the degree centrality, betweenness centrality, closeness centrality, eigenvector centrality, and overall centrality. All centrality is measured based on the group-crime case data as the example given in Table 5.

**Table 5.** Crime group case sample

Date	Actors	Relation	Relationship Type	Crimes	Articles
30/01/2013	A	A-B	Family	Gambling	303 KUHP
	B	A-C	Friendship		
	C	A-D	Family		
	D	B-C	Friendship		
		B-D	Friendship		
		C-D	Friendship		

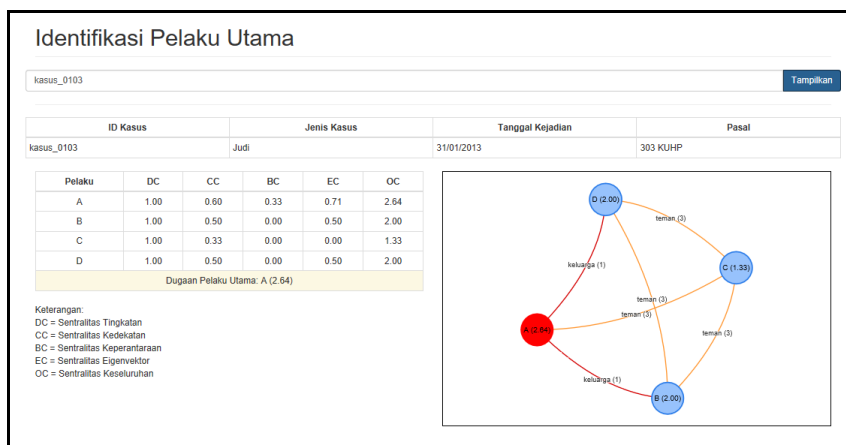
Table 5 shows a data sample from real cases. There were four actors that were connected each other with different relationship. They all committed crime that violated Indonesian law based on KUHP articles 303. The relationship between actors then is measured using Equation (1), (2), (3), and (4) to find their degree centrality, betweenness centrality, closeness centrality, and eigenvector centrality. Overall centrality can be measured based on the other four centrality measures. The result from this network is shown in Table 6.

**Table 6.** Centrality measures

Actors	Degree	Betweenness	Closeness	Eigenvector	Overall
A	1.00	0.33	0.60	0.71	2.64
B	1.00	0.00	0.50	0.50	2.00
C	1.00	0.00	0.33	0.00	1.33
D	1.00	0.00	0.50	0.50	2.00

Table 6 shows that all actors have same centrality degree, which means they all know each other or are connected in this crime network. Actor A has the highest betweenness centrality. It means that A is an information courier or middle man that is supposed to pass information from an actor to others. A also has the highest closeness centrality, which means that A has closer relationship to other actors. Closer relationship means A can influence or affect other actors' decision. The highest eigenvector centrality means A is the most important actor among the others in this network based on their network. A has 2.64 overall centrality value which is the highest overall centrality value. This meanshe is the central actor of this network. If there are more than one actor that gets the highest overall centrality value, then whoever have the highest centrality value from other centrality measures will be the central actor.

Visualization of this procedure can also help police to identify central actor faster than reading the text-based report. In the visualization shown in Figure 1, the blue nodes are actors while the central actor is marked with red node. The edge (line) shows the relationship type and weight between actors.



**Figure 1.** Network Visualization

Evaluation for this method using accuracy measure and it has good performance by correctly showing 82 main actors from 102 crime group cases. One of the main limitations of this method is there are same relationships or kinships between actors that cause same similarity measure for some actors, which makes it hard to decide the central actor for some crime group cases.

### 3.4. Actor Engagement

Network built from the data can also show other actor engagements although he was not involved in a case specifically. In other words, the network also carries the actors and cases connection. This feature help police for further analysis of the case. Number of crime cases committed by an actor varies from a single case into multiple cases. Relationship between actors can build a larger network and can merge with other network cases. Figure 2 shows that actor D did engage with actor A, B, and C in case number 103. This sample shows the ability to visualize and the possibility to detect any other related criminal actors to a central actor of a case.

