IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 13, Issue 2 Ver. V (Mar. - Apr. 2017), PP 47-51
www.iosrjournals.org

Applying Graph Theory to Modeling Investigations

Chuck Easttom[1]
[1](*Collin College United States)*

Abstract: *This paper presents a methodology for applying the elements of graph theory to modeling forensic investigations. This methodology uses well established principles of graph theory to model any forensic investigation and thus mathematically evaluate the elements of a case, including the probabilities associated with specific suspects*

Keywords: *Forensics, Graph Theory, Mathematical Modeling*

I. Introduction

The purpose of this paper is to fully describe how graph theory can be applied to mathematically model a forensic investigation. The goal is to use the well-established principles of graph theory to mathematically model the elements in an investigation in order to provide mathematical clarity for investigations. This paper is an expansion of the paper Applying Graph Theory to Evidence Evaluation[1] This paper significantly expands the mathematics introduced in the earlier paper, as well introduce new applications of graph theory to modelling forensic investigations. That paper provided a general overview of graph theory; this paper assumes the reader already understands fundamental graph theory. For those readers who might need a refresher in elementary graph theory you can refer to the aforementioned paper, or to Balakrishnan[2].

II. Review Of Literature

Graph theory is a robust tool for examining relationships between any set of objects. The essentials of graph theory are relatively easy to grasp. Put formally: A finite graph G (V, E) is a pair (V, E), where V is a finite set and E is a binary relation on V[3]. Graph theory has been previously applied to evaluating network traffic[4,5,6,7]. In that context, graph theory was used to evaluate network traffic patterns to identify issues in a network. Wang's approach in his 2010 dissertation was to utilize graph theory to categorize and aggregate network evidence in order to present a cohesive and comprehensible map of the network traffic. Given that networks consist of a set of nodes that are connected, it is natural to represent the nodes as vertices, and the connections as edges. In the case of network traffic, the nodes are servers, routers, switches, client computers, and other network devices. The level of traffic between two nodes (or vertices) can be represented by assigning a weight to the edge.

There have been limited attempts to apply graph theory to narrowly defined subsets of forensic investigation. In one study, graph theory was applied to the study of heroin seized in drug arrests[8] In this study, graph theory was used to evaluate heroin and the cutting agents used in producing heroin that is sold to consumers. The authors of the study state "An application of graph theoretic methods has been performed, in order to highlight the possible relationships between the location of seizures and co-occurrences of particular heroin cutting agents. An analysis of the co-occurrences to establish several main combinations has been done." Graph theory was used in this instance, to recognize patterns in the applications of cutting agents used in heroin. This was a significant advance in the application of graph theory to forensic investigations, but this study was very limited to a specific, highly focused, application. The current paper expands into a broader application of graph theory to any forensic examination.

Graph theory has been suggested as a methodology for the study of data in the examination of unstructured data in emails[9]. In their paper, the authors describe analysing the strength of relationships in the unstructured data, via the application of graph theory. The data elements are represented as nodes, and the vertices are used to describe both the presence of a relationship as well as the strength of that relationship. While this particular study is more narrowly defined that the current paper, it does suggest that ability to apply graph theory to the investigative process.

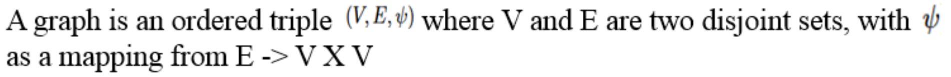
As can be seen in the above references, the literature already contains a few narrowly focused applications of graph theory to specific forensic questions. But aside from my own earlier paper, there is not a generalized methodology for applying graph theory to any forensic examination. In this paper, such a methodology is described. In addition to that methodology, the mathematics discussed in my previous paper are expanded upon. This current paper should provide a firm basis for utilizing graph theory to create accurate mathematical models of forensic investigations.

III. The Methodology

A significant challenge in collecting and categorizing digital evidence is to appropriately attribute evidence. While this can be an issue in any forensic investigation, it is a particular problem with digital forensics. In digital forensics, it is insufficient to simply trace a given attack vector to a specific network, or even a specific computer. Any number of individuals could have been utilizing that network or even the specific computer. Therefore, graph theory is particularly applicable to digital forensics. This is also consistent with the many studies wherein graph theory has been applied to analysing network traffic. It is a natural progression to move from analysing network traffic to analysing network attacks. In other forensic disciplines the goal can be to identify all potential suspects, pieces of evidence, and victims and to mathematically model the crime. By creating a graph representation of the case wherein each relevant entity is a vertex and the connections between vertices are edges, the forensic analyst can then apply graph theory to evaluating the evidence. The current methodology involves applying graph theory to create a mathematical model of the investigation, irrespective of the specific nature of the investigation. This can assist in attribution, but can also provide a robust view of the entire investigation and all elements therein. Essentially a graph is a set of vertices and the edges that connect them. This is shown mathematically as follows:

