Utilizing Graph Theory to Model Forensic Examinations

Mathematical modeling of forensic data

May 6, 2018

Who is the speaker

- 26 books (Including 3 on forensics, 6 on computer security, 1 on cryptology)
- Authored over 40 research papers
- Over 40 industry certifications including several forensic certifications
- 2 Masters degrees (3rd in progress); D.Sc. ***In progress***
- 13 Computer science related patents
- Over 25 years experience, over 15 years teaching/training
- Helped create CompTIA Security+, Linux+, Server+. Helped revise CEH v8. Created the ECES certification
- Associate Member of the American Academy of Forensic Sciences
- Created the OSForensics Certified Examiner course and test (OSCFE)
- Frequent speaker at security and forensic conferences including AAFS, ADFSL, IAFLS, ISC2 Security Congress, Secure World, Secure Jordan, Defcon, and others
- Frequent consultant/expert witness

www.chuckeasttom.com

chuck@chuckeasttom.com

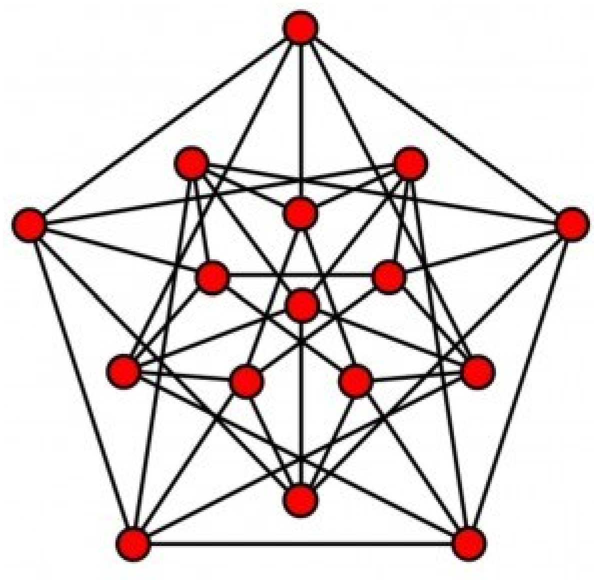
The presentation outline

- Overview of Graph Theory Basics
- Basic Modeling with Graph Theory
- More details on graph theory and evaluating models

The Genesis of this Technique

- Complexity
- Mathematics <-> Forensics

Graph Theory Basics



Graph Theory

Graph theory is an important part of discrete mathematics. It is a way to examine objects and the relationship between those objects mathematically . 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.

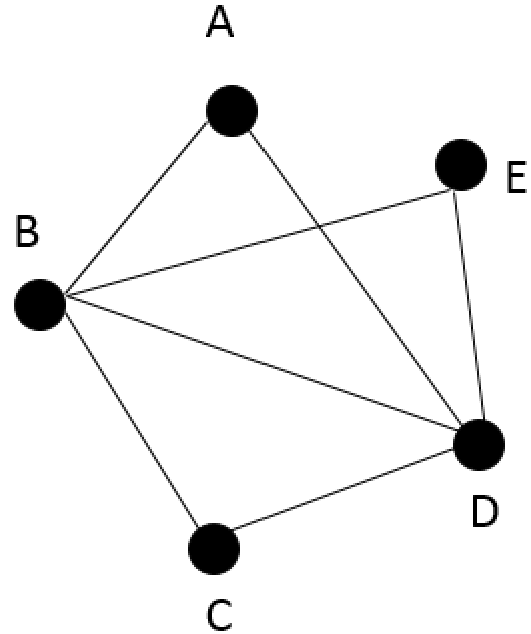
$$G = (V, E, \psi)$$

Now let's examine that definition in more reader-friendly terms. A graph starts with vertices or nodes, what I previously referred to as objects. These objects can be anything. The edges are simply lines that show the connection between the vertices. The edges are ordered pairs and not necessarily symmetrical. In other words, the connection between two vertices may or may not be one-way.