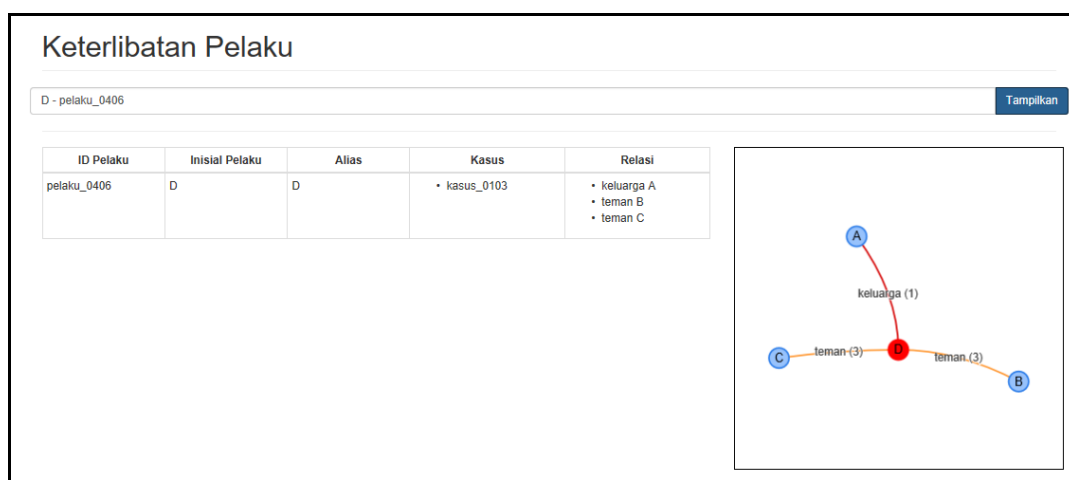


Figure 2. Actor Engagement

## 4. Conclusion

Based on the results of the research and the results of system testing, it can be concluded that the use of Semantic Social Network Analysis (SNA) is able to identify the main actors with an accuracy of 80.39%. This is due to the small amount of data used in the research and the relationships between the actors. Many actors have the same relationship, resulting in the same overall centrality each actor has in one case. This can be overcome by further studies on the closeness of social relations and applied to social network analysis.

The use of overall centrality in the semantic SNA has succeeded in narrowing the chances of an actor being the main actor. This measurement of centrality successfully combines the value of each centrality measure, so that the actor who has the highest accumulated value can be considered to be the main actor.

## 5. References

- [1] G. Marciani, M. Porretta, M. Nardelli, and G. F. Italiano, "A data streaming approach to link mining in criminal networks," *Proc. - 2017 5th Int. Conf. Futur. Internet Things Cloud Work. W-FiCloud 2017*, pp. 138–143, 2017.
- [2] M. a. Tayebi and U. Glasser, "Organized Crime Structures in Co-offending Networks," *2011 IEEE Ninth Int. Conf. Dependable, Auton. Secur. Comput.*, pp. 846–853, 2011.
- [3] S. M. Alam, N. Islam, and M. S. Hosain, "Detecting most central actors of an unknown network using friendship paradox," in *2016 International Conference on Informatics and Computing, ICIC 2016*, 2017, pp. 343–348.
- [4] A. Rasheed, U. K. Wiil, and M. Niazi, "Evaluating PEVNET: A framework for visualization of

