# Application of Graph-Based Technique to Identity Resolution

Hassan Kazemian, Mohammad-Hossein Amirhosseini, Michael Phillips

# Application of Graph-Based Technique to Identity Resolution

Hassan Kazemian[1], Mohammad-Hossein Amirhosseini[2] and Michael Phillips[3]

[1] Intelligent Systems Research Centre
London Metropolitan University, London, N7 8DB, UK
h.kazemian@londonmet.ac.uk
[2] Intelligent Systems Research Group
University of East London, Docklands Campus, London, E16 2RD, UK
[3] Intelligent Systems Research Centre
London Metropolitan University, London, N7 8DB UK

**Abstract.** These days the ability to prove an individual identity is crucial in social, economic and legal aspects of life. Identity resolution is the process of semantic reconciliation that determines whether a single identity is the same when being described differently. The importance of identity resolution has been greatly felt these days in the world of online social networking where personal details can be fabricated or manipulated easily. In this research a new graph-based approach has been used for identity resolution, which tries to resolve an identity based on the similarity of attribute values which are related to different identities in a dataset. Graph analysis techniques such as centrality measurement and community detection have been used in this approach. Moreover, a new identity model has been used for the first time. This method has been tested on SPIRIT policing dataset, which is an anonymized dataset used in SPIRIT project funded by European Union's Horizon 2020. There are 892 identity records in this dataset and two of them are 'known' identities who are using two different names, but they are both belonging to the same person. These two identities were recognized successfully after using the presented method in this paper. This method can assist police forces in their investigation process to find criminals and those who committed a fraud. It can also be useful in other fields such as finance and banking, marketing or customer service.

**Keywords:** Identity Resolution, Identity Model, Graph Analysis, Community Detection, Centrality Measurement.

## 1 Introduction

### 1.1 Importance of Identity Resolution

Identity can be described as a set of identifiable characteristics that can distinguish one individual from another. Nowadays electronic records are replacing paper-based documents and identity records can be generated easily. Therefore, duplicate and false

identity records are quite common in electronic systems and databases because of lack of sufficient verification or validation during data entry processes [1]. In this situation, finding an effective solution to address this issue is extremely critical and it can facilitate fighting crime, terrorism or enforce national security. Li and Wang [1] pointed out that criminals and terrorists try to hide their true identity via using fake identities. There are some cases documented by government reports which are showing terrorists in different countries have committed different identity crimes such as falsifying passports or birth certificates to facilitate their travelling or their financial operations [2,3]. The problem of multiple identities for an individual can mislead police and law enforcement investigators [4]. Identity resolution is a pathway to tackle problems when it becomes intensely difficult to determine if the resultant identity is same when criminals describe it differently.

### 1.2 Identity Model

There should be a clear identity model before starting identity resolution process. Based on the identity theories from the social science literature, an individual's identity is considered to have two basic components, namely a personal identity and a social identity. A personal identity is one's self-perception as an individual, whereas a social identity is one's biographical history that builds up over time [5]. These two aspects of identity have been considered by researchers for identity resolution. But the previous identity models have been suffering from some limitations which can affect on the accuracy of the results. In fact, individuals are not isolated but interconnected to each another in a society. The social context associated with an individual can be clues that reveal his or her undeniable identity. Recognizing the limitations of personal attributes, many recent studies have started exploiting social context information for identity resolution.

For instance, Ananthakrishna et al. [6] introduced a method that eliminates duplicates in data warehouses using a dimensional hierarchy over the link relations. This method can improve the performance of the matching technique by only comparing those attribute values that have the same foreign key dependency. For instance, the similarity of two identity will be analyzed only when both of them live in the same city. Afterwards, Kalashnikov et al. [7] combined co-affiliation and co-authorship relationships and created a new resolution model for reference disambiguation. In another research, Köpcke and Rahm [8] categorized entity resolution methods into context matchers and attribute value matchers. They explain that attribute value matchers rely on descriptive attributes, while context matchers consider information inferred from social interactions which is represented as linkages in a graph.