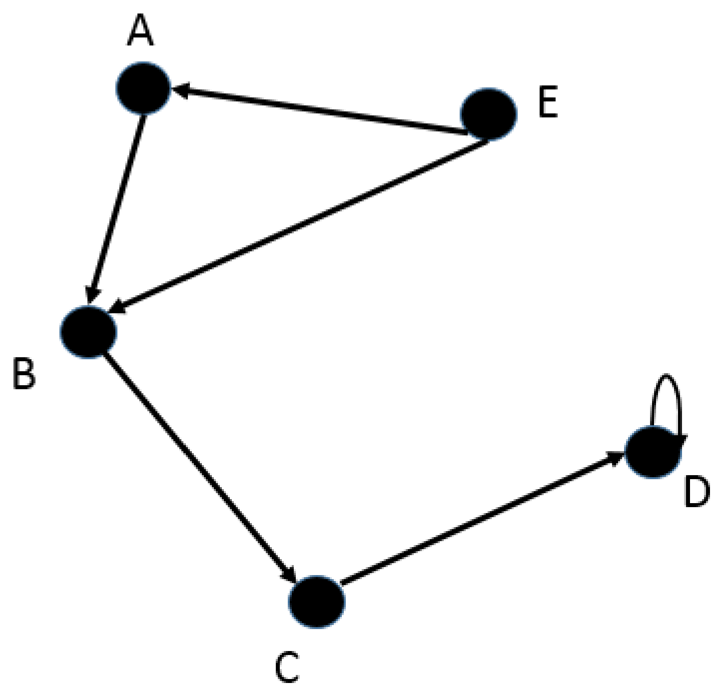
Outline

I. Essentials of graph theory
II. Applying graph theory to forensic examinations
III. Area's for further research

A Basic Graph



A Basic Digraph



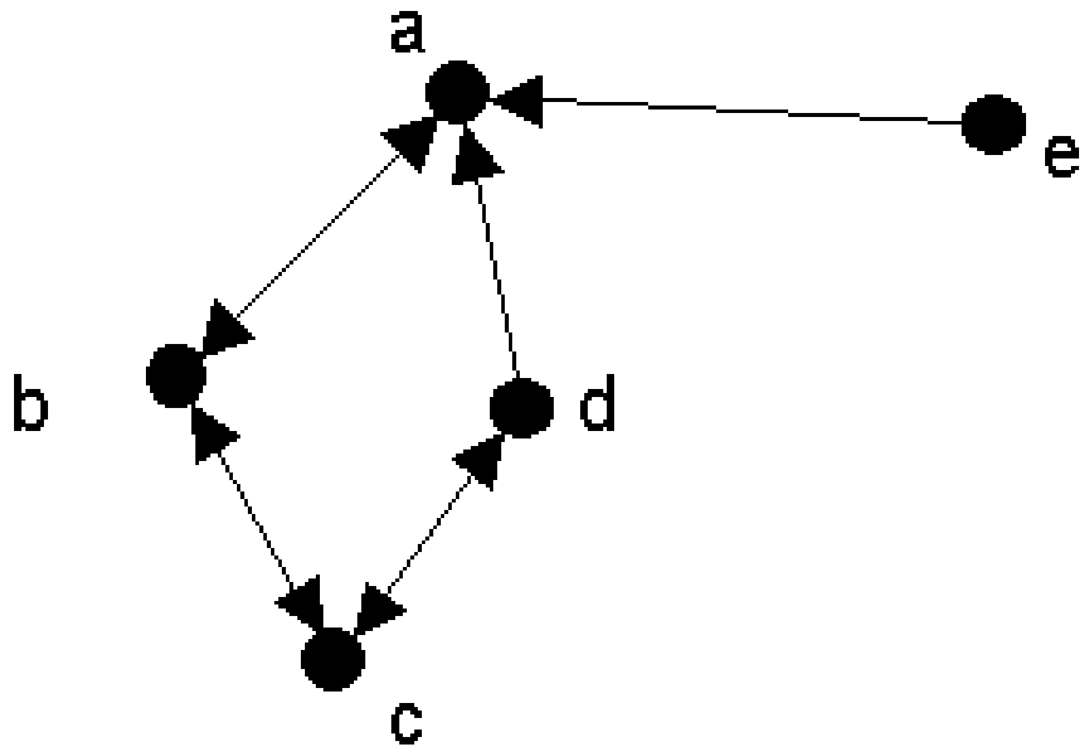
Degree of a vertex: Number of edges incident to the vertex. Nodes of a digraph can also be said to have an *in* degree and an *out* degree. The in degree is an edge pointing toward the vertex. The out degree is an edge pointing away.

Adjacency: Two vertices connected by an edge are adjacent.

