Identifying Key-Players in Online Activist Groups on the Facebook Social Network

Mariam Nouh
Centre for Doctoral Training in Cyber Security
Department of Computer Science
University of Oxford, UK
Email: mariam.nouh@cs.ox.ac.uk

Jason R. C. Nurse
Cyber Security Centre
Department of Computer Science
University of Oxford, UK
Email: jason.nurse@cs.ox.ac.uk

Abstract—**Online social media applications have become an integral part of our everyday life. Not only are they being utilised by individuals and legitimate businesses, but also recently several organised groups, such as activists, hactivists, and cyber-criminals have adopted them to communicate and spread their ideas. This represents a new source for intelligence gathering for law enforcement for instance, as it allows them an inside look at the behaviour of these previously closed, secretive groups. One possible opportunity with this online data source is to utilise the public exchange of social-media messages to identify key users in such groups. This is particularly important for law enforcement that wants to monitor or interrogate influential people in suspicious groups. In this paper, we utilise Social Network Analysis (SNA) techniques to understand the dynamics of the interaction between users in a Facebook-based activist group. Additionally, we aim to identify the most influential users in the group and infer their relationship strength. We incorporate sentiment analysis to identify users with clear positive and negative influences on the group; this could aid in facilitating a better understanding of the group. We also perform a temporal analysis to correlate online activities with relevant real-life events. Our results show that applying such data analysis techniques on users online behaviour is a powerful tool to predict levels of influence and relationship strength between group members. Finally, we validated our results against the ground truth and found that our approach is very promising at achieving its aims.**

Keywords-**Social Network Analysis; Influential Users; Cyber Security; Sentiment Analysis, Online Social Networks;**

I. INTRODUCTION

The popularity of online social networks (OSNs) has significantly increased in the last decade. Individual people, businesses, and news agencies use OSNs such as Twitter and Facebook to discuss topics, promote and advertise services, spread news and political views. Over the last few years, several organised groups including terrorists, activists, hacktivists have been utilising OSNs to communicate their ideas, plan or brag about crimes, and distribute messages [1]. According to experts in the criminology field, the existence of organised cyber-criminals in the online world is growing rapidly [2]. Although the use of online social services provide these organised groups with a medium to communicate and voice their views, it also provides law enforcement and the research community with a window into their world to better understand the behaviour of these communities.

Social Network Analysis (SNA) is a method used to investigate social structures by utilising graph theory concepts [3], [4]. SNA techniques have proven to be particularly useful in studying and analysing the structure and behaviour of social groups [5]. SNA is typically used to study real-world networks, either using static techniques that analyse the structural properties of the network and/or using dynamic techniques that use statistical methods to model different network processes over time. Furthermore, using SNA metrics such as centrality measures [6] provide insights into the community structure and key players within a network.

The aim of this paper is to investigate SNA metrics that can aid in identifying key players within a given organised group, mainly of activists. National and international activist groups often use web forums to promote movements and distribute propaganda materials. Although some of these activist groups organise peaceful activities, some escalate to hostile movements which may cause disruption and financial losses to targeted organisations. The identification of key players in a given organised group of interest can help authorities save resources spent on investigating the whole network especially when the network is huge and complex [7]. Additionally, this can serve as a proactive measure to predict the occurrence of any potentially disruptive offline action. For instance, last year the Australia and New Zealand Banking Group (ANZ) head office was in lock-down as around 80 people aggressively protest the bank's funding of fossil fuel projects [8]. Another example is when two activists who broke into a coal mine and scaled equipment, obstructed the work of mining equipment resulting in financial losses [9].

Much of the existing efforts that aim to identify key individuals within organised criminal groups rely on a single type of analysis such as SNA metrics, or text mining. Relying on a single method may fail to identify all influential people as well as it can misidentify them as influential while they are not (false positives). Therefore, a hybrid approach that combines several analysis techniques to identify key influential individuals will give more accurate and reliable results. In this paper, we perform an empirical study to identify key players within activist groups in the Facebook OSN. When analysing key players we must make the

distinction between the most active member of the group and the most influential. A user who is the most active in an OSN group in terms of posting activities, is not necessarily the most influential on the group. Our goal is to investigate the inference of such influential actors in real life through their social network interactions. Additionally, we intend to identify key players in suspicious activist networks (both most active and most influential) where targeting these individuals may (1) result in destruction of the network, and (2) aid in collecting intelligence about the network and mediating the flow of information.

Furthermore, after identifying the most influential individuals we utilise sentiment analysis on posts and comments to identify whether the identified individual has a mainly positive or negative influence on the group. Thus, we identify key players that can potentially persuade the majority to accept or refuse a given proposition. Furthermore, we initially study how the group evolves and grows over time by performing time analysis to study when and why the group activity level increases or decreases. We correlate the observed trends with related real-world events and try to identify patterns in offline and online behaviour. These additional analysis techniques will aid in gaining useful group insights.

The main contributions of our work are summarised as follows:

- Understand the dynamics of the interactions in potentially suspicious activist networks.
- Apply SNA techniques to identify the key players in organised activist groups. This includes the most active and most influential.
- Inference of trust relations between actor pairs within a social network, based on structural properties and sentiment analysis information to gain further group insights.
- Perform a temporal analysis of the network posting structure and compare it over time. In addition, we investigate the correlation of the amount of online activity with related real-world events.

The remainder of this paper is organised as follows: Section II provides a brief background and review of related work. Section III and Section IV describe the methodology we follow and introduce the dataset used in our research. Section V details the analysis that we conducted followed by a validation and discussion of the results in Section VI. Finally, we conclude this paper in Section VII, and present ideas for future work.

