ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

CISRI - Crime Investigation System Using Relative Importance: A Survey

Nilam Prakash Sonawale[1], B. W. Balkhande[2]

Student of M.E. Computer Engineering, Bharati Vidyapeeth College of Engineering, Mumbai University, Navi Mumbai, Maharashtra, India[1]

Senior Consultant, Infostep India Pvt. Ltd., Client, Reserve Bank Of India , Mumbai, India[1]

Assistant Professor, Dept. of Computer Engineering, Bharati Vidyapeeth College of Engineering, Mumbai University, Navi Mumbai, Maharashtra, India[2]

ABSTRACT: CISRI forensic analysis systems that helps forensic investigation to determine the most influential member of a criminal group who are related to the known member of group for the purpose of investigation. It describe the structure relationships between the members of a criminal group in terms of graph. In a graph a node represents a member of criminal group , an edge represents relationship between two members and weight of an edge represents degree of relationship between two members. A Graph Constructed from Mobile Communication Data (MCD).
CISRI Overcomes limitation by Computing the contribution of the overall set of query nodes to the importance of node under observation and doing a tight constraint calculation that checks how much each query node contributes to the relative importance of node under consideration.[5]

KEYWORDS: Mobile Communication Data (MCD); Relative Importance(RI); Graph Of Interest (GOI);Social Network Analysis (SNA); Criminal Network Analysis (CNA); Number Of Hubs (NOH)

I. INTRODUCTION

Cloud Digital Forensics has always been an evolving field of research due to the constantly changing devices and technologies. Forensic department have spent great effort and time in analysing and studying new techniques and systems. Electronic devices like Smartphone's, Tablets store huge amount of information as a data which can be use to analyse the involvement of the owners in the crime events. Digital Forensic facing most serious concerns like time and effort to analyse the huge amount of data which can complicates the process of identifying relevant data.

Mobile Communication Data (MCD) which includes data of phone calls, messages and locations which can collected from mobile devices or from the mobile network providers. MCD helps for lead to research, analyse and deduce useful patterns. Patterns from MCD permits Digital Forensic department to understand human behaviour and domain of interest to derive useful conclusion. MCD has also taken a vital role to find the dynamics of criminal networks. To analyse how the members are related , how the information flows on the network and which members have essential roles. It requires information needs to be represents in unified structural schema. One of the most common schemas for representing networks is Graphs. A graph consists of a set of nodes that are linked by a set of edges. In MCD node represents contact and edge communication between contacts via call, message or any way of communication.

Current approaches of relative importance determine the relative importance of a node under consideration by estimating the contribution of each query node individually which results in low precision. CISRI overcomes this limitation by computing the contribution of the overall set of query nodes to the importance of a node under consideration, and adopting a tight constraint calculation that considers how much each query node contributes to the relative importance of a node under consideration.

CISRI provides system which analyses criminal networks and determines the relative importance of their members with other members. CISRI provides tight constraint calculation of relative importance which ensure accurate identification of members in criminal network.[5]

II. LITERATURE SURVEY

A branch of forensic science, i.e. Digital forensics includes the recovery and investigation of data found in digital devices, which are in relation to computer crime. The term digital forensics has expanded to cover investigation of all devices efficient of storing digital data. The technical feature of an investigation is divided into several sub-branches, relating to the type of digital devices involved; computer forensics, network forensics, forensic data analysis and mobile device forensics. Forensic process includes the seizure, forensic imaging and analysis of digital media and the production of a report into collected proof. As well as identifying direct proof of a crime, digital forensics can be used to attribute proof to specific suspects, confirm alibis or statements, determine intent, identify sources (for example, in copyright cases), or authenticate documents. There are four stages of forensics Processes :- Identification of Digital Evidence, Preservation of Digital Evidence, Analysis of Digital Evidence , Presentation of Digital Evidence.

CISRI is closely related to Social Network Analysis (SNA). Centrality measure have been used in SNA which is used to find the members of the network based on the importance and role.(Degree Centrality, Betweenness Centrality, Closeness Centrality) [2].

Digital forensics does Criminal Network Analysis (CNA) by some monitoring systems like LogAnalysis, Blockmodeling, k-clique. [3]

Relative Importance Analysis measure how central a node is with respect to other nodes in graph.

SNA	CNA	RIA
Centrality measures have been used intensively in SNA.	CNA tends to have diverse shapes, sizes & goals. 1) Highly Structured 2) Loosely Structured	Can be done using tech known as relative centrality measure.
Measures used are 1) Degree Centrality 2) Betweenness Centrality 3) Closeness Centrality	Methods : 1) Log Analysis 2) Clique	Methods : 1) PageRank 2) Hits
Community detection is also a part of SNA	It also uses shortest path algorithm and Block modeling	Probabilities are used to calculate RI
Example of social structures commonly visualized through social network analysis include social media networks, friendship and acquaintance network	Identification of core members	To Rank important web pages according to their importance.