A new identity model for identity resolution has been used in this research. In addition to physical and social aspects of an identity, this new identity model is considering two more aspects of an identity which results in considering more attributes and can improve the accuracy and reliability of the identity resolution. This will be explained in methodology section.

### 1.3 Identity resolution methods

Existing identity resolution methods can be categorized into two groups which are (1) rule-based and (2) machine learning methods. Most of the rule-based identity resolution methods have been developed based on the matching rules. As an example, for a simple rule, two identity records match only if their first name, surname, and date of birth values are identical [9]. Li and Wang [1] explained that matching rules try to have high precision, but they usually suffer from low sensitivity in detecting true matches. This is because of data quality issues such as missing data, entry error and deceptions. They also discussed that the most important challenge for a rule-based method can be creation of the rule set because creating an effective and comprehensive rule set can be very complicated and time consuming and the rules may not be portable and applicable across different contexts.

Creating a comprehensive rule sets can be highly time consuming and expensive. Another issue might be portability as some of rules could be dependent to a specific domain and not portable across different domains. In this situation, machine learning can be considered as an alternative approach to manual rule coding because it can automatically recognize patterns in training data with matching pairs. This will help to build a resolution model for new identity records. Li and Wang [1] explained that when there is a pair of identity records, distance measures can be defined for different descriptive attributes and then they can be combined into an overall score. The overall distance score will be compared to a pre-defined threshold and the pair should be considered as a match if this score is below or above the threshold.

One of the first identity resolution methods was a data association method for linking criminal records that possibly refer to the same suspect [10]. This method was comparing two different records and calculating an overall distance measure as a weighted sum of the distance measures of all corresponding feature values. In another attempt, Wang et al. [11] proposed a record linkage method which was detecting misleading identities by comparing four attributes and combining them into an overall distance score. These attributes were (1) name, (2) date of birth, (3) social security number, and (4) address. They used a supervised learning method to determine a threshold for match decisions. This was done via using a set of identity pairs which were labelled by an expert. Wang et al. [12] discussed that these methods perform based on a limited number of descriptive attributes and the most important issue in this case is that they tend to fail if one or more of considered attributes contains missing values.

In another study, a graph-based method for entity resolution was proposed by Bhattacharya and Getoor [13]. This new method defined a distance measure that combined graph-based relational similarity with corresponding attribute similarities between each entity reference pair. They extended this approach later and proposed a collective entity resolution method. As a result, instead of simply making pair-wise entity comparisons, they could derive new social information and incorporate it into further resolution process repeatedly.

There were some other researchers tackled the issue of one person having several profiles on different social media platforms and some techniques for matching user profiles have been developed. For instance, a CRF-based approach was proposed by

Bartunov et al. [14]. They used two user graphs which were created using both user profile attributes and social linkages and then they combined these two graphs. These researchers have successfully demonstrated that social information can help to improve the performance of identity resolution, when incorporated in matching algorithms.

## 2 Methodology

### 2.1 SPIRIT Policing Dataset

SPIRIT policing dataset is an anonymized dataset which has been used in SPIRIT project funded by European Union's Horizon 2020. This dataset includes 891 identities, and each identity has 30 different attributes. Eight of these attributes will be considered in this research which are (1) postcode, (2) date of birth, (3) town, (4) offence, (5) gender, (6) street name, (7) district, and (8) ethnicity. There are two 'known' identities in this dataset who are using two different names 'Billy Smith' and 'Mariet Snehh' but they are both belonging to the same person.

### 2.2 Identity Model

Identity refers to those attributes that enable us to recognize an individual from others. A new identity model has been used in this research for the first time, which includes four categories of attributes. The first category is physical identity which includes characteristics which an object or person is definitively recognizable or known by. The second category is official identity which is the identity that carries a legal status, usually issued by governments to their citizens. The third category is virtual identity which is the identity created by human user that acts as an interface between physical person and virtual person that other users see on their computer screen. It is a model for self expression, and tools for virtual interaction and a representation of a user in a virtual world. Finally, the fourth category is social identity which is a set of behavioral or personal characteristics by which an individual is recognizable as a member of a group.