II. Background and Related Work

Previous research has stressed the need to identify key players within different organised groups. Mainly, the focus has been on criminal groups such as drug dealers, and terrorists with the aim to understand the structure of these groups to be able to effectively disrupt their behaviour and help law enforcement capture them [10]–[13]. There are two main streams of research in criminal network literature. The first is the study of the main enablers of link-formation, in other words, what motivates individuals to form or break links [2]. Some of these studies suggest that prison for instance might be one of these enablers as people who spent prison time together are more likely to form a link in the future and commit a crime [14]. Others suggest that similar individual attributes such as age, race, and religion are key in forming links and co-offending [15]. The second stream is about utilising statistical methods and social network analysis to measure these networks, identify key individuals, and measure tie strength between them.

Social Network Analysis (SNA) can be used to identify key nodes in the network using multiple centrality measures [4]. One of the most used measures is degree centrality, which measures the number of direct neighbours connected to a given node. Thus, the higher the degree centrality the more critical this node is to the network as it could represent a hub for information and resource flowing within the network. Networks with a few centralised hubs are commonly referred to as centralised networks. Launching a random attack on such networks in an attempt to disrupt their behaviour will most likely result in targeting exterior nodes, and thus not having a significant impact on the network. However, targeting the identified hubs is likely to have significant impact and will disconnect the network, which could result in an effective disruption of, for instance, a criminal group's activities. Another important centrality measure is betweenness centrality, which measures the number of shortest paths that pass through a given node. This measures the indirect contacts that surround a particular node, which gives an indication of the importance of this node for the information flow in the network. These are examples of measures that can be useful in achieving our aims to identify key players in a given network of interest.

Borgatti argues that in order to effectively measure centrality, one should have an understanding of why centrality is important [16]. He distinguishes between two different reasons to measure centrality and identify key players in a network. The first reason is to maximize disruption to a given network, while the second reason is to maximize collection of information. Additionally, Borgatti recognises that typically effective targeting of central nodes is often done on multiple nodes in the network rather than a single central node. One limitation of Borgatti's model is that it does not incorporate weights for nodes nor links when measuring centrality. This was identified by Schwartz *et al.* [11] and they built on it and incorporate both node weights and link weights into Borgatti's approach.

SNA techniques have been utilised to understand and disrupt several organised groups and networks such as terrorists, arm dealers, drug smuggling, and wildlife trafficking networks [12]. Several data sources are used to construct the criminal network. Examples include intelligence from law

enforcement, police arrest reports, and courts sentencing. Bright et. al., used judges' sentencing comments to conduct SNA in order to understand criminal networks in Australia. The aim of their study was to examine whether SNA based on judges' sentencing comments can provide insights into the structure and operation of Australia's methamphetamin market [10]. Furthermore, L'Huillier *et al.* used SNA techniques in addition to text mining in order to identify key-members in virtual communities of interests, where members share common interest in particular topic [17]. Another important research track that aids in identifying key influential individuals within organised groups is tie strength. This concept was first presented by Granovetter when he discussed the strength of weak ties [18]. Following this, other work in the literature focused on incorporating his theory to study and predict tie strength between users in social networks [19]. This is particularly important when analysing organised groups to identify trust relations between pairs of users and identify the strong ties connected to the key players in a given network.

Prior studies have put forward several initiatives to identify key individuals and group leaders [11], [20], [21]. Some used SNA metrics and algorithms, such as centrality measures and Page Rank algorithm [22], while others used text mining techniques [23]. However, few studies tried to combine several techniques to identify key and influential members in suspicious groups [24]. In our study, we adopt a hybrid approach as we utilise several SNA metrics to identify key players in the group, then we perform text sentiment analysis and temporal analysis to gain further insights.

III. Methodology

In order to identify key players in a given social group, several SNA metrics can be utilised. In graph theory, centrality measures identify the most important vertices in a graph [6]. The simplest and most obvious centrality measure that can help identify the most important node in a graph is Degree Centrality (DegCent) [12]. As mentioned earlier, DegCent measures the number of direct links a given node has. The higher the DegCent of a node, the more connected the node is in the network. While DegCent focuses on the number of direct connections a node has, it does not consider the global localization the node has in the network and does not take in consideration how powerful those connections are in the graph. Thus, another centrality measure that focuses on the influence of the connected nodes is Eigenvector Centrality (EigenCent). EigenCent concentrates more on identifying the most connected intermediary [25]. A node that has a high eigenvector score is one that is adjacent to nodes that themselves have high scores [6]. Additionally, Betweenness Centrality (BetCent) is a measure that indicates the intermediary with the most control over information flow. In our empirical study, we apply these centrality measures to an activist dataset to identify the most connected, and

influential users in each activist social group. We summarise our methodology as follows:

1) Pre-process the data in order to create different graphs, such as a bipartite graph for users and their created posts, and a directed user-to-user relation graph.
2) Apply SNA techniques to the data to identify the most important and influential users in the network based on the different centrality metrics.
3) Use sentiment analysis on users' posts and comments and assign to each post either a positive or negative value. Then we label each user as having a negative or positive influence on the group based on the average sentiment of all their posts.
4) Apply clustering algorithms to identify potentially hidden communities within the larger network dataset.
5) Perform analysis of the network posting structure over time and correlate observed trends with related real-world events.

IV. Online Activist Dataset

The online activist dataset that was used for our study consists of Facebook posts from a UK-based activist group in the Facebook online social network. For ethical reasons we are not able to mention the group's name but will refer to it as FB-Activist Group (FB-AG). The group is mainly focused on sharing information related to their cause, promoting for several movements and organising protests. The FB-AG dataset incorporates 670 Facebook group posts posted by 274 users. These posts include 370 original posts and 300 comments on those posts. Note that our dataset did not contain any replies to comments on original posts. For each post, we have access to the number of likes, number of comments, and number of shares each post received. Additionally, the posts are time-stamped which allows us to reconstruct the time line of the posts. Moreover, the dataset contains generic information about the users who created each post. This information includes, user ID, age, gender, and location. The dataset was collected from the period of March to May, 2015.