Limitation :

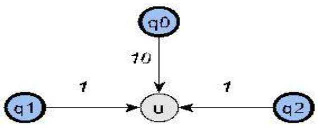
Incomplete Contribution of the Query Nodes : Query nodes does not contributes any value to the relative importance then problem arise.

Fig. 1. Illustration of the incorrect relative importance nodes, where one or more query node contributes 0 to the relative importance of a node (as in b).

Total relative importance of node u reflects it's importance to node q0 and q1 which is incorrect.

Incorrect Relative Importance Nodes : Due to incomplete contribution of a qury node this problem arises. In above diagram when we apply equation will result in ranking node v higher than u. However node v should not be ranked as it is only linked with q0 and q1.

Inconsistent Contribution of the Query Nodes : When the query nodes does not contribute similarly to the total relative importance of a node.



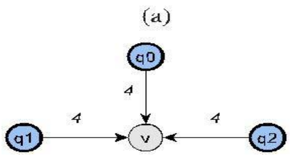


Fig. 2. Illustration of the inconsistent contribution of query nodes, where the query nodes do not contribute equally to the relative importance of a node (as in a).

Both nodes u and v have equal relative importance values (12) , the relative importance of node u is more biased to the query node q0. With node v all query nodes have contributed equally.

CISRI Overcomes Limitations :

By using tight constraint calculation CISRI overcomes on limitations. Logical And operator will be used in programming. CISRI overcomes the inconsistent contribution limitation by expected contribution from each query is multiplied by 1/ | Q | .

| Q | denotes number of nodes.

if Q = { q0 , q1 , q2 }

then each query node is expected to contribute 1/3 of the total relative importance of a node.

III. METHODOLOGIES

System Model :

1. Mobile Communication Data
2. Mobile Communication Data (MCD) Graph
3. Extracting Graph Of Interest (GOI)
4. Analyse Each Node in GOI
5. Identifying Relative Importance of each Node
6. Calculating Barycentric Distance to Overcome Inconsistent Contribution Problem
7. Average Number of Edges
8. Normalizing the Factor to assign a value in the range [0,1]
9. Ranking Node According to their Relative Importance

1. Mobile Communication Data :

Data than can be collected from mobile devices like phone call records, messages which describes patterns of communication between entities.

2. Mobile Communication Data (MCD) Graph :

To identify person who involved in an incident, Represent the MCD data network in unified structural schema i.e., Graph. Consist set of nodes that are linked by set of edges, Where node is contact and edge is communication between contacts.

3. Extracting Graph of Interest :

MCD graphs are large in size hence CISRI extracts in to sub graph called as Graph Of Interest (GOI). CISRI constructs the GOI by identifying the nodes that have significant relationship with the query nodes. GOI excludes redundant nodes that cannot be within the top k nodes, Which enhances the computational performance and time complexity.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

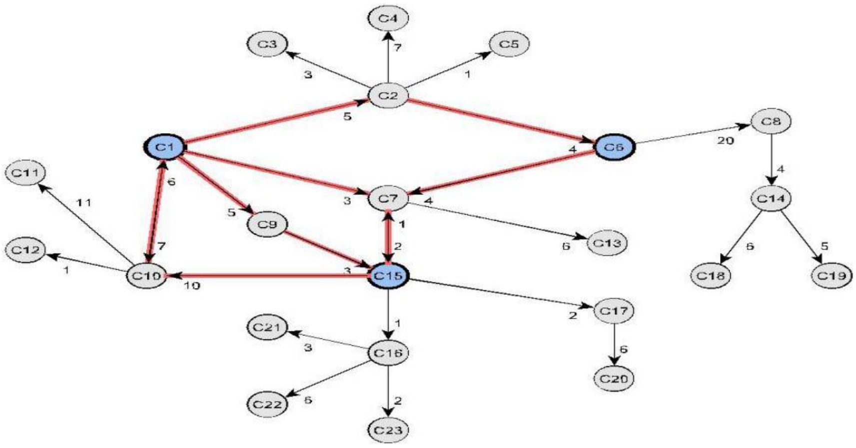


Fig. 3. Directed, weighted MCD graph where nodes represent contacts, edges represent the communication between contacts with weights relative to the amount of communication attempts. Nodes with bold borders are query nodes. weighted MCD graph where nodes represent contacts, edges represent the communication between contacts with weights relative to the amount of communication attempts. Nodes with bold borders are query nodes.