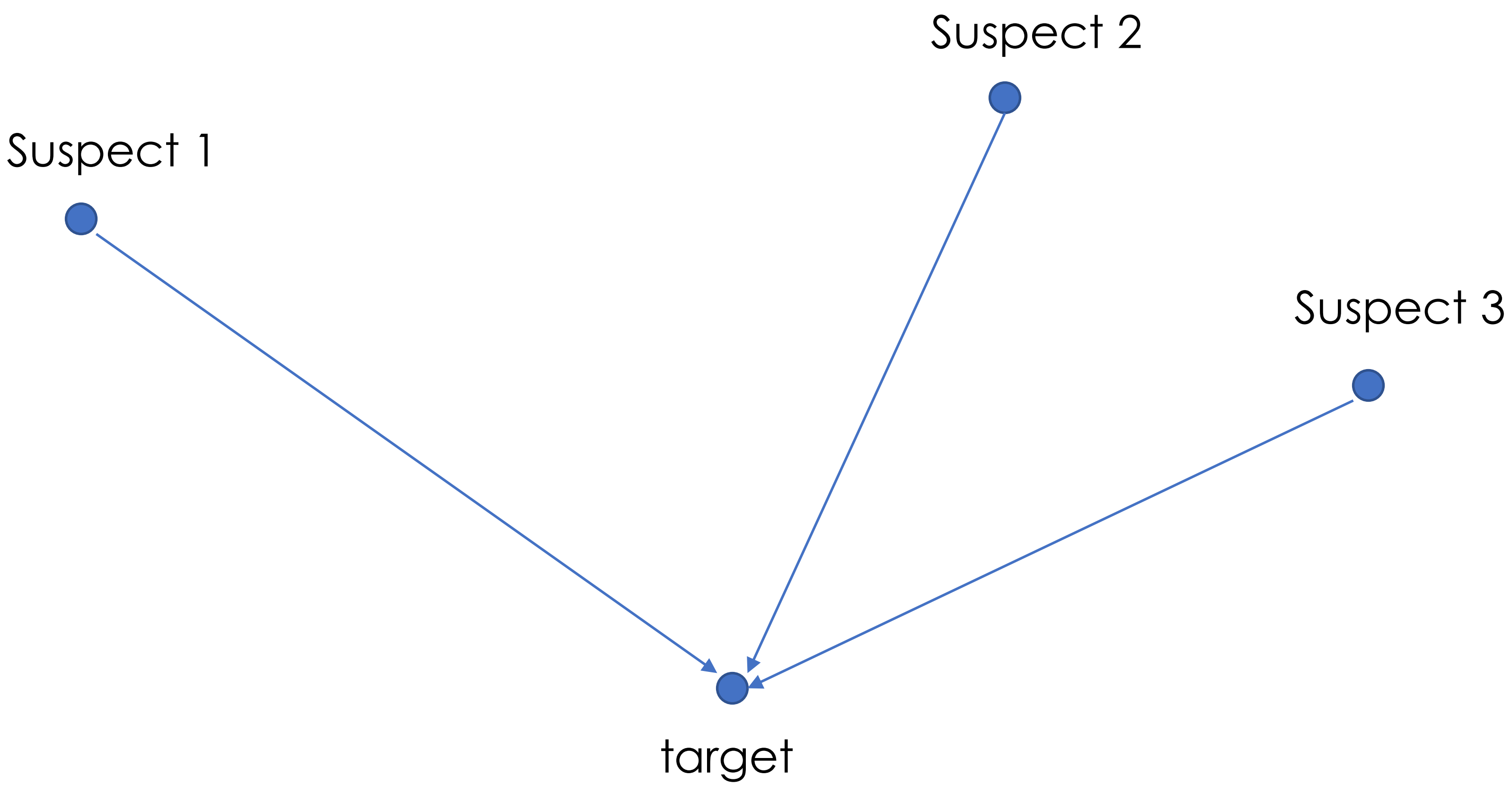
Directed graph: In a directed graph, often called a digraph, the edges have a direction.

Weighted digraph: This is a directed graph that has a "cost" or "weight" associated with each edge. In other words, some edges (that is, some relationships) are stronger than others.