V. Analysis of The Facebook Activist Group

A. Social Network Analysis

1) Research Questions: Before we demonstrate how we apply SNA metrics to identify key players in the network, we provide the research questions that we aim to answer in our analysis:

1) By analysing the network, can we identify the most active, and most influential user(s)? Are they the same?
2) Can we identify sub-communities within the network, and predict tie strength between members?
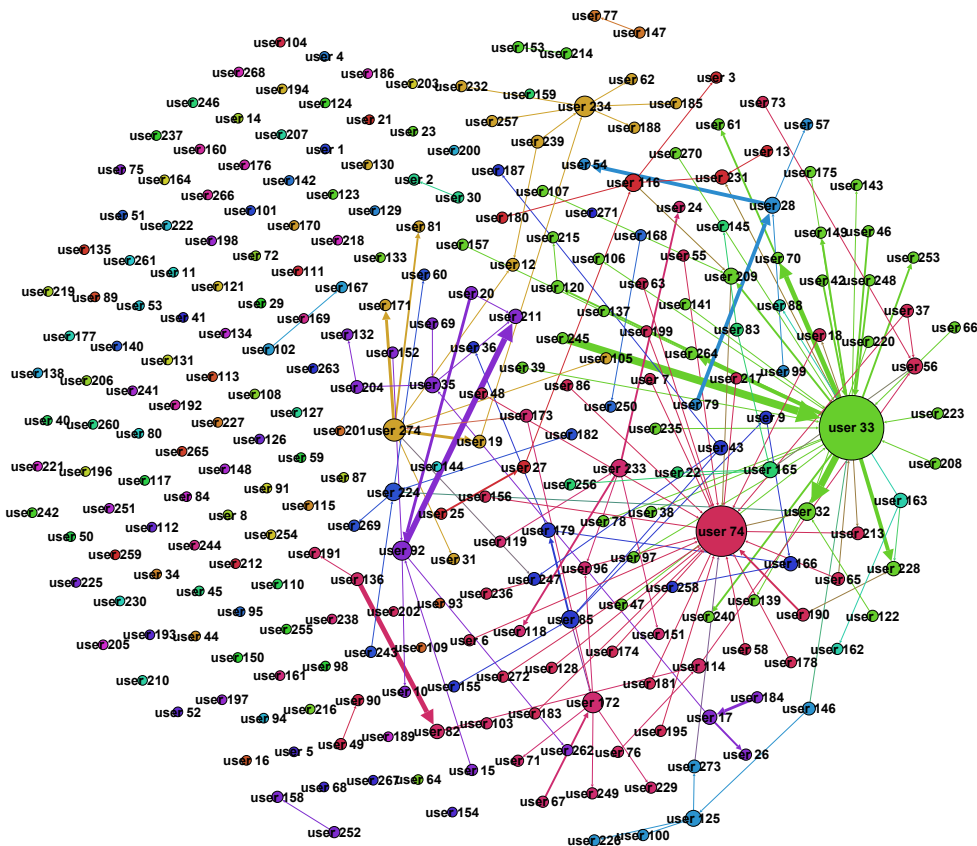3) How does the network grow over time? Is it getting more dense or sparse?

Figure 1. The Facebook-AG Comments Network. In this network, the node size represents the nodes' degree centrality measure and the nodes' colour represents the detected communities or clusters. Edge thickness represents the frequency of communication between nodes and the arrows represent the direction of communication. Here the number of nodes (N) is 274 and number of edges (E) is 162.

To be able to answer the above questions, we construct different graphs from the FB-AG dataset. First, we generate a user-to-user relationship graph. In this graph we say there is a relation between two users if one of them comments on the other's post. Formally, we define the graph as a directed weighted graph $G = (U, E)$, where U represent the users who initiated a comment action, and E represent the edges set. An edge e exist between two user nodes A and B, if user A comments on a post published by user B. The edge weight w represents the frequency of interaction between the two users. We will refer to this graph as the Comments Network and it is shown in Figure 1.

2) Applying SNA Metrics: In order to analyse the topology of the Comments Network we first introduce a number of key SNA metrics. The five most important measures for network topology are [26]:

Centrality: As described in Section II, there are several centrality measures and each gives a different interpretation to the network. We apply the following centrality measures: Degree centrality (DegCent), to identify how many people this user can reach directly; Betweenness centrality (Bet-Cent), to identify how likely this user is to be the most direct route between two other users; Eigenvector centrality (EigenCent), to measure how well this user is connected to other well connected users.

Network Density: This is defined as total number of actual connections in the network divided by total number of possible connections. This measure provides insight into how effective information spreads among the network users. In the Comments Network, the network density is 0.002, which is considered low. This suggests that information does not spread effectively in the activist group as peripheral nodes depend on their connection to other central nodes to receive information.

Average Degree: This measures the number of average connections a node has in the network. Again the average degree for a node in the Comments Network equals to 1, which is considered quite low as the network consist of 274 users. This suggests that only few users have high degree measure and many users comment on their posts. This matches our previous observation that users in the activist group are not well connected to each other.

Average Path Length: This is the average value of the number of shortest paths between all pairs of nodes. This measure represents the average number of steps it takes to get from one node in the network to another. For the Comments Network, on average indirectly connected users can reach each other in around 2 steps. This means that information can spread in the network fairly quickly if it goes through one of the hubs in the network.

Network Diameter: This is the longest of all the calculated shortest paths in the network. In the Comments Network, the diameter equals to 5. In other words, within the FB-AG network the longest distance it takes for information to flow between two users in the network is 5 hops.