3.1. Algorithm

```
Algorithm  : Computing GOI (iterGraph)
Input: G = (Sm , E, R, λG), Q, k
Output : GOI
function IterGraph (G,Q,k)
GOI = Graph()                    > Intiliaze a New Graph
Sx[ ] = { }
AM=AdjacencyMatrix(G))
TAM= AM | AMT
if (|Q| > 1) then
   for (q1,q2) in getPairs(Q)
    paths[ ] = find ALL Paths(q1,q2, TAM) >
    Find all Paths between node q1 and q2 from the
    Adjacency matrix TAM.
    for p in paths
      Sx.append(getPathNodes(p))        > No
      Duplication of nodes.
    end for
   end for
end if
N O H = 0
while (|SX| < k) in
 t[ ]={ }
 N O H = N O H +1
 for node in Sx
   t.append(getNeighbours(node, N O H)
 end for
 Sx = Sx + t                    > Merge the two lists
end while
Sg= Sx
addNodesToGraph(G O I, Sg)          >Add all nodes in
```

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

List Sg to graph G O I and all corresponding edges.
25:**return** G O I
26:**end function**

The resulted GOI generated by the iterGraph algorithm are as below :

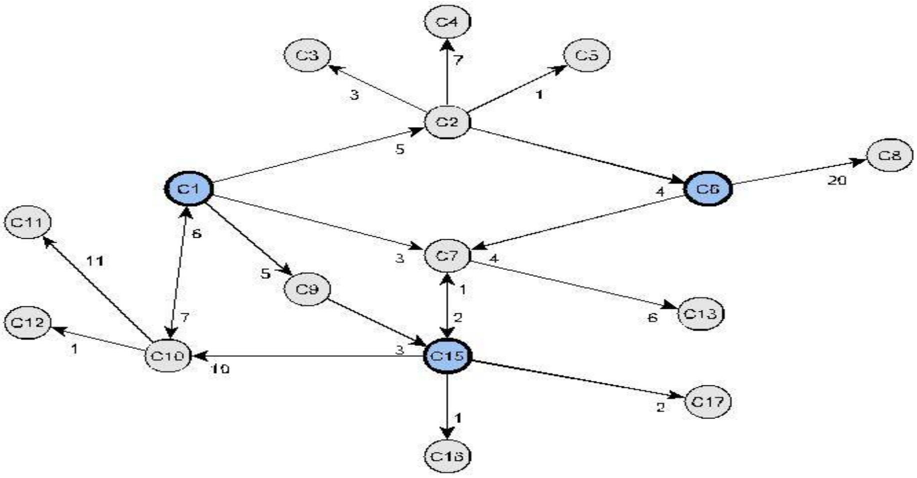


Fig. 4. The resulted GOI generated by the iterGraph algorithm.

4. Analyse each Node in GOI :
CISRI analyses each node in the GOI individually, Importance of each node is measured with respect to all query nodes.

5. Identify the Relative Importance of Each Node :
Individual relative importance iRI is a value which reflects the relation between the query node and node under consideration. We can show as iRI(u|q) is the relation between the node u and q. If a node does not have direct edge to anyone of the query nodes , the node will be eliminated. IF iRI(u|q)=0 then q will be eliminated.
Total relative importance is the overall relative importance of a node (tRI) . It is accumulative relationship that have been shared between each query node and node under consideration.
tRI(u|q) - Total relative importance of node u with respect to set of query nodes Q.
Illustration of the effect of the fractional penalty on the iRI Calculation :

$$tRI(u|Q) = \sum_{i=0, q1\in Q}^{|q|} \mathrm{iRI}(u|qi)$$

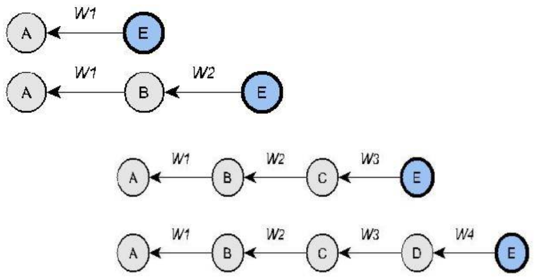


Fig. 5. Illustration of the effect of the fractional penalty on the iRI calculation. It illustrates, along with Table I, the exponential behavior that the fractional exponent penalty exhibits. It results in giving a higher iRI to nodes

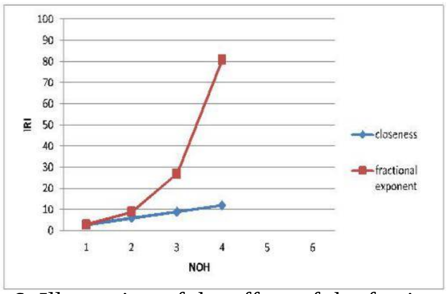


Fig. 6. Illustration of the effect of the factional exponent penalty and the closeness centrality penalty. The exponent increase results in a higher iRI for nodes with less NOH than nodes with high NOH.