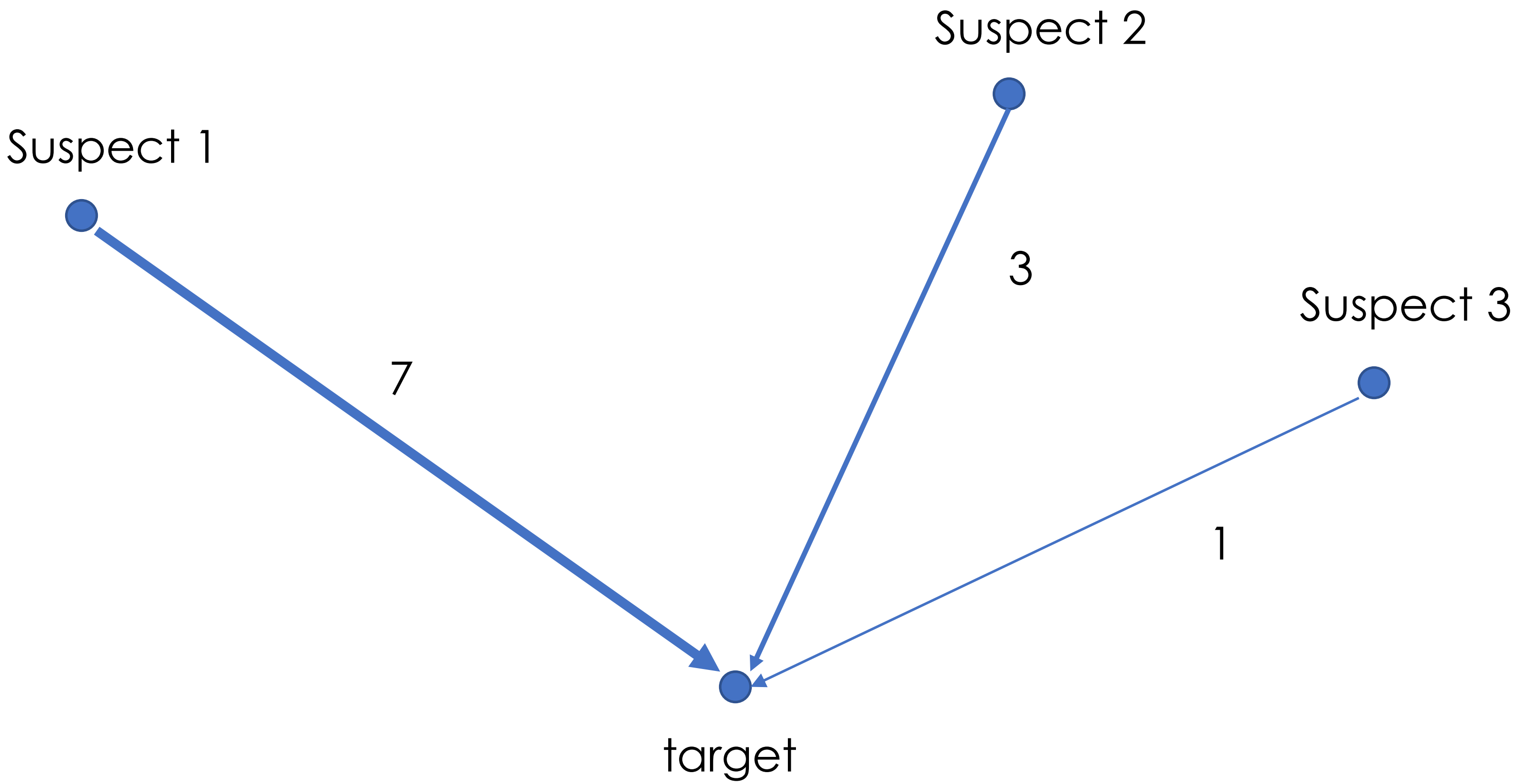
Basic Modeling with Graph Theory



Step 1 in applying Graph Theory
We can create a modified graph that includes a vertex for each of the points (suspect machines, target, any source of traffic, etc.)
The directed graphs showing relationships



Step 2
Take the weightings you found earlier and annotate each directed graph
Perhaps even thicken the arrow based on weight

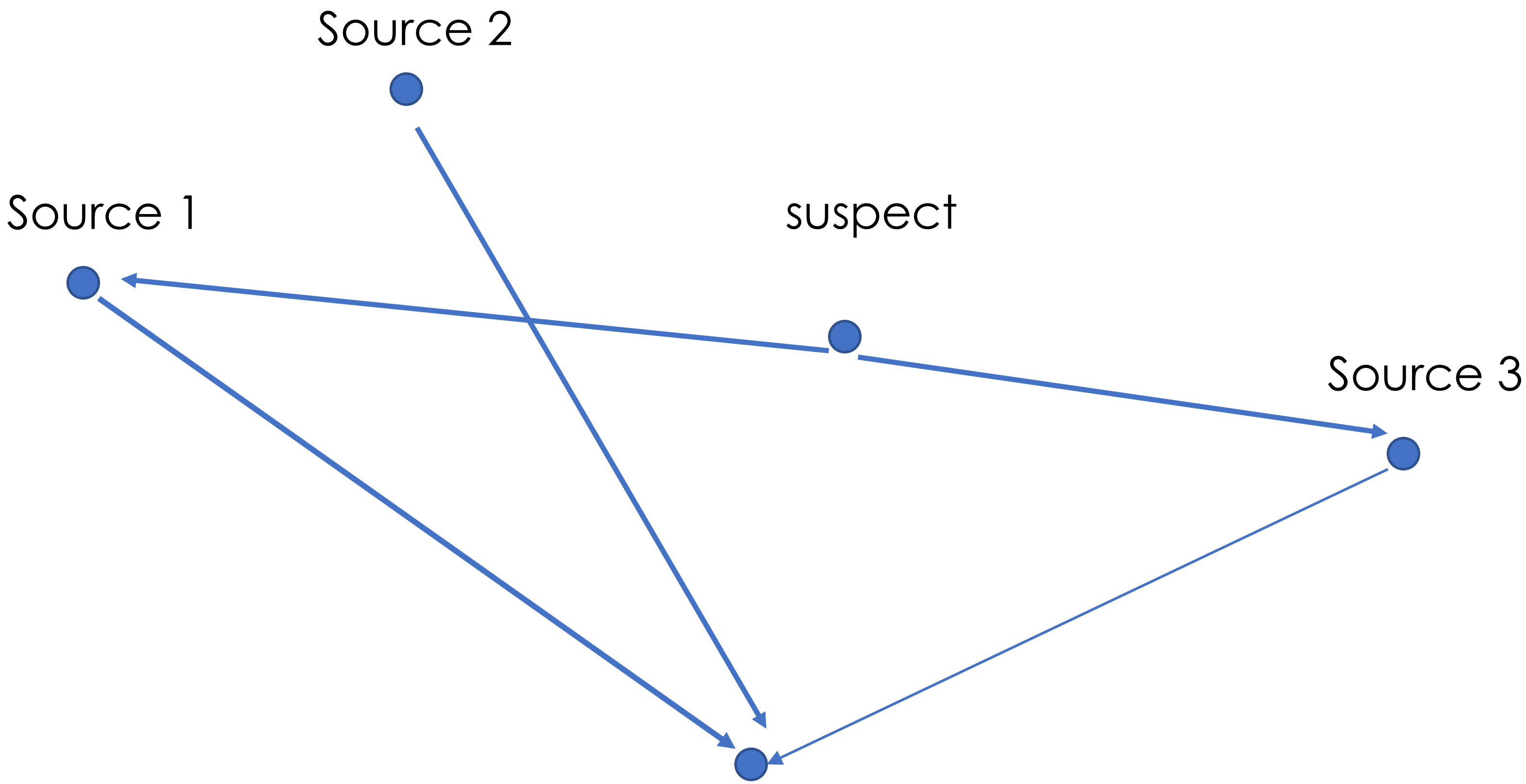


How to weight connections

If a digraph's edges have a specific cost or weight associated with each edge, then the graph is considered to be a *weighted digraph* (Trudeau, 1994). For the purposes of evaluating relationships between objects, weighted digraphs are very effective.