### 2.3 Graph Creation

Eight graphs will be created after selecting 8 highly valued attributes which were mentioned in the previous section. In this case, for instance if 4 identities have the same postcode, the graph shows that these 4 identities are connected to each other. In these graphs, the nodes will be presenting a person with his/her first name and surname, and each edge will be showing that there is a similarity between two nodes (person).

### 2.4 Community Detection Algorithm

After graph creation step, the Louvain algorithm will be used for community detection based on the selected attributes. This method is a very efficient method for identifying communities in large networks. Blondel et al. [15] mentioned that the Louvain

method has been used successfully for analyzing different type of networks and for sizes up to 100 million nodes and billions of links. They also pointed out that the analysis of a typical network of 2 million nodes takes 2 minutes on a standard PC. In fact, this method is a greedy optimization method which tries to optimize the modularity of a partition of the network [15]. Modularity is a metric that can be used to quantify the quality of an assignment of nodes to communities. In other words, modularity can be defined as a value between −1 and 1 that measures the density of links inside communities compared to links between communities [16]. For a weighted graph, modularity is defined as:

$$M = \frac{1}{2k}\sum_{xy}\left[Q_{xy} - \frac{p_x p_y}{2k}\right]\delta(n_x, n_y) \quad (1)$$

In this equation, $Q_{xy}$ represents the edge weight between nodes $x$ and $y$. $p_x$ and $p_y$ are the sum of the weights of the edges attached to nodes $x$ and $y$. $k$ is the sum of all of the edge weights in the graph. $n_x$ and $n_y$ are the communities of the nodes and $\delta$ is a Kronecker delta which is a function of tow variables. It is 1 if the variables are equal and it is 0 if the variables are not equal. Equation 2 explains this.

$$\delta_{wt}\begin{cases}0 & xf\ w \neq t\\ 1 & xf\ w = t\end{cases} \quad (2)$$

In the Louvain algorithm, optimization will be performed in two steps. In the first step, small communities will be found by optimizing modularity locally. Then in the second step, the nodes which are belonging to the same community will be cumulated and a new network will be built where its nodes are the communities. These steps will be repeated until a maximum of modularity is achieved and a hierarchy of communities is produced [15]. In other words, in the first step, each node in the network will be assigned to its own community. Then for each node $x$, the change in modularity will be calculated for removing node $x$ from its own community and moving it into the community of each neighbor $y$ of $x$. Equation (3) explains the process of inserting $x$ to the community of $y$.

$$\Delta M = \left[\frac{\sum_{in} + 2p_{x,in}}{2k} - \left(\frac{\sum_{tot} + p_x}{2k}\right)^2\right] - \left[\frac{\sum_{in}}{2k} - \left(\frac{\sum_{tot}}{2k}\right)^2 - \left(\frac{p_x}{2k}\right)^2\right] \quad (3)$$

In this equation, $\sum_{in}$ is the sum of all the weights of the links inside the community that node $x$ is moving into. $\sum_{tot}$ is the sum of all the weights of the links to nodes in the community that node $x$ is moving into. Moreover, the weighted degree of node $x$ is represented by $p_x\ and$ the sum of the weights of the links between node $x$ and other nodes in the community that $x$ is moving into, is represented by $p_{x,in}$. Finally, $k$ is the sum of the weights of all links in the network. Table 1 shows some of the most important studies used Louvain method for community detection.

## 2.5 Investigating Potential Targets

One of the most important things in network analysis is finding the most important nodes in a graph. Newman [17] explained that 'centrality' is a term that can be used to describe importance of individual nodes in a graph and 'degree of a node' is the number of edges that it has. The nodes with more connections are more influential and important in a network. As a result, the person with more friends in a social graph, is the one that is more central. Thus, in the next step of our method, eight different lists of names will be provided based on the measurement of centrality and the degree of nodes in each graph. Top 20 nodes based on their degree (number of connections that they have) will be recorded in each list. Then these lists will be compared with each other to find the similar identities. If any identity is repeated in at least five lists, it will be recorded in a new list as a potential target. Following this step, a new list of all related identities to the potential targets will be provided.