The iRI with penalty is be calculated using following formula :

$$iRI(u|Q) = \left(\sum_{i=0}^{n-2} w(\text{ui}, \text{ui}+1)\right)^{1/n-1}$$

6. Calculating Barycentric Distance :

Optimal Contribution overcomes the problem that arises when each iRI contributes equally to the overall relative importance of a node. Optimal contribution situation occurs when each iRI contributes 1/|q| of the tRI this factor is called Barycentric Distance.

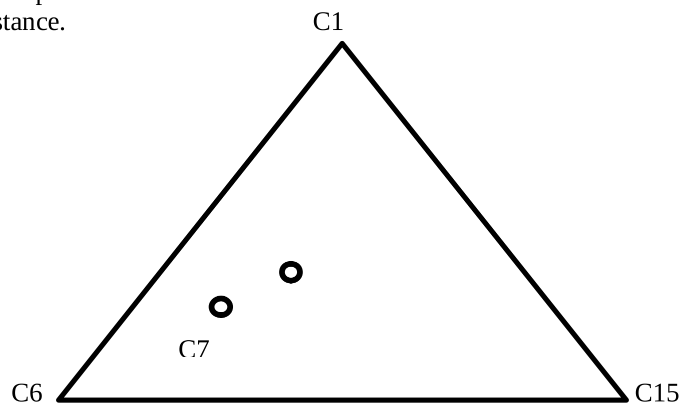


Fig. 7. Representing node C7 in the Barycentric Coordinate system where the vertices of the triangle correspond to the query nodes. The barycenter is represented as a blue vertex in the middle of the triangle. The distance between C7 and the barycenter is the barycentric distance

Contribution Of(q) = iRIq / t RI

Euclidean point (x,y) can be calculated as :

$$x = a \times P1_x + b \times P2_x + c \times P3_x + \ldots + k \times Pn_x$$
$$y = a \times P1_y + b \times P2_y + c \times P3_y + \ldots + k \times Pn_y$$

The Barycentric coordinate system, where a point in the system is represented by coordinates with reference to vertices of a simplex (triangle in this case). The barycenter is a point when all vertices contribute the same to the system.

7. Average Number of Edges :

It is essential to distinguish between an i RI calculated with long paths between a node and a query node, and an i RI calculated with short paths between a node and a query node.

If i RI (u|q) = 2 where only one edge separates node u from node q, then this node should have a higher ranking than

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

a node i RI (v |q) = 2 where 3 edges separate node v from node q.

8. Normalizing the Factor :

Total relative importance, Barycentric Distance and average number of edges are normalized in such a way that each factor is assigned a value in the range [0,1].

9. Ranking Node According to their Relative Importance :

Once the factors are normalized a generic equation is used to rank all nodes in the GOI. If the list of ranked nodes exceeds the requested k number of nodes, then the first k nodes will be selected from the ordered list.

IV. CONCLUSION

This CISRI system helps us to analyse the criminal networks by using the relative importance of their members. CISRI system uses tight constraint calculation which ensures accurate identification. Compare to other systems CISRI takes less time and efforts find the accurate results.

REFERENCES

[1]. F. Calabrese, M. Colonna, P. Lovisolo, D. Parata, and C. Ratti, "Realtime urban monitoring using cell phones: A case study in rome," IEEE Trans. Intell. Transp. Syst., vol. 12, no. 1, pp. 141–151, Mar. 2011.

[2]. L. C. Freeman, "Centrality in social networks conceptual clarification," Social Netw., vol. 1, no. 3, pp. 215–239, 1979.

[3]. E. Ferrara, P. De Meo, S. Catanese, and G. Fiumara, "Detecting criminal organizations in mobile phone networks," Expert Syst. Appl., vol. 41, no. 13, pp. 5733–5750, 2014.

[4]. S. P. Borgatti, "Identifying sets of key players in a social network," Comput. Math. Org. Theory, vol. 12, no. 1, pp. 21–34, Apr. 2006.

[5]. H. Wang, C. K. Chang, H.-I. Yang, and Y. Chen, "Estimating the relative importance of nodes in social networks," J. Inf. Process., vol. 21, no. 3, pp. 414–422, 2013.

BIOGRAPHY

Nilam Prakash Sonawale is Student of M.E. Computer Engineering, Bharati Vidyapeeth College of Engineering, Mumbai University, Navi Mumbai, Maharashtra, India. She is an Senior Consultant at Infotep India Pvt. Ltd.,
Her area of interest is Digital Forensic.

B.W.Balkhande is a Professor in the Department of Computer Engineering, Bharati Vidyapeeth College of Engineering, Mumbai University, Navi Mumbai, Maharashtra, India. Her area of interest is Digital Forensic, algorithms and programming.