Measures	Values
Number of Nodes	274
Number of Edges	162
Average Degree Centrality	1.000
Average Path Length	2.012
Density	0.002
Diameter	5
Number of weakly connected components	123
Number of strongly connected components	274
Modularity	0.774
Number of communities	131

We apply each of these measures on the Comments Network and summarise the results in Table I. The network is visualised in Figure 1, where the nodes' colours represent the detected communities in the graph. We used a modularity algorithm [27] to measure the network modularity and partition the network into communities. Additionally, the nodes' size represent the DegCent measure. The bigger the node the more active the associated user is in the network. The edge thickness represents the frequency of interaction between any given two nodes. Moreover, the Comments Network shows that there are a number of isolated nodes without any incoming or outgoing edges. This represents users who have created posts in the Facebook group but did not receive any comments on them. Similarly, those users did not engage in any discussion in the group since they do not have any outgoing edges.

Furthermore, we can see from the graph that there are several communities of small size (two nodes). Clearly, those small disconnected communities are not relevant to our analysis of influential users therefore we filter them out to get the more concise graph presented in Figure 2. It is this graph that we use for further analysis.

Identifying Influential Users – Previous research has identified that network robustness may depend among other factors on high degree nodes (i.e., hubs) [28]. Targeting those hubs will lead to the destruction of the network into sub-networks, thus, losing its power and effectiveness. This is true if the network is scalefree. This means that if the network degree distribution follows a power law, then identifying hubs in the network will aid in interrupting the flow of information and destroy the network [13]. We measure the network degree distribution of the Comments Network and find that it follows a power law distribution realised by the heavy tailed distribution as shown in Figure 3.
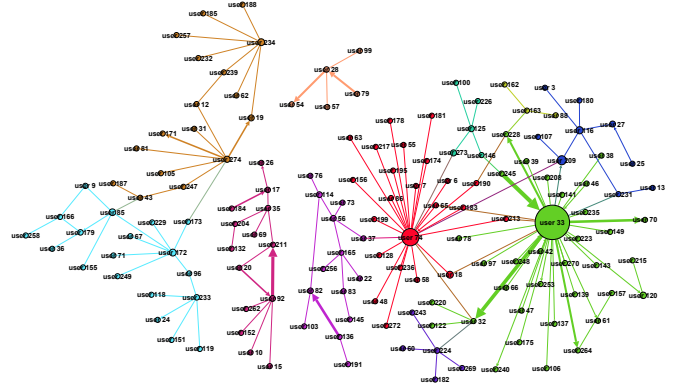

Figure 2.   Facebook-AG Filtered Comments Network. The nodes size represent the betweenness centrality measure.

In the context of our study of activist groups, this means that if we aim to disrupt the flow of information in the Comments Network, we would target nodes with high degree centrality. For example, isolating the users with the highest DegCent would most likely result in preventing the activist group from organising large protests.
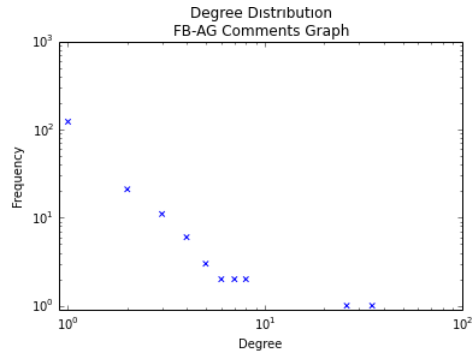

Figure 3.   Facebook-AG Comments Network Degree Distribution.

We can see in Figure 2 that FB-AG network consist of four disconnected components of varying sizes. The largest component (located in the right side of the graph) has two main central nodes (User 33, and User 74). These two nodes have the highest DegCent and BetCent measures in the entire network. It is interesting though, that these two hubs do not have any direct links to each other. They are only connected through indirect links. One explanation for this is that the two users are the leaders of the activist group each within their own small community. Both have the responsibility to spread messages and answer members questions, which does not require them to comment on each other's posts.

When using SNA metrics, it is important to combine different measures to be able to draw conclusions about the network [29]. Moreover, based on the nature of these groups we assume that there are additional influential people hidden within the network. To discover these individuals, we apply additional SNA metrics to the network. EigenCent is an important measure that can identify how well a user is

connected to other highly connected people (See Figure 4). This measure reveals that User 32 has the highest eigenvector centrality value followed by User 213. This shows that both of them have a powerful position in the network as they are mainly connected to other well connected users. In other words, User 32 and User 213 have a strong relation with User 33 and User 74.
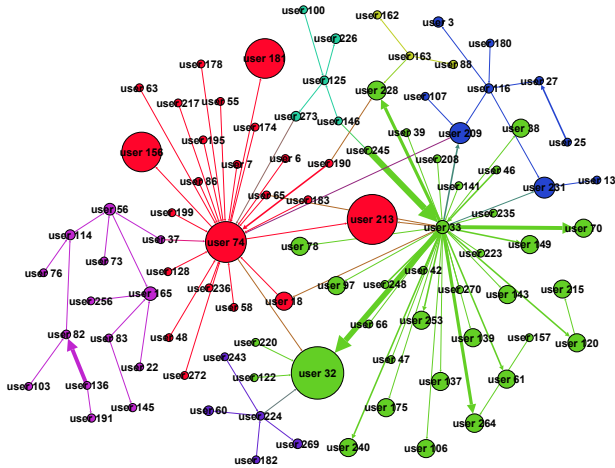

Figure 4. Facebook-AG Comments Network- Nodes size represent the nodes Eigenvector Centrality

Moreover, additional interesting metrics that revealed more insights to identify influential users in the network are PageRank scores [30] and HITS Authorities and Hubs [31]. By applying the PageRank algorithm to our network the following users scored the highest: User 74, User 33, User 165, and User 156. The first two users are expected to show up in the top score list as they have the highest DegCent in the network. However, another two users were identified by PageRank algorithm as highly important users. The importance of applying PageRank algorithm is that it takes in consideration not only the quantity of links between nodes (like DegCent), but also the links' qualities. Furthermore, the HITS authority score showed that User 74 and User 33 are the most authoritative in the network (many users commenting on their posts), while the HITS hub scores showed that User 74 and User 209 are the top hubs (they comment heavily on other users' posts). Table II summarises the top three ranked users for each SNA measure.