For the purposes of evaluating evidence, an ordinal or interval measurement will be the most accurate. Ordinal measurements merely require a ranking, without specific interval spacing (Gibilisco, 2004). An example ranking for a connection between a given computer user and a particular web server could be 1) casual and infrequent visits; 2) casual and frequent visits; 3) significant interaction with the web server; and 4) either administrative access to the web server or deliberate hacking of the web server. This type of ordinal ranking actually provides the investigator with a much clearer understanding of the nature of the user's interaction with the web server. An ordinal approach, with a small number of ordinals, also simplifies the weighting process. For the purposes of analyzing digital evidence, ordinal approaches with 5 or fewer ordinals is recommended.

More complex views will include multiple points and multiple connections, this is where graph theory will really play a part. We have multiple sources of traffic, some seem to connect to the suspect, some do not.



The pictorial representation of the graph is not even necessary. Various matrices can be utilized

	A	B	C
A			1
B			1
C	1	1	

Adjacency Matrix

0	0	1	1	0
1	0	0	1	1
0	1	0	0	0
0	0	1	0	0
1	0	1	1	0

Incidence Matrix

One can even show weight in an incidence matrix.

0	0	1 (2)	1 (3)	0
1 (4)	0	0	1 (1)	1 (1)
0	1(3)	0	0	0
0	0	1(1)	0	0
1(1)	0	1(3)	1(1)	0

Weighted Incidence Matrix

So how does this help

First it allows you, the investigator to really see the evidence, and avoid tunnel vision.

Second, it can provide very good visual aids for trial.

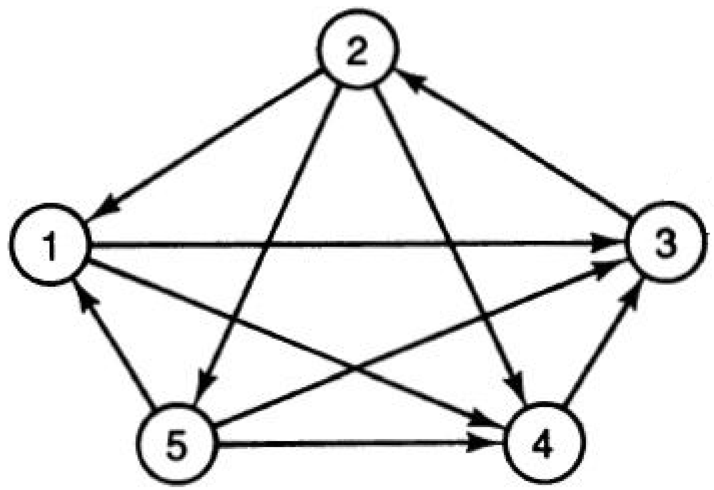
Weighting

An ordinal methodology is recommended

For example, when weighting the connection between an employee at the victim company and the infected website can be expressed as an ordinal value, such as

1) visited the web site very infrequently and is not known to have downloaded anything;
2) visited the website with some frequency;
3) routinely visited the website;
4) is known to have downloaded files from the infected website.

An incidence matrix



An incidence matrix

0	0	1	1	0
1	0	0	1	1
0	1	0	0	0
0	0	1	0	0
1	0	1	1	0

Incidence Matrix

Evaluating Graphs

An ordinal methodology is recommended

For example, when weighting the connection between an employee at the victim company and the infected website can be expressed as an ordinal value, such as

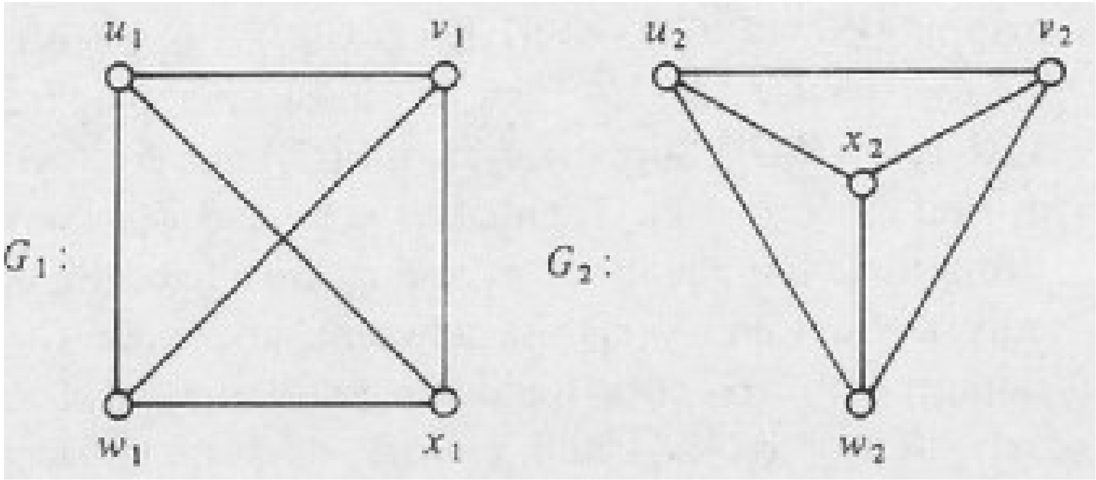
1) visited the web site very infrequently and is not known to have downloaded anything;
2) visited the website with some frequency;
3) routinely visited the website;
4) is known to have downloaded files from the infected website.

