Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 108C (2017) 2388–2392

Procedia Computer Science

International Conference on Computational Science, ICCS 2017, 12-14 June 2017, Zurich, Switzerland

Feasibility Study of Social Network Analysis on Loosely Structured Communication Networks

Jan William Johnsen and Katrin Franke

Norwegian University of Science and Technology (NTNU), Norway
jan.w.johnsen@ieee.org and kyfranke@ieee.org

Abstract

Organised criminal groups are moving more of their activities from traditionally physical crime into the cyber domain; where they form online communities that are used as marketplaces for illegal materials, products and services. The trading of illicit goods drives an underground economy by providing services that facilitate almost any type of cyber crime. The challenge for law enforcement agencies is to know which individuals to focus their efforts on, in order to effectively disrupting the services provided by cyber criminals. This paper present our study to assess graph-based centrality measures' performance for identifying important individuals within a criminal network. These measures has previously been used on small and structured general social networks. In this study, we are testing the measures on a new dataset that is larger, loosely structured and resembles a network within cyber criminal forums. Our result shows that well established measures have weaknesses when applied to this challenging dataset.

Peer-review under responsibility of the scientific committee of the International Conference on Computational Science

Keywords: Digital forensics; Social network analysis; Centrality measures; Criminal networks

1 Introduction

Law enforcement agencies report that cyber crime activity is growing and become more aggressive and technically proficient [3, 7] – although the majority of cyber criminals in online marketplaces have relatively low technical skills and capabilities. This suggests that a minority of cyber criminals use marketplaces to sell easy access to sophisticated tools and expetise, through a business model called Crime-as-a-Service (CaaS) [3]. Which allow lesser skilled cyber criminals to have more impact and success in their cyber attacks. A focus on identifying and disrupting criminals in the smaller and more technical skilled group will have a larger impact on stopping illegal activities in underground marketplaces. Because their skills and expertise are difficult to replace by the larger group, with lower technical skills.

Social Network Analysis (SNA) methods has been proposed [9] for the application of identifying central individuals within criminal networks. More specifically, centrality measures are used to determine central individuals by analysing their position in a network [8], represented

1877-0509 © 2017 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the scientific committee of the International Conference on Computational Science
10.1016/j.procs.2017.05.172

by a graph as defined in Section 2. In previous research, centrality measures has been used to analyse relational structures in organisations [4, 5, 2] and terrorist groups [6]. The network size in these studies are between 30 and 150 individuals. Centrality measures have shown promising results to find central individuals in small and organised networks – although the networks has been incomplete or is just a sample from the total population.

However, real world datasets are neither small nor organised, and they often requires data preprocessing before they can be analysed. Although centrality measures has performed good on networks of smaller sizes by finding interesting individuals, this does not mean they will also perform good on larger and more loosely structured [1] networks. This paper is guided by the research question: *How can graph-based methods be applied to identify important individuals within a real-world online communication network?* Our research question seeks to determine the feasibility of centrality measures in applying it to the area of civil and criminal investigations.

2 Methodology

We extracted information to represent the communication within Nulled.IO as graphs: users and the messages between them, represented as vertices and edges respectively. It has not been pre-filtered and is used in its original form (detailed in Section 3) except for separating public and private messages; which results in two graphs with public communication between 26.11.2012 - 06.05.2016 and private communication between 14.01.2015 - 06.05.2016.

The reason for this division is twofold: (i) communication patterns is likely to be different between them, and (ii) civil investigators only have access to public communication in their investigation, whereas criminal investigators will have access to both.

The four centrality measures under evaluation are: *degree*, *betweenness*, *closeness* and *eigenvector*. They differ in the interpretation of *important*, thus different individuals will be ranked as more important in the same network; illustrated in Figures 1 - 4.

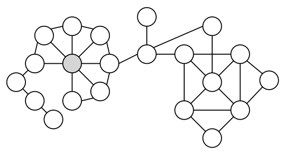

Figure 1: Largest degree centrality

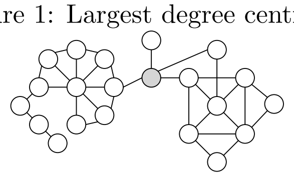

Figure 2: Largest betweenness centrality

A (undirected) graph $G = (V, E)$, where V is the set of vertices and E is the set of edges, is represented in terms of the binary adjacency matrix A. Degree centrality is the most basic measure as it only counts directly adjacent vertices. For a vertex $v \in V$, it is defined by $C_D(v) = \sum_{u=1}^{n} A_{v,u}$, where $n = |V|$. The centrality measures discussed in this paper do not consider the diagonal elements in A [8], where $v = u$, because the relationship to oneself is not important.

Betweenness centrality looks at how often a vertex sits in the *geodesic* (shortest path) between two other vertices. A vertex is considered more important because it can act like a *broker* – i.e. arrange or negotiate plans and deals – and have more influence on the network by choosing to withhold or distort information [8]. Figure 2 highlights the vertex in the network with the highest betweenness centrality score, because it sit in between two large subgraphs and one vertex. Betweenness centrality for a vertex v is defined by $C_B(v) = \sum \frac{\partial_{u,v,w}}{\partial_{u,w}}$, where $\partial_{u,w}$ is the total number of shortest paths between vertex u and w, and $\partial_{u,v,w}$ is the number of those paths that pass through v, and $u \neq v \neq w$.