$G = (V, E)$

This definition is over simplified, and should be expanded. Later in this paper, the role of incidence functions in modelling investigations will be explored in more detail. For now, it is assumed the reader is aware that an incidence function maps a specific edge to the vertices on either end of the edge. The addition of incidence functions leads to a more complete mathematical description of a graph:



A graph is an ordered triple (V, E, ψ) where V and E are two disjoint sets, with ψ as a mapping from E -> V X V

Graph theory can be applied to a variety of different aspects of forensic science. The current methodology is concerned with describing the evidence in question and evaluating the connections between individual evidence items, suspects, victims, and any other entities relevant to a given investigation The methodology presented in this paper analyses all three elements of the graph: the vertex set, the edge set, and the incidence function that relates edges to vertices.

The issue of directionality of an arc, for forensic examination how to model direction is an important issue. The mathematics of graph theory simply state that an arc can be incident from one vertex to another. However, graph theory does not indicate how one determines that the arc begins at one vertex, rather than the other. For the purpose of forensic examination direction should always be from the initiator to the target. For example, if a given suspect visits the scene of a crime, then the arc would be from the suspect to the crime scene. In the case of digital forensic investigations, the issue for forensic examination is not the direction of the flow of data, but rather who initiated the flow of data. For example, if a given individual downloads a document from a website, the direction of the arc is from the individual to the website. Even though data flowed from the website to the individual, it was the individual who initiated the data flow.

The first step in applying graph theory to any investigation is to identify the various entities involved in the incident in question. These entities will be represented as vertices. When any two entities have any connection, that connection is represented as an edge. The edge should always be an arc that models not only the connection, but the initiator of that connection. It is possible, even likely, that some vertices will have multiple arcs. For example, if the investigation involves confidential information that was stolen from company A and subsequently found in company B, one can represent each company as a vertex, and relevant personnel as vertices. In most investigations, the more granular approach will be more effective.

Continuing with the previous example, each employee at company A that had access to the data in question would need to be represented. Then any connections these employees had to company B would be represented as edges or arcs. If evidence shows that the data passed through some intermediate entity, such as a hacker external to either company or perhaps a dark web market, then each employee in either company A or company B would have any relationships to that third-party entity represented as arcs.

The representation of entities and connections as a graph is a relatively simple process. The key issue will be to ensure that all entities and all of the connections are represented. As with any modelling tool, the model can only be as accurate as input allows. Graph theory allows or edges and arcs to be either weighted or not. For the purposes of modelling forensic investigations, weighted arcs will be most useful. Weighting of arcs must utilize a consistent approach. For this reason, ordinal measurements are recommended, rather than ratio measurements. One might assume ratio measurements would be more effective, however, in an investigation it is likely that many details will not be amenable to ratio measurements. Ordinal measurements merely require a ranking, without specific interval spacing[10]. The specific data used in such ordinal weighting will be dependant upon the particular investigation. However, some measurements will be consistent throughout any investigation. For example, criminal investigations usually consider the suspects means, motive, and opportunity[11]. A suspects arc to an element of the crime could be weighted as 1 to 3 depending on how many of these three aspects (means, motive, and opportunity) the suspect presented. Another approach to weighting would involve ranking the level of connection between two vertices, again with an ordinal measurement. For example, if a theft of a business is being investigated, an individual who had once been in the vicinity of the business would have an arc to the business weighted a 1. An individual who had occasionally visited the business would be weighted a 2. An individual with routing visitation of the business would be weighted a 3. An individual who was known to have access to not only the building but the items that were stolen would be weighted a four.

To illustrate the application of graph theory to forensic investigations, let us begin with a scenario involving a simple crime. In this scenario, a specific company has been subjected to a cyber-attack. A former employee is suspected of initiating the attack. However, the investigation has also shown this company is an industry that has recently been the target of nation state based cyber-attacks. The suspect, the hypothetical foreign based attacker, and the victim network all form vertices. The next step in modelling is to represent any connections between these vertices as edge. At this point we have a very simple graph. The suspect is vertex A, the infected website is vertex B, and the victim company is vertex C. This simple graph is shown in figure 1. This graph is particularly simple, to illustrate the process. However, real investigations would not only involve many more vertices, but would likely involve multiple edges or arcs between vertices.

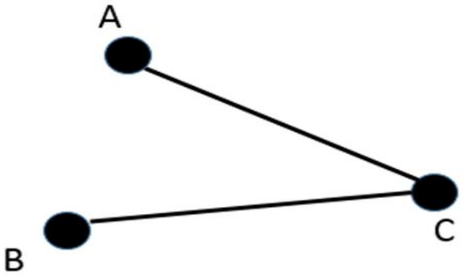


Figure 1: Graph of hypothetical crime

At this point no weighting nor directionality has been included in the graph. That would be an additional step the investigator would then take, that we will explore momentarily. While this can be illustrated with a traditional graph, an adjacency matrix can also be a useful method for modelling the investigation.