A weighted incidence matrix

0	0	1 (2)	1 (3)	0
1 (4)	0	0	1 (1)	1 (1)
0	1(3)	0	0	0
0	0	1(1)	0	0
1(1)	0	1(3)	1(1)	0

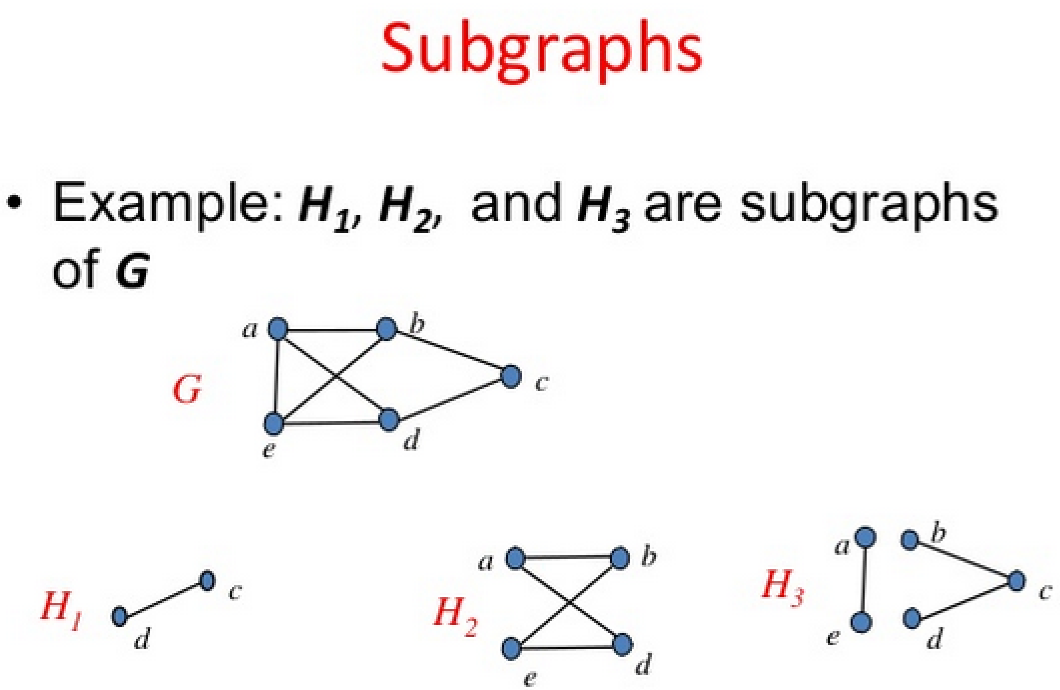
Weighted Incidence Matrix

Deeper with Graph Theory and Modeling



Subgraph

A subgraph of a graph *G* is another graph formed from a subset of the vertices and edges of *G*. The vertex subset must include all endpoints of the edge subset, but may also include additional vertices. A spanning subgraph is one that includes all vertices of the graph; an induced subgraph is one that includes all the edges whose endpoints belong to the vertex subset.



Isomorphisms

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

Two graphs which contain the same number of graph vertices connected in the same way are said to be isomorphic. Formally, two graphs G and H with graph vertices $V_n = \{1, 2, \ldots, n\}$ are said to be isomorphic if there is a permutation p of V_n such that $\{u, v\}$ is in the set of graph edges $E(G)$ iff $\{p(u), p(v)\}$ is in the set of graph edges $E(H)$.

-Wolfram Mathworld

Isomorphisms – Forensic Implications

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.

Isomorphisms – Forensic Implications

This can be applied to serial killers. If you have a complete and accurate graph of known crimes of a given serial killer, it should be the case that these graphs are largely isomorphic, or at least have significant subgraphs that are isomorphic.
These graphs of known crimes can then be compared to graphs of new crimes to determine if it is likely they were perpetrated by the same serial killer.