**Table 1.** Louvain Method for Community Detection

| Project | Number of nodes | Source |
|---|---|---|
| Twitter social network | 2.4M | Divide and Conquer: Partitioning Online Social Networks [18]. |
| Mobile phone networks | 4M | Tracking the Evolution of Communities in Dynamic Social Networks [19]. |
| Flickr | 1.8M | Real World Routing Using Virtual World Information [20]. |
| LiveJournal | 5.3M | |
| YouTube | 1.1M | |
| Citation network | 6M | Subject clustering analysis based on ISI category classification [21]. |
| LinkedIn social network | 21M | Mapping search relevance to social networks [22]. |

## 2.6 Phonetic Algorithms

In the next step, a cascading method will be used for applying three phonetic algorithms on the potential targets and their relevant identities in order to detect any possible human errors during data entry or wrong information given by the person. Three phonetic algorithms have been implemented for indexing names by sound. They are (1) Soundex, (2) Metaphone, and (3) Jaro-Winkler that will be applied following each other on the potential targets and their relevant identities. This means that in the first cycle, Soundex method will be applied. Then in the second cycle Metaphone will be applied on the results of Soundex method. Finally, in the third cycle the Jaro-Winkler method will be applied on the results of Metaphone method. Thus, we narrow down the results to get the best output. As a result, all similar first names and surnames to the potential target identities and their relevant identities, with a potential of being manipulated will be detected to be considered in the next step.

### 2.7 Comparison Process

After applying Soundex, Metaphone and Jaro-Winkler algorithms using a cascading method, all potential targets and their relevant identities as well as similar identities with a potential of using manipulated forenames and surnames will be added to a new dataset for comparison purpose. All attributes of the potential targets and their relevant identities in the new dataset will be compared separately and similarity will be scored. This means that all 30 attributes for each one of these identities in SPIRIT policing dataset will be considered for comparison and scoring process. The same identities will be investigated based on the similarity scores.

## 3 Results and discussion

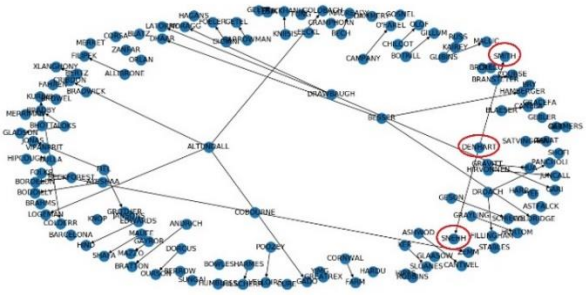

**Fig. 1.** Graph based on the same postcode

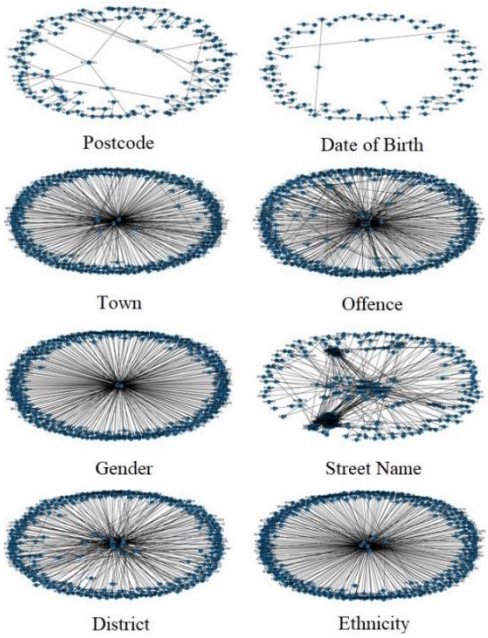


**Fig. 2.** Graphs for eight selected attributes

As it was explained in the methodology, eight attributes in SPIRIT policing dataset were selected and they are including (1) postcode, (2) date of birth, (3) town, (4) offence, (5) gender, (6) street name, (7) district, and (8) ethnicity. As a result, in the first step 8 graphs were created. Figure 1 shows one of these graphs, which was created based on the same postcodes and it shows that Billy Smith, Lorret Denhart and Mariet Snehh have been using the same postcode. Moreover, Figure 2 shows all eight created graphs.