Table II
TOP 3 INFLUENTIAL USERS IN DIFFERENT SNA METRICS.

Metric	First Rank	Second Rank	Third Rank
Betweenness Centrality	User 33	User 74	User 209
Eigenvector Centrality	User 32	User 213	User 74
PageRank	User 74	User 33	User 32
HITS - Authority	User 74	User 33	User 32
HITS - Hub	User 74	User 209	User 234

From the table it is clear that User 74 has the highest

influence in the network followed by User 33. Moreover, User 32 and User 209 seem to have important role in the network. For instance, User 32 scored highest in EigenCent and third in PageRank and Authority. He has a strong tie with User 33 which is realised by the high number of comments from User 33 to User 32. Additionally, he has a tie with User 74 although it is not as strong as the other one. On the other hand, looking at User 209 position in the graph, we can see that the user acts as a middle connection point between the two most influential users (User 33 and User 74). Similarly, based on the community detection algorithm, User 209 is the linking node between three different communities which makes him a critical node for the spread of information within those different communities. This information can be useful for authorities who aims to disrupt the activist group network by targeting a minimum number of users.

Analysing Tie Strength – To be able to analyse and predict tie strength between users in the activist group we rely on several variables. The most intuitive and obvious tie strength variable is the frequency of interaction between the linked users. If two users constantly comment on each other's posts we might assume that they have a strong relation. Identifying tie strength between users is particularly important as it provides insight into the social structure of the activist group. In most social groups, users who have strong relations with leaders of the group are considered influential themselves. Thus, identifying strong ties can aid in identifying key influential members of the group.

There is a number of SNA measures that can aid in better understanding relations between nodes and the level of trust between them. One of these measures is neighbourhood overlap, which is defined as a ratio as follows [32]: The neighbourhood overlap of an edge (A, B) is:

$$\frac{number\ of\ nodes\ adjacent\ to\ both\ A\ and\ B}{number\ of\ nodes\ adjacent\ to\ at\ least\ one\ of\ A\ or\ B}$$

Neighborhood overlap measures closeness to being a local bridge. The numerator in the above definition of neighbourhood overlap represent another important measure called edge embeddedness. Both of these measures are considered an indication of the level of trust in the relation between A and B [32]. Moreover, previous literature in the sociology field [33]–[35] declare that if two users are connected by an embedded link then they will more likely trust each other and have increased confidence in their relations.

We applied those two measures on our network to detect and understand the potential level of trust between the users in the Facebook activist group. The results showed that for neighbourhood overlap and edge embeddedness, the highest scoring links were between User 20 and User 211, User 20 and User 92, and User 92 and User 211, respectively. These three links form a triangle between the three users which may suggest strong trust relation between them. All other
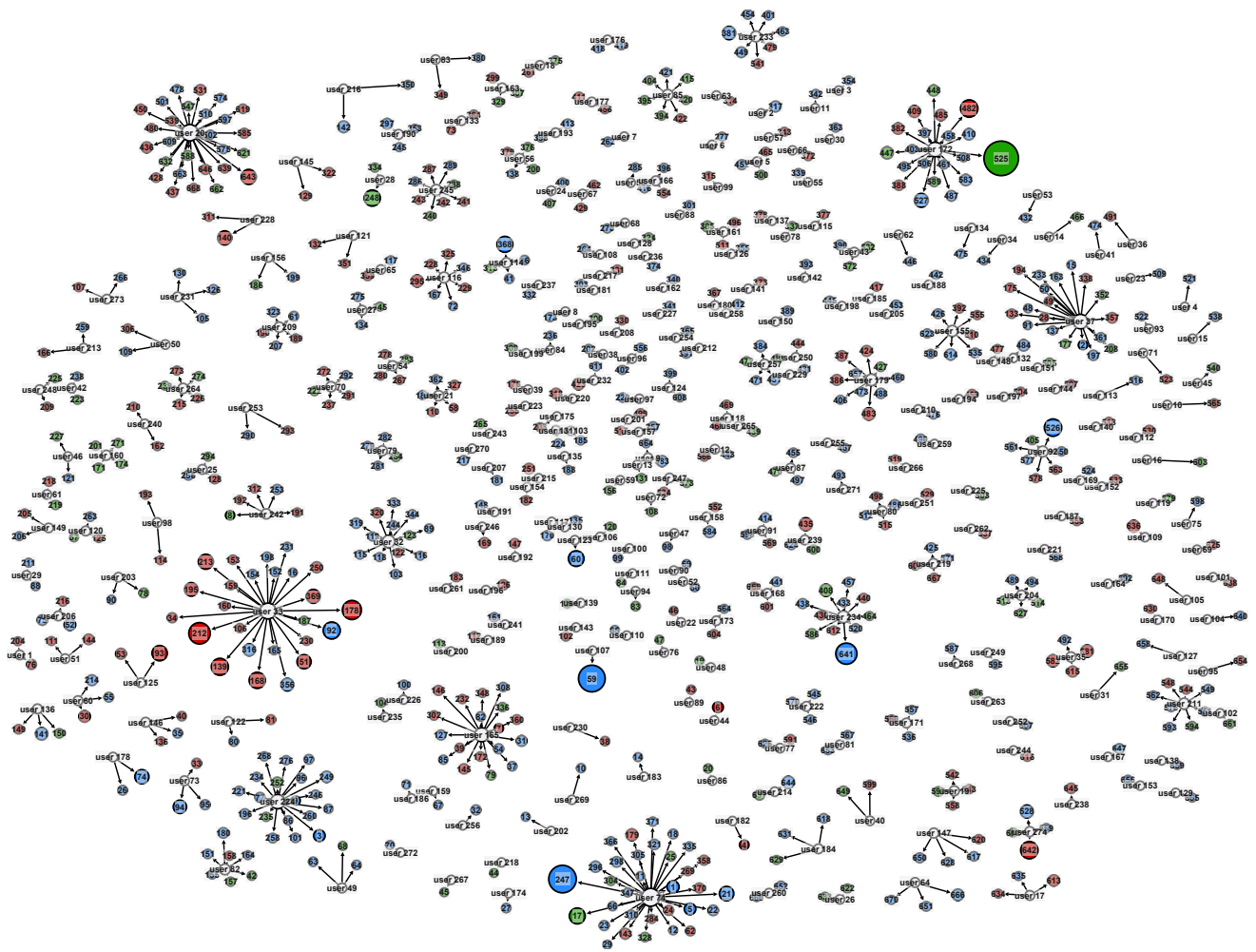
Figure 5. User-Post bipartite graph for the FB-AG group. Users are visualized as white nodes where the size represent how active the user is. Posts are coloured (red, green, blue) based on the sentiment analysis of the text (negative, positive, neutral, respectively).