Partial Isomorphism

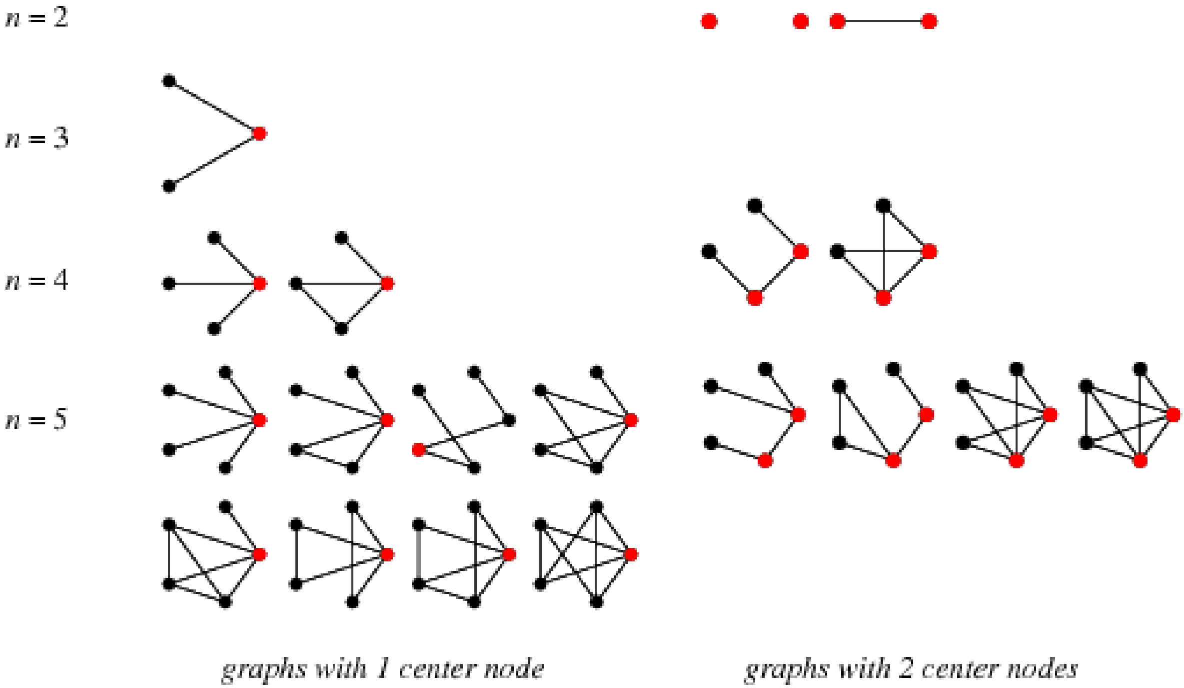
The degree of isomorphism is defined as the percentage to which two graphs are isomorphic. This is expressed as a percentage, rather than as an integer value. To calculate the degree of isomorphism between two graphs requires a rather simple formula. The percentage of identical vertices added to the percentage of identical edges, that sum divided by two, yields the percentage of isomorphism between the two graphs. To put this in a more mathematically rigorous format, Given G1 =(V1, E2, ψ1) and G2 = (V2, E2, ψ2), the formula in figure 4 illustrates how to compute the degree of isomorphism (note that %I is the percentage of isomorphism between the two graphs).

$$\frac{\left(\frac{\sum_{i=1}^{n} G_{1i}V_{1i} = G_{2i}V_{2i}}{n} + \frac{\sum_{i=1}^{n} G_{1i}E_{1i} = G_{2i}E_{2i}}{n} \right)}{2} = \%I$$

Centers

The center of a graph is the vertex (s) with minimal eccentricity

- Eccentricity is defined as the distance between a given vertex and the vertex(s) farthest from it.
- A graph can have more than one center.



-From Wolfram Mathworld

Centers – Forensic Implications

If the center(s) of a graph are devices in a network intrusion investigation, then these are the most important locations to seek evidence.

If the center(s) of a graph are individuals in any investigation, there is a reasonable chance the these individuals are victims or involved in the crime.

It will always be the case that centers in a graph of an incident are points of interest and one should focus the investigation on these points.

Incidence Functions

Essentially what connects vertex A to vertex B. Put another way, why is there an edge (or an arc) at that location?

These could be actual mathematical functions, but are more likely to be textual descriptions of the relationship between the vertices. This should be directly related to the weighting.

Weighting – Additional Considerations

- Edge Weighting
- Vertex Weighting
- Both

Weighting - issues

- Somewhat subjective
- Greatest potential for bias
- This is why an ordinal approach

This tool and a Daubert Challenge

First introduced at SecureWorld Dallas September 2016
A nascent version published in November 2016: Easttom, C. (2016). Multi-Dimensional Analysis for Cyber Investigations. Forensic Examiner Journal,25 (4).

An expanded version published as an open source article: Easttom, C. (2016). Applying Graph Theory to Evidence Evaluation. *Research Gate*. DOI: 10.13140/RG.2.2.23391.0528

Presented as an invited speaker at a forensic conference in Cairo in January 2017: Easttom, C. (2017). Utilizing Graph Theory to Model Forensic Examinations- Presentation for the 2nd Annual International Congress of the International Association of Law and Forensic Science (IAFLS). - Cairo Egypt, January 2017. DOI: RG.2.2.33025.66407

Presented at ISC2 Security Congress 2017.