In the second step, the Louvain algorithm was used for community detection based on the 8 selected attributes. Figure 3 shows the community detection graphs.

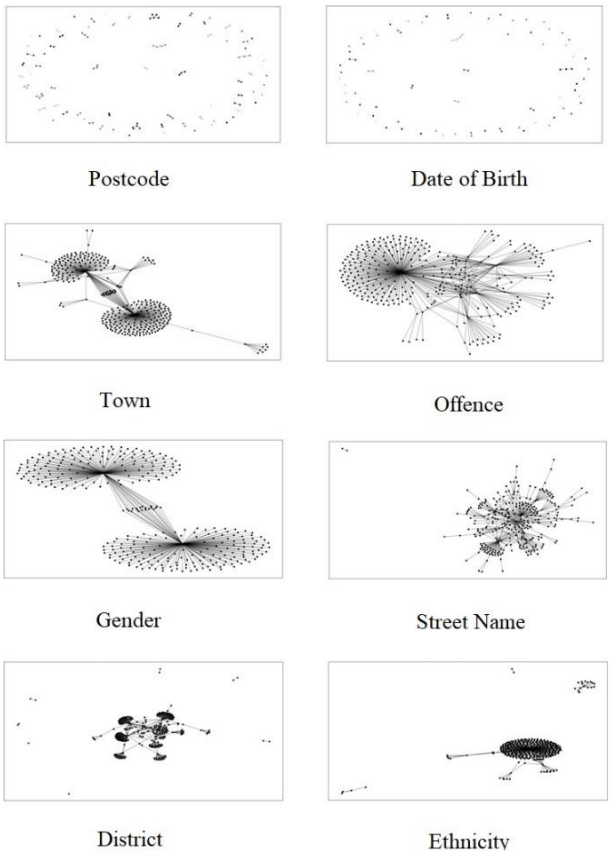


**Fig. 3.** Community detection graphs

Following this step, centrality and the degree of nodes in each graph were measured, and top 20 nodes in each graph were recorded in eight different lists. These lists were compared, and those identities which were repeated in at least 5 lists were recorded in a new list for potential targets. Table 3 shows the surnames which were repeated in at least five lists.

**Table 2.** Surnames which were repeated at least in 5 lists.

| Top 20 nodes in | | | | |
|---|---|---|---|---|
| **postcode graph** | **town graph** | **date of birth graph** | **offence graph** | **street name graph** |
| ALTUNDALL | AINSBURY | **SMITH** | AINSBURY | BESSER |
| DROACH | ADRE | VELLOIRS | ALTUNDALL | ADRE |
| FIEL | ARTHURS | BOWLES | AMENT | GLAASOW |
| GROTHER | BATISTE | DENHART | ALOKS | DICIANI |
| BECKFOREST | **BORDELON** | GLAASOW | VELLOIRS | BALOW |
| BESSER | BOSSERMAN | KER | ADRE | **SMITH** |
| BLAESER | DENHART | MERRET | **SNEHH** | CRECO |
| **BORDELON** | GAYROR | ANCHORRS | ASTFALCK | **SNEHH** |
| DICIANI | VELLOIRS | ASTFALCK | BECKFOREST | FETTES |
| DORCUS | **SMITH** | **SNEHH** | BALOW | BRAHMS |
| DRAWBAUGH | **SNEHH** | BARCELONA | **BORDELON** | CRECO |
| GAYROR | BECH | BECH | ALOI | DORCUS |
| GELTZ | GROTHER | BECKFOREST | CRECO | FAHREN |
| GETEL | FETTES | BERROW | FAHREN | KER |
| JUNCALL | BHOTT | BHOTT | KER | **BORDELON** |
| KER | BUTLAND | BLAESER | **SMITH** | GETEL |
| **SNEHH** | 'CORSA | BLATZ | GOSNEL | JUNCALL |
| KNIISIS | DEGNER | **BORDELON** | BARCELONA | GAYROR |
| POOZEY | DORCUS | BRADWICK | GAYROR | KNIISIS |
| **SMITH** | FONES | BRAHMS | FILLINGHAM | SATVINGRER |
| | | BROWEL | | |