links in the graph scored zero in both measures, which suggests that they are local bridges in the graph [32]. According to previous studies that focused on understanding social collective behaviour, local bridges are considered powerful in spreading awareness and general information. However, they are not as effective in spreading strong behaviour such as political or protesting movements [36]. This means that in the FB-AG network, if one of the three users (User 20, User 211, User 92) calls for a protest or a movement then it is more likely that the other two users will join the movement.

Another important measure to study is edge betweenness centrality, which computes the degree to which an edge makes other connections possible. The results showed that the following links had the highest betweenness measure in the graph: User 33 and User 18, User 18 and User 74, User 74 and User 156, User 209 and User 74, and User 74 and User 181. This means that those users relations facilitate the flow of information within the network. Moreover, they act as bridges to connect different communities of highly interconnected people. Using this information, law enforcement can disrupt the network and stop a riot for example by

targeting these particular links.

B. Sentiment Analysis

Now that we have an understanding of the most influential users in the network, we construct a second graph from the dataset which is more focused on analysing user-to-post relationship. We generate a bipartite graph $biG = (U, V, E)$ consisting of two node sets (U and V), and an edge set E which contains the edges of the graph. The two sets U and V represent two types of nodes, *Users* and *Posts*, respectively. An edge e between a given user u and a post p is formed, if user u created the post p. We will refer to this graph as the user-post graph.

Figure 5 shows the bipartite User-Post graph for FB-AG group. The *Users* are represented by a white colour while *Posts* are presented in either blue, red, or green colours. The colour code for the nodes in the *Posts* set represents the result of sentiment analysis on the posts' text. We used automated tools to calculate the sentiment score for each post and classify it as a negative, positive, or neutral sentiment [37]. A post is assigned a negative sentiment if it contains aggressive

language, or calls for protests and is presented in a red colour. A post with a positive sentiment is one that contains positive and encouraging text, and is depicted by green colour. Finally, a post is assigned a neutral sentiment if it contains generic content such as questions, URLs, images, etc. and it is depicted by blue colour. The size of the nodes in the *Users* set represent how active the user is in the network (i.e., the Out Degree). On the other hand, the size of the nodes in the *Posts* set represent the total number of likes, comments, and shares that post received divided by the total number of group members. We will refer to this measure as the post-engagement level.

We can see from the User-Post graph in Figure 5 that there are ten main active hubs in the group. The top three most active users are User 74, User 20, and User 33. User 74 has mostly neutral sentiment posts (i.e., has neutral influence on the group), while User 20 and User 33 usually post negative sentiment posts (i.e., have negative influence). A summary of the top ten users is presented in Table III along with the distribution of posts' sentiments. This finding supports our previous conclusion from the analysis on the Comments Network that User 74 and User 33 are the leaders of the group. It also adds that User 33 is more inclined to organise or lead a protest with User 20, while User 74 role is more related to answering questions or posting neutral content that can be related to raising awareness to the group's cause.

Table III
USER-POST BIPARTITE GRAPH: 10 MOST ACTIVE USERS

Number of Nodes: 944 , Number of Edges: 670				
User ID	Out Degree	Negative Sentiment	Positive Sentiment	Neutral Sentiment
User 20	27	13	5	9
User 32	13	2	1	10
User 33	25	15	1	9
User 37	20	7	3	10
User 74	32	8	4	20
User 165	18	9	2	7
User 172	20	5	4	11
User 179	10	4	1	5
User 224	19	0	2	17
User 234	11	3	3	5

C. Time Analysis

In this section, we perform a temporal analysis of the network's posting and commenting structure. This analysis will aid in identifying hidden activity patterns in the dataset. Figure 6 shows the group posting activity levels per day. On average, the group posts around 10 posts everyday. Some days had zero posting activity, while other days had a spike in the number of posts to a maximum of 56. It is important to mention here that even though some days had zero posts this does not mean that the group was completely in active. In fact, this means that no new posts were created, but a like or a share action of an existing post may have occurred.

As shown in the graph in Figure 6, there were three time periods that had an increased activity level with the highest
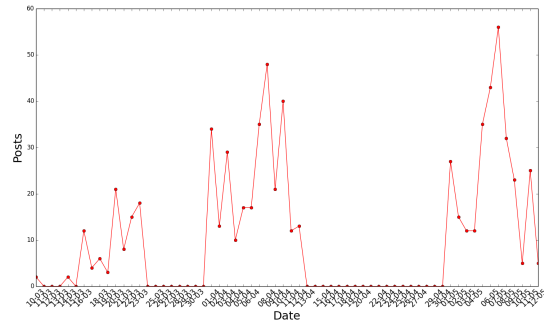

Figure 6.   Facebook Activist Group posts per day

being in around May 7th. This is likely to be correlated with the fact that on this particular day there was the United Kingdom general election, which is of interest to this particular activist group as it would have an impact on their cause. Additionally, the spike occurring in March 20 coincides with a call for a national day of action called for by other activist groups who share the same interest.