Presented at Enfuse Conference 2017

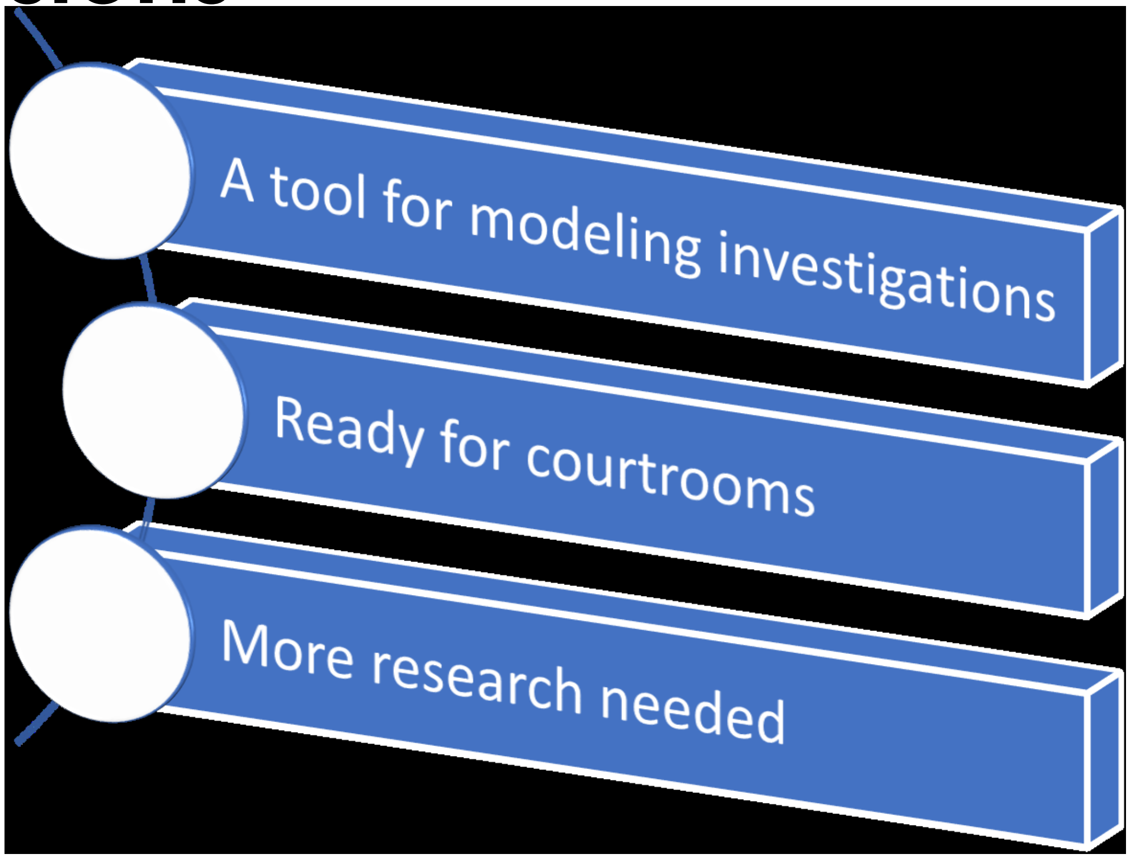
An expanded version published in February 2017 - Easttom, C. (2017). Utilizing Graph Theory to Model Forensic Examination. International Journal of Innovative Research in Information Security (IJIRIS), 4(2).

An expanded version is currently undergoing peer review for yet another journal.

Areas for Further Research

- Isomorphism's
- Other forensics disciplines
- Weighting Methodologies
- Case Studies

Conclusions



Any questions/comments?

www.ChuckEasttom.com
chuck@chuckeasttom.com

References

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

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

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

Bollobás, B. (2013). *Modern graph theory (Vol. 184)*. Springer Science & Business Media

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

Catanese, S. A., Fiumara, G. (2010, October). A visual tool for forensic analysis of mobile phone traffic. *In Proceedings of the 2nd ACM workshop on Multimedia in forensics, security and intelligence* (pp. 71-76). ACM.

Chartrand, C. (1985). *Introductory Graph Theory*. New York City, NY: Dover Publication.

References

Chaski, C. (2005). Who's at The Keyboard? Authorship Attribution in Digital Evidence Investigations. *International Journal of Digital Evidence, 4*(1).

Clark, J., & Holton, D. A. (1991). *A first look at graph theory (Vol. 1).* Teaneck, NJ: World Scientific.

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

Easttom, C. (2016). Applying Graph Theory to Evidence Evaluation. *Research Gate*. DOI: 10.13140/RG.2.2.23391.0528

Easttom, C. (2017). Utilizing Graph Theory to Model Forensic Examinations- Presentation for the 2nd Annual International Congress of the International Association of Law and Forensic Science (IAFLS). - Cairo Egypt, January 2017. DOI: RG.2.2.33025.66407

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

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

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

References

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.

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

Peterson, G., Shenoi, S. (2011). Advances in Digital Forensics VII: 7th IFIP WG 11.9 International Conference.

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

Wang, Wei, (2010). A Graph Oriented Approach for Network Forensic Analysis. Graduate Theses and Dissertations. Paper 11736

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.