- criminal networks,” *Commun. Comput. Inf. Sci.*, vol. 540, no. Asonam, pp. 131–149, 2015.
- [5] K. Taha and P. D. Yoo, “SIIMCO: A forensic investigation tool for identifying the influential members of a criminal organization,” *IEEE Trans. Inf. Forensics Secur.*, vol. 11, no. 4, pp. 811–822, 2016.
- [6] S. Ahajjam, M. El Haddad, and H. Badir, “Influentials identification for community detection in complex networks,” *Colloq. Inf. Sci. Technol. Cist*, pp. 111–115, 2017.
- [7] H. Sarvari, E. Abozinadah, A. Mbaziira, and D. Mccoy, “Constructing and Analyzing Criminal Networks,” *2014 IEEE Secur. Priv. Work.*, pp. 84–91, 2014.
- [8] R. Maddox, R. Davey, R. Lovett, A. Van Der Sterren, J. Corbett, and T. Cochrane, “A systematic review protocol: Social network analysis of tobacco use,” *Syst. Rev.*, vol. 3, no. 1, pp. 1–5, 2014.
- [9] D. Jarman, E. Theodoraki, H. Hall, and J. Ali-Knight, “Social network analysis and festival cities: an exploration of concepts, literature and methods,” *Int. J. Event Festiv. Manag.*, vol. 5, no. 3, pp. 311–322, 2014.
- [10] O. Dziadkowiec, S. Wituk, and D. Franklin, “A social network analysis of South Central Kansas Workforce Innovations in Regional Economic Development,” *J. Place Manag. Dev.*, vol. 8, no. 1, pp. 6–22, 2015.
- [11] B. K. Wichmann and L. Kaufmann, “Social network analysis in supply chain management research,” *Int. J. Phys. Distrib. Logist. Manag.*, vol. 46, no. 8, pp. 740–762, 2016.
- [12] A. Mangas-Vega, R. Gomez-Diaz, and J. A. Cordon-Garcia, “Approach to self-publishing with a combination of bibliometric study and social network analysis techniques,” *Electron. Libr.*, vol. 34, no. 6, pp. 902–914, 2016.
- [13] V. L. M. Schropfer, J. Tah, and E. Kurul, “Mapping the knowledge flow in sustainable construction project teams using social network analysis,” *Eng. Constr. Archit. Manag.*, vol. 24, no. 2, pp. 229–259, 2017.
- [14] B. J. Kim, H.-G. Park, and J.-B. Chung, “Social network analysis of the Korean disaster-safety industry,” *Disaster Prev. Manag.*, vol. 27, no. 1, pp. 28–42, 2018.
- [15] A. El Kassiri and F. Z. Belouadha, “Towards a unified semantic model for online social networks analysis and interoperability,” *2015 10th Int. Conf. Intell. Syst. Theor. Appl. SITA 2015*, 2015.
- [16] F. Bisio, C. Meda, R. Zunino, R. Surlinelli, E. Scillia, and A. Ottaviano, “Real-time monitoring of Twitter traffic by using semantic networks,” *Proc. 2015 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Min. 2015 - ASONAM '15*, pp. 966–969, 2015.
- [17] C. Fu, L. Chuang, Y. Jiahai, and X. Yuemei, “A study on online social networks theme semantic computing model,” *Proc. - 2016 IEEE Int. Conf. Web Serv. ICWS 2016*, pp. 244–251, 2016.
- [18] P. S. Raji and S. Surendran, “RDF approach on social network analysis,” *Int. Conf. Res. Adv. Integr. Navig. Syst. RAINS 2016*, pp. 6–9, 2016.
- [19] P. Seidler, R. Adderley, A. Badii, and M. Raffaelli, “MOSAIC: Criminal network analysis for multi-modal surveillance and decision support,” *ASONAM 2014 - Proc. 2014 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Min.*, no. Asonam, pp. 257–260, 2014.
- [20] G. Oatley and B. Ewart, “Data mining and crime analysis,” *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 1, no. 2, pp. 147–153, 2011.
- [21] K. Taha and P. D. Yoo, “A System for Analyzing Criminal Social Networks,” *Proc. 2015 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Min. 2015 - ASONAM '15*, pp. 1017–1023, 2015.
- [22] M. A. Tayebi, L. Bakker, U. Glässer, and V. Dabbaghian, “Locating Central Actors in Co-offending Networks,” *2011 Int. Conf. Adv. Soc. Networks Anal. Min.*, pp. 171–179, 2011.
- [23] C. Morselli, *Inside Criminal Networks*, vol. 8. 2009.
- [24] T. Dörfler, E. Stollenwerk, and J. Schibberges, *Uncovering Covert Structures: Social Network Analysis and Terrorist Organizations*. 2019.
- [25] G. S. Lee and M. A. Djauhari, “An Overall Centrality Measure: The Case of U . S Stock



- Market,” *Int. J. Basic Appl. Sci.*, vol. 12, no. 06, pp. 99–104, 2012.
- [26] R. Zafarani, M. A. Abbasi, and H. Liu, “Social Media Mining An Introduction,” *Cambridge Univ. Press*, p. 382, 2014.
- [27] M. Naderifar, H. Goli, and F. Ghaljaie, “Snowball Sampling: A Purposeful Method of Sampling in Qualitative Research,” *Strides Dev. Med. Educ.*, vol. 14, no. 3, Sep. 2017.
- [28] A. von der Fehr, J. Sølberg, and J. Bruun, “Validation of networks derived from snowball sampling of municipal science education actors,” *Int. J. Res. Method Educ.*, 2018.
- [29] J. Illenberger, M. Kowald, K. W. Axhausen, and K. Nagel, “Insights into a spatially embedded social network from a large-scale snowball sample,” in *European Physical Journal B*, 2011.
- [30] R. Miller and J. Brewer, “Sampling, snowball: accessing hidden and hard-to-reach populations,” in *The A-Z of Social Research*, 2016.