	A	B	C
A			1
B			1
C	1	1	

Adjacency Matrix

As the modeling process continues, in addition to adding additional arcs and vertices in order to accurately model the investigation, the arcs should be weighted. The modeling process can best be described by returning to our previous scenario of a company that has been the victim of a cyber-attack. A former employee is a suspect, but there is also some chance that a foreign based attacker is responsible. In our previous graph the former employee is represented by vertex A, the hypothetical foreign attacker as verted B, and the victim network as vertex C. To expand the model, let us assume that vertex B has no direct connection to verted C, but rather has multiple connections to a fourth vertex, D. Vertex D could represent a second company network that has been compromised and used as a basis for other attacks. Vertex D might have numerous connections to Vertex C. This is depicted in the graph shown in figure 2.

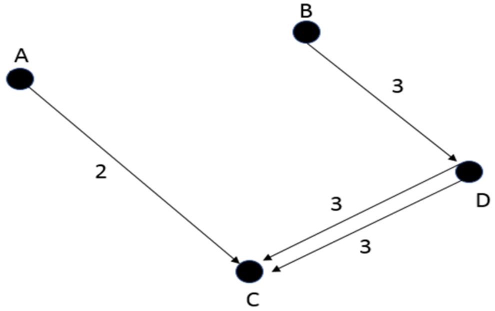


Figure 2 Graph with weighted arcs.

In this scenario, which is growing more complex, graph theory can be a valuable tool for evaluating the evidence. Furthermore, it may be more useful to use the adjacency matrix rather than a pictorial description of the graph. The adjacency matrix, as well as other algebraic forms of a graph[13] are more readily introduced into computer algorithms or spreadsheets, making analyses of the data easier.

An incidence matrix can also be useful in evaluating forensic evidence. An incidence matrix records edges that are incident from a given vertex. In the case of a forensic investigation, incidence graphs indicate the connections between vertices. Consider the graph from figure 2:

Note that the edges have direction, so that vertex D is incident to vertices C, twice (there are 2 arcs). And vertex B is incident to vertex D. To create an incidence matrix, the vertices are the rows and the edges are the columns. For directed graph, an edge is only considered if it is incident from a given vertex. For example, vertex D is incident to vertex C, but vertex C is not incident to D. The incidence matrix is shown below.

0	0	1	0
0	0	0	1
0	0	0	0
0	0	2	0

Incidence Matrix for directed graph

Just reviewing the incidence matrix one can readily ascertain that the strongest connection in this graph is via vertex D to vertex C. And this incidence matrix does not account for weights, merely for the number of connections. Traditionally incidence matrices in graph theory do not account for weighting. However, for the purposes of forensic investigation, one could add the weights of the connections in parentheses. Thus, continuing our hypothetical scenario, the following would be a weighted incidence matrix.

0	0	1(2)	0
0	0	0	1 (3)
0	0	0	0
0	0	2 (6)	0

Weighted Incidence Matrix

This modified incidence matrix provides the investigator with a clearer view of the relationships between entities in the case. This provides a mathematical model of the case, that can be readily analysed. Now it becomes even more apparent than the connection from vertex D to vertex C is the strongest. And we already know the vertex D is strongly connected to vertex C. At this point vertex B, the foreign attacker, is now emerging as a much stronger suspect than the former employee. While the graph at this point is significantly more detailed than when the modelling began, it can be expanded even further. The modelling described to this point is just a very basic modelling and should be considered the very minimum acceptable for modelling a forensic investigation.

One item to consider when evaluating a given vertex's importance in the mathematical model of an investigation is centricity. The center of a graph is the vertex (s) with minimal eccentricity, with eccentricity defined as the distance between a given vertex and the vertex(s) farthest from it[13]. A graph can have more than one center. Finding the center(s) of a graph, when modelling a forensic investigation, can be useful in determining which entities (as represented by vertices) are most involved with the overall case. Put another way, a given center may not be the perpetrator of a given crime, but clearly is of importance in understanding the case.

Another issue to address is the incidence function for a given edge or arc. The incidence function associates each edge with an unordered pair of vertices[14]. For example, if you have graph G with edge A, that connects vertices u and v, the incidence function is defined as follows:

$$\psi G(a) = uv$$

This does expand our previous definition of a graph to the following, which was introduced earlier in this paper:

$$G = (V, E, \psi)$$

Typically, in graph theory the exact nature of the incidence function is not a primary concern. The concern is that vertex u and vertex v are connected in some manner, thus creating the edge a. However, in modelling a forensic investigation the incidence function takes on a more important role. To simply establish that there is a connection between vertex u and vertex v is inadequate. So far, the methodology presented has added direction and weighting, but that does not describe the incidence function. For the purposes of modelling a forensic investigation, the actual function that connects vertex u to vertex v must be describes. In many, if not most situations, this may not be a typical mathematical function. Rather the incidence function is most likely to be a narrative description of the relationship. For example, arc a connects vertex u and v, because the individual represented by vertex u visited the website represented by vertex v.