Moreover, there is a clear pattern in the increase and decrease of the group activity levels across the three-month period. Therefore, we wanted to investigate who are the users who post just before the spikes in the activity occur. This might give us insight into identifying influential users, for instance if there are a specific group of people who tend to always post before a spike in the activity occurs. We identified three main dates in each month where the activity was at its maximum level (March 20, April 8, and May 7). Then, we gathered all users who posted during the three previous days. Our assumption here is that the three days leading up to each peak may serve as the reason for the spikes. Our results showed that User 74 and User 37 were consistently active before most spikes. This supports our previous results regarding the high influence of User 74, and allows us to infer that User 37 has a similar high influence.

VI. RESULTS VALIDATION AND DISCUSSION

In the previous section we used several SNA metrics, sentiment and temporal analysis to understand and identify the most influential people in a group of organised activists. Our results from each type of analysis yielded a set of users as most influential. Some of these results overlapped, which indicates that the identified users possess a strong influence over the group. In such research, it is usually difficult to get access to the ground truth in order to validate the results. In our case, we were able to work with a security consultancy company to acquire the ground truth about the activist group since they had more comprehensive intelligence information about them. Their ground truth was based on monitoring the group activities online as well as offline information.

They have identified seven users as being the most influential in the group. By comparing our results to the ground truth we find that through our analysis we were able to

identify five (User 74, User 33, User 20, User 37, and User 179) of those seven users. The remaining two users did not have an active presence in our dataset which explains why they were not flagged in our analysis. Additionally, through our analysis we identified in total eight important links, of these we found a strong relation between User 20 (one of the influential members), and User 211 and User 92. Although the latter two users were not flagged in our analysis as influential, their strong connection to User 20 could be grounds for an increased importance.

In our study, we were able to understand the dynamics of interactions between users in the group by applying SNA metrics, such as network diameter, average paths length, and degree centrality. Moreover, we identified the key-players in the group by using different centrality measures (DegCent, BetCent, EigenCent) and tie strength analysis (Neighbourhood overlap, Embeddedness, Edge Betweenness). Through our analysis, we found that the most active users are not necessarily the most influential. Additionally, by applying sentiment analysis on users' posts we were able to identify which users have positive and negative influence on the group. This can help law enforcement to focus on monitoring key users with negative influence on the group to be able to predict future hostile protests.

Furthermore, we were able to identify sub-communities in the activist group as well as predict trust relation and tie strength between members through several SNA methods. Additionally, we performed time analysis that showed several increases and decreases in the group posting activity levels. Some of these increases coincide with multiple real-life events such as UK general elections, and calls for national day of action and protests.

It is important to acknowledge here that the size of our dataset is relatively small, which makes it difficult to generalise the findings. Thus, further analysis is needed on larger datasets that cover longer periods of time. We are currently pursuing this for our next line of research.

VII. Conclusion and Future Work

In this paper, we explored the use of different SNA methods in order to identify key and influential individuals within online activist groups. Since online social networks have gained increased popularity, they have been adopted by several organised groups as a medium for fast communication and wide spread of their messages. This gives us, the research community, a chance to analyse the behaviour of these groups. We analysed the dynamics of the interactions in a Facebook activist group using SNA and time-analysis techniques. This allowed us to identify the most active and most influential members. Additionally, we adopted sentiment analysis to identify users with both positive and negative influences on the group. Trust relations were investigated using link analysis methods. We validated our results against a ground truth, which showed that we were able to identify the main influential people based on structural properties of their online behaviour.

For our future work, we aim to evaluate our methods in a more extensive study with a larger sample collected over a longer period of time. Additionally, we aim to compare the behaviour of similar groups across different social media e.g., Facebook and Twitter. This may help in identifying additional relations between members. Similarly, we aim to compare the interaction-based relations to the actual social (friendship) relations of the users. This may allow us to identify hidden friendships, where users try to cover their friendship by not creating a direct link between them. This type of analysis is valuable to the field of cybercrime investigation, particular in combination with work such as [38], [39], where we do not have enough information about the group hierarchy, however, may have access to their communication.

References

[1] H. Alexander, "Tweeting terrorism: How al Shabaab live blogged the Nairobi attacks," September 2013, last accessed 25-June-2015. [Online]. Available: www.telegraph.co.uk/news/worldnews/africaandindianocean/kenya/10326863/Tweeting-terrorism-How-al-Shabaab-live-blogged-the-Nairobi-attacks.html

[2] Canadian Centre for Intelligence and Security Studies, "A framework for understanding terrorist use of the internet," Integrated Terrorism Assessment Centre (ITAC), Tech. Rep. Volume 2006-2, 2006.

[3] E. Otte and R. Rousseau, "Social network analysis: a powerful strategy, also for the information sciences," *Journal of Information Science*, vol. 28, no. 6, pp. 441–453, 2002.

[4] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, M. Granovetter, Ed. Cambridge University Press, 1994.

[5] A. Basu, "Social network analysis: A methodology for studying terrorism," in *Social Networking*, ser. Intelligent Systems Reference Library, M. Panda, S. Dehuri, and G.-N. Wang, Eds. Springer, 2014, vol. 65, pp. 215–242.

[6] S. P. Borgatti, "Centrality and network flow," *Social Networks*, vol. 27, no. 1, pp. 55–71, 2005.

[7] P. Williams, *Transnational Criminal Networks. Networks and Netwars: The Future of Terror, Crime, and Militancy*, 2007.