According to Table 2, three surnames including Smith, Snehh and Bordleon were existed in at least five lists for top 20 nodes. These lists are related to (1) postcode graph, (2) town graph, (3) date of birth graph, (4) offence graph, and (5) street name graph. The relevant surnames to these surnames were detected in the next step based on their connections in different graphs and they were added to a new dataset after applying the phonetic algorithms. Table 3 shows these identities.

**Table 3.** Potential target names and their identities

| |
|---|
| Loret Denhart |
| Billy Smith |
| Nizie Bordelon |
| Mariet Snehh |
| Jasmalinne Beckforest |
| Kemp Bech |

This new dataset was included all 30 attributes for each identity. These 30 attributes were mentioned in section A of the methodology. Figure 4 shows a screenshot from a part of the new dataset, which is created after applying Soundex, Metaphone and Jaro-Winkler algorithms.

As figure 4 shows, some of the names in this dataset were repeated. The reason is that some of the values of their different attributes are different. For instance, there are two rows for Kemp Bech. This is because there are two different postcodes related to this name and this person was committed two different offences. As a result, there are two rows related to this person with different values for two attributes including postcode and offence. Finally, every single row of this dataset was compared with other rows and based on the number of attributes which had the same value, a score was assigned to show the similarity between each two identities. After comparing these scores, it was realized that two identities including Billy Smith and Mariet Snehh have the most

similarity. As it was explained in section A of the methodology, these two were 'known' identities in SPIRIT policing dataset who had two different names, but they were both belonging to the same person. Thus, the system was successful in resolve their identities.