Once an investigation is completely modeled with an accurate graph, isomorphisms can be used to compare this current incident to other similar incidence. Two graphs are isomorphic if they have the following properties:

1. Same number of vertices
2. Same number of edges
3. The vertices are of the same degree

If you have a complete and accurate graph of a given incident, then any incident that produces an isomorphic graph may be related. For example, if you create a graph of a known nation state sponsored breach of a network, then while investigating a new and separate breach, you find the graph of the new breach is isomorphic with the graph of the nation state attack, then this would make it more likely that the new breach is related to the first and possibly perpetrated by the same individuals.

Traditionally, graph theory does not have a concept of partial isomorphisms. However, for a forensic investigation, a partial isomorphism is still of interest. For example, the graphs of two separate crimes might not be isomorphic, but if they are 90% isomorphic that would still indicate a common perpetrator. Partial isomorphisms can be evaluated with a very simple calculation. The percentage of identical vertices multiplied by the percentage of identical edges yields the percentage of isomorphism between the two graphs. Put mathematically:

$\%E(\%V)=\%I$

For example, if 95% of the vertices are identical, and 90% of the edges are identical then the two graphs of the two crimes would be considered 85.5% isomorphic. It would also be possible to further detail the examination of isomorphism by examining the incidence functions for edges/arcs to see if those functions are identical.

IV. Conclusions

Graph theory is a well-established, mathematical method for evaluating relationships. Applications to graph theory for network modeling and electrical engineering are also well established and widely used. There have been some tentative steps towards applying graph theory to model narrowly defined, highly specific areas of forensic investigation. In this paper, a methodology has been described for a generalized approach of modeling forensic investigations that utilizes well-established principles of graph theory. This methodology can be applied to any investigative process to provide a comprehensive mathematical model of all of the elements of a given investigation along with the relationships between those elements.

The entire graph, including vertices, arcs, and incidence functions must be evaluated in order to have an accurate model of the investigation. It is also possible to compare two diverse crimes, by evaluating the degree of isomorphism of the graphs of those investigations. In this paper, a simple formula for evaluating partial isomorphism was introduced.

References

[1]. Easttom, C. (2017). Utilizing Graph Theory to Model Forensic Examination. International Journal of Innovative Research in Information Security (IJIRIS), 4(2).

[2]. Balakrishnan, V.K. (2010). Introductory Discrete Mathematics. Mineola, New York: Dover Publications

[3]. Deo, N. (2016). Graph Theory with Applications to Engineering and Computer Science. Mineola, NY: Dover Publications

[4]. Wang, W., & Daniels, T. E. (2006, September). Diffusion and Graph Spectral Methods for Network Forensic Analysis. In Proceedings of the 2006 workshop on New security paradigms (pp. 99-106). ACM.

[5]. Ahlswede, R., Cai, N., Li, S. Y., & Yeung, R. W. (2000). Network information flow. IEEE Transactions on information theory, 46(4), 1204-1216

[6]. Holme, P. (2003). Congestion and Centrality in Traffic Flow on Complex Networks. Advances in Complex Systems, 6(02), 163-176

[7]. Amaral, L. A., & Ottino, J. M. (2004). Complex networks. The European Physical Journal B- Condensed Matter and Complex Systems, 38(2), 147-162.

[8]. Zufferey, A., Ratle, F., Ribaud, O., Esseiva, P, & Kanevski, M. (2006). Pattern Detection in Forensic Case Data Using Graph Theory: Application to Heroin Cutting Agents. Forensic Science International 167 (2-3), pp 242–246.

[9]. Haggerty, J., Karran, A., Lamb, D., & Taylor, M. (2011). A Framework for the Forensic Investigation of Unstructured Email Relationship Data. International Journal of Digital Crime and Forensics, 3(3), 1-18.

[10]. Gibilisco, S. (2004). Statistics Demystified. New York City, NY: McGraw-Hill

[11]. McKendall, Marie A., and John A. Wagner III. "Motive, opportunity, choice, and corporate illegality." Organization Science 8.6 (1997): 624-647.

[12]. Godsil, C., & Royle, G. F. (2013). Algebraic graph theory (Vol. 207). Springer Science & Business Media

[13]. Trudeau, R. (1994). Introduction to Graph Theory. Mineola, New York: Dover Publications

[14]. Bondy, A., Murty, U. (2008). Graph Theory. New York City, NY: Springer Publishing