[8] V. Validakis, "ANZ HQ in lock-down as fossil fuel protest gets aggressive," October 2014, last accessed 25-June-2015. [Online]. Available: http://www.miningaustralia.com.au/news/anz-hq-in-lock-down-and-climate-protest-gets-aggre

[9] C. Latimer, "Coal Protestors Convicted," July 2013, last accessed 25-June-2015. [Online]. Available: http://www.miningaustralia.com.au/news/coal-protestors-convicted

[10] D. A. Bright, C. E. Hughes, and J. Chalmers, "Illuminating dark networks: a social network analysis of an australian drug trafficking syndicate," *Crime, Law and Social Change*, vol. 57, no. 2, pp. 151–176, 2012.

[11] D. Schwartz and T. Rouselle, "Using social network analysis to target criminal networks," *Trends in Organized Crime*, vol. 12, no. 2, pp. 188–207, 2009.

[12] T. C. Haas and S. M. Ferreira, "Federated databases and actionable intelligence: using social network analysis to disrupt transnational wildlife trafficking criminal networks," *Security Informatics*, vol. 4, no. 2, 2015.

[13] P. A. C. Duijn, V. Kashirin, and P. M. A. Sloot, "The relative ineffectiveness of criminal network disruption," *Sci. Rep.*, vol. 4, 02 2014.

[14] H. B. Milward and J. Raab, "Dark Networks as Organizational Problems: Elements of a Theory," *International Public Management Journal*, vol. 9, no. 3, pp. 333–360, 2006.

[15] A. J. Reiss, "Why Are Communities Important in Understanding Crime?" *Crime and Justice*, vol. 8, 1986.

[16] S. Borgatti, "Identifying sets of key players in a social network," *Computational and Mathematical Organization Theory*, vol. 12, no. 1, pp. 21–34, 2006.

[17] G. L'Huillier, H. Alvarez, S. A. Ríos, and F. Aguilera, "Topic-based social network analysis for virtual communities of interests in the dark web," *SIGKDD Explor. Newsl.*, vol. 12, no. 2, pp. 66–73, 2011.

[18] M. S. Granovetter, "The Strength of Weak Ties," *American Journal of Sociology*, vol. 78, no. 6, pp. 1360–1380, 1973.

[19] E. Gilbert and K. Karahalios, "Predicting tie strength with social media," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '09. New York, NY, USA: ACM, 2009, pp. 211–220.

[20] B. Hoppe and C. Reinelt, "Social network analysis and the evaluation of leadership networks," *The Leadership Quarterly*, vol. 21, no. 4, pp. 600 – 619, 2010, leadership Development Evaluation.

[21] P. Duijn and P. Klerks, "Social network analysis applied to criminal networks: Recent developments in dutch law enforcement," in *Networks and Network Analysis for Defence and Security*, ser. Lecture Notes in Social Networks, A. J. Masys, Ed. Springer, 2014, pp. 121–159.

[22] J. Heidemann, M. Klier, and F. Probst, "Identifying key users in online social networks: A pagerank based approach." in *ICIS*. Association for Information Systems, 2010, p. 79.

[23] S. A. Ríos and R. Muñoz, "Dark web portal overlapping community detection based on topic models," in *Proceedings of the ACM SIGKDD Workshop on Intelligence and Security Informatics*, 2012, pp. 2:1–2:7.

[24] C. C. Yang, X. Tang, and B. M. Thuraisingham, "An analysis of user influence ranking algorithms on dark web forums," in *ACM SIGKDD Workshop on Intelligence and Security Informatics*, ser. ISI-KDD. ACM, 2010.

[25] T. Opsahl, F. Agneessens, and J. Skvoretz, "Node centrality in weighted networks: Generalizing degree and shortest paths." *Social Networks*, vol. 32, no. 3, pp. 245–251, 2010.

[26] R. A. Hanneman and M. Riddle, *Introduction to social network methods*, University of California, 2005.

[27] V. Blondel, J. Guillaume, R. Lambiotte, and E. Mech, "Fast unfolding of communities in large networks," *J. Stat. Mech*, 2008.

[28] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," *Nature*, vol. 406, no. 6794, pp. 378–382, 2000.

[29] S. S. Everton, "Tracking, destabilizing and disrupting dark networks with social networks analysis," *Dark Networks Course Manual*, 2008.

[30] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," in *Seventh International World Wide Web Conference*, 1998.

[31] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *J. ACM*, vol. 46, no. 5, pp. 604–632, Sep. 1999.

[32] E. David and K. Jon, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. New York, NY, USA: Cambridge University Press, 2010.

[33] M. Granovetter, "Economic action and social structure: The problem of embeddedness," *American Journal of Sociology*, pp. 481–510, 1985.

[34] S. Coleman and J. S. Coleman, "Social capital in the creation of human capital," *Amer. J. Sociol*, pp. 95–120, 1988.

[35] J. S. Coleman, *Foundations of social theory*. Belknap Press of Harvard University, 1990.

[36] D. Mcadam, "Recruitment to High-Risk Activism: The Case of Freedom Summer," *The American Journal of Sociology*, vol. 92, no. 1, pp. 64–90, 1986.

[37] J. Perkins, "Natural Language Text Processing APIs," 2010, last accessed 06-Sept-2015. [Online]. Available: http://text-processing.com/docs/sentiment.html

[38] S. Creese, M. Goldsmith, J. R. C. Nurse, and E. Phillips, "A data-reachability model for elucidating privacy and security risks related to the use of online social networks," in *11th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*. IEEE, 2012, pp. 1124–1131.

[39] J. R. C. Nurse, J. Pumphrey, T. Gibson-Robinson, M. Goldsmith, and S. Creese, "Inferring social relationships from technology-level device connections," in *12th International Conference on Privacy, Security and Trust*. IEEE, 2014, pp. 40 – 47.