| Surname | Forename | Datae_of_birth | Postcode | Town | Offence |
|---|---|---|---|---|---|
| BECH | KEMP | 25/02/2003 00:00 | B35 6DE | CARSINGTON | THEFT FROM MOTOR VEHICLE |
| BECH | KEMP | 25/02/2003 00:00 | B23 6PU | CARSINGTON | BURGLARY OTHER BUILDING |
| BECH | LGAH | 18/03/1945 00:00 | B75 7EN | TURNBURY | ATTEMPT BURGLARY DWELLING |
| BECKFOREST | JASMALINNE | 29/03/1989 00:00 | B44 0TD | CARSINGTON | CRIMINAL DAMAGE TO DWELLING |
| BECKFOREST | JASMALINNE | 29/03/1989 00:00 | B44 0TD | CARSINGTON | CRIMINAL DAMAGE TO DWELLING |
| BORDELON | NIZIE | 01/01/1989 00:00 | B44 0TD | CARSINGTON | ASSAULT OCCASION ABH |
| BORDELON | NIZIE | 01/01/1989 00:00 | B44 0TD | CARSINGTON | COMMON ASSAULT |
| BORDELON | NIZIE | 01/01/1989 00:00 | B23 7UP | CARSINGTON | HARASSMENT |
| BORDELON | NIZIE | 01/01/1989 00:00 | B23 7UP | CARSINGTON | HARASSMENT |
| BORDELON | NIZIE | 01/01/1989 00:00 | B23 7UP | CARSINGTON | PUTTING PEOPLE IN FEAR OF VIOLENCE |
| BORDELON | NIZIE | 01/01/1989 00:00 | B23 7UP | CARSINGTON | HARASSMENT |
| BORDELON | NIZIE | 01/01/1989 00:00 | B23 7UP | CARSINGTON | HARASSMENT |
| BORDELON | NIZIE | 01/01/1989 00:00 | B44 0TD | CARSINGTON | CRIMINAL DAMAGE TO DWELLING |
| BORDELON | NIZIE | 01/01/1989 00:00 | B44 0TD | CARSINGTON | COMMON ASSAULT |
| BORDELON | NIZIE | 01/01/1989 00:00 | B44 0TD | CARSINGTON | CRIMINAL DAMAGE TO DWELLING |
| BORDELON | NIZIE | 01/01/1989 00:00 | B44 0TD | CARSINGTON | OTHER CRIMINAL DAMAGE |
| BORDELON | NIZIE | 01/01/1989 00:00 | B44 0TD | CARSINGTON | BURGLARY DWELLING |
| BORDELON | NIZIE | 01/01/1989 00:00 | B44 0TD | CARSINGTON | THEFT DWELLING NOT MACHINE/METER |
| BORDELON | NIZIE | 01/01/1989 00:00 | B44 0TD | CARSINGTON | OTHER CRIMINAL DAMAGE |
| BORDELON | NIZIE | 01/01/1989 00:00 | B44 0TD | CARSINGTON | CRIMINAL DAMAGE TO DWELLING |
| BORDELON | NIZIE | 01/01/1989 00:00 | B43 7BX | YARNFORTH | CRIMINAL DAMAGE TO DWELLING |
| BORDELON | NIZIE | 01/01/1989 00:00 | B44 0TD | CARSINGTON | CRIMINAL DAMAGE TO DWELLING |
| BORDELON | NIZIE | 01/01/1989 00:00 | B44 0TD | CARSINGTON | BURGLARY DWELLING |
| BORDELON | NIZIE | 01/01/1989 00:00 | B43 7BW | YARNFORTH | BURGLARY DWELLING |
| DENHART | LORRET | 01/06/1994 00:00 | SF19 9NF | CAERLEON | BURGLARY DWELLING |
| DENHART | LORRET | 01/06/1994 00:00 | SF19 9NF | CAERLEON | BURGLARY DWELLING |
| DENHART | LORRET | 01/06/1994 00:00 | B18 4AS | CARSINGTON | BURGLARY DWELLING |

**Fig. 4.** New dataset including all potential targets and their relevant identities to be used for comparison process.

## 4 Conclusion

This research introduces a new graph-based approach for identity resolution. In this approach, graph analysis techniques such as community detection and centrality measurement have been used. Furthermore, this research introduces a new identity model which represents four different types of attributes including (1) physical attributes, (2) social attributes, (3) official attributes, and (4) virtual attributes. SPIRIT policing dataset was used for testing this method. This dataset is an anonymized dataset which has been used in SPIRIT project funded by European Union's Horizon 2020 and includes 892 identity records. Two of these identities are 'known' identities which both are belonging to the same person, but they are using two different names. The methodology presented in this paper successfully recognized these two identities in SPIRIT policing dataset and the expected results were exactly the same as actual results. This identity resolution approach can effectively facilitate the investigation process for police forces and assist them to find criminals and individuals who committed a fraud. It can also be useful for other similar datasets which are containing identity records related to other fields such as finance and banking, customer service or marketing.

**Acknowledgement**

This research was supported by European Union's Horizon 2020, grant no: 786993. We thank our colleagues from SPIRIT project consortium who provided insight and expertise that greatly assisted the research.

## References


[1] J. Li, A.G. Wang, A framework of identity resolution: evaluating identity attributes and matching algorithms, *Security Informatics*, vol. 4, no. 6, 2015, [online] Available: https://doi.org/10.1186/s13388-015-0021-0

[2] T.H. Kean, C.A. Kojm, P. Zelikow, J.R. Thompson, S. Gorton, T.J. Roemer, J.S. Gorelick, J.F. Lehman, F.F. Fielding, B. Kerrey, The 9/11 Commission Report, 2004, [online] Available: http://govinfo.library.unt.edu/911/report/index.htm

[3] U.S. Department of State: Country Reports on Terrorism, 2006, [online] Available: http://www.state.gov/j/ct/rls/crt/2006/

[4] J. Li, G.A. Wang, H. Chen, Identity matching using personal and social identity features, *Information Systems Frontiers*, vol. 13, pp. 101-113, 2010.

[5] J.M. Cheek, S.R. Briggs, Self-consciousness and aspects of identity, *Journal of Research in Personality*, vol. 16, pp. 401-408, 1982.

[6] R. Ananthakrishna, S. Chaudhuri, V. Ganti, Eliminating Fuzzy Duplicates in Data Warehouses. In *Procceeding of 28th International Conference on Very Large Data Bases*, Hong Kong, China, pp. 586-597, 2002.

[7] D. V. Kalashnikov, S. Mehrotra, Z. Chen, Exploiting relationships for domain-independent data cleaning. In *Proceeding of 2005 SIAM International Conference on Data Mining*. Newport Beach, CA, pp. 262-273, 2005.

[8] H. Köpcke, E. Rahm, Frameworks for entity matching: a comparison, *Data and Knowledge Engineering*, vol. 69, pp. 197-210, 2010.

[9] B. Marshall, S. Kaza, J. Xu, H. Atabakhsh, T. Petersen, C. Violette, H. Chen, Cross-Jurisdictional Criminal Activity Networks to Support Border and Transportation Security. In *Proceedings 7th Int IEEE Conference Intelligent Transportation Systems*, Washington, D.C., pp. 100-105, 2004.

[10] D.E. Brown, S.C. Hagen, Data association methods with applications to law enforcement, *Decision Support Systems*, vol. 34, pp. 369-378, 2003.

[11] G.A. Wang, H. Chen, H. Atabakhsh, Automatically detecting deceptive criminal identities. *Communications of the ACM*, vol. 47, pp. 70-76, 2004.

[12] G.A. Wang, H.C. Chen, J.J. Xu, H. Atabakhsh, Automatically detecting criminal identity deception: an adaptive detection algorithm. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems Humans*, Vol. 36, pp. 988-999, 2006.

[13] I. Bhattacharya, L. Getoor, Entity resolution in graphs, in Min graph data, Wiley-Blackwell, Hoboken, 2006.

[14] S. Bartunov, A. Korshunov, S. Park, W. Ryu, H. Lee, Joint Link-Attribute User Identity Resolution in Online Social Networks. In *Proceeding of the 6th International Conference on Knowledge Discovery and Data Mining, Workshop on Social Network Mining and Analysis*, Beijing, China, 2012.

[15] V.D. Blondel, J.L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, *Journal of Statistical Mechanics: Theory and Experiment.* Vol. 10, 2008.

[16] F. Hua, Z. Fang, T. Qiu, Modeling Ethylene Cracking Process by Learning Convolutional Neural Networks, *Computer aided chemical engineering*, vol. 44, pp. 841-846, 2018.

[17] M. Newman, *Networks: An Introduction*, Chapter 7: Measures and Metrics, Oxford University Press, pp. 168-234, 2010.

[18] J.M. Pujol, V. Erramilli, P. Rodriguez, Divide and Conquer: Partitioning Online Social Networks, 2010.

[19] D. Greene, D. Doyle, P. Cunningham, Tracking the Evolution of Communities in Dynamic Social Networks, *International Conference on Advances in Social Networks Analysis and Mining*, 2010.

[20] P. Hui, N.R. Sastry, Real World Routing Using Virtual World Information, *International Conference on Computational Science and Engineering*, 2009

[21] L. Zhang, X. Liu, F. Janssens, L. Liang, W. Glänzel, Subject clustering analysis based on ISI category classification, *Journal of Informatics*, vol. 4, no. 2, 2010.

[22] J. Haynes, I. Perisic, Mapping search relevance to social networks, *Proceedings of the 3rd Workshop on Social Network Mining and Analysis*, 2010.