Lunds Universitet
Statsvetenskapliga institutionen

UNDK02
HT 2023
Handledare: Tony Ingesson

En Social Nätverksanalys av ett Hackerforum

Albin Erlander

Abstract

The spread and societal dependence on the internet has increased the threat posed by criminal hackers. While often thought of as lone wolfs, hackers often exist in communities and a widespread subculture. Due to the nature of hacker activity, these communities are often situated online, such as in forums, which provide an excellent opportunity for observation and study.

An increased threat level brings an increased need for protection and countermeasures, but also for cyber threat intelligence (CTI) to indicate what groups, people, and tools could come to pose a threat. Social network analysis has the potential to play a key role in in transforming CTI into something more proactive than it is today.

The goal of this study was to use SNA to map and analyse an online hacking forum to determine its general and structural properties, as well as investigate its central members. The forum was determined to be comparatively large, cohesive, and decentralized. One user stuck out as most central in terms of all metrics but one, some ranked high for several metrics, and some where only standing out in some respects, all with implications for their characteristics.

Nyckelord: Social Nätverksanalys, SNA, Hackerforum, Cyberhotsunderrättelser
Antal ord: 10 044

Innehållsförteckning

1 Inledning

1.1 Introduktion

Spridningen av digital teknik och det samhälleliga beroendet av internet har signifikant ökat hotet från kriminella hackers. Det finns en vanlig uppfattning om att hackers och cyberkriminella är antisociala ensamvargar, men ofta existerar de i en egen gemenskap och subkultur i högre utsträckning än vad den allmänna bilden antyder (Holt et al. 2012). Det är en på vissa sätt paradoxal värld, där meritokratins tendenser mot att bevisa sig och skryta blandas med hackerkulturens betoning på sekretess och anonymitet (Holt et al. 2012; Jordan & Taylor 1998). Dessa grupper existerar även fysiskt, men kanske framförallt online i forum och chattrum, vilket kan innebära goda förutsättningar för att studera dem.

Ett växande hot innebär ett växande behov av skydd och motåtgärder, men också av cyberhotsunderrättelser (engelska; Cyber Threat Intelligence, CTI) om individer och verktyg som utgör seriösa hot. Forskningen om hackers och hackergrupper online med social nätverksanalys (SNA) och andra metoder kan vara en nyckel i att transformera framställningen av cyberhotsunderrättelser till något mycket mer proaktivt än det är idag (Grisham et al. 2017).

1.2 Syfte

Syftet med denna studie är att undersöka ett hackerforum med hjälp av tekniker från SNA. Detta genomförs för att ge en förståelse och inblick i hur ett sådant forum kan se ut, både mer övergripande och strukturella aspekter samt vad som karaktäriserar centrala individer i nätverket. Studien ämnar inte att definiera eller pröva generalisande teori, utöver användningen av SNA, men förhoppningen är att en detaljerad analys av ett hackerforum i framtiden kan användas som en del i en bredare förståelse av dessa.

1.3 Frågeställning

Frågeställningen för denna studie är tvådelad, delvist fokuserat på det övergripande nätverket, och delvist fokuserad på individuella aktörer i nätverket. Frågeställningarna lyder:

1. Vilka övergripande och strukturella egenskaper har nätverket?
2. Vilka individer framträder som mer signifikanta och vad karaktäriserar dem?

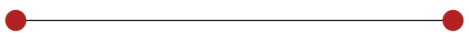
2 Teori: Social nätverksanalys

SNA innefattar en stor mängd specifika metoder, mått och tillvägagångsätt. Den teori som presenteras i denna del är de som används i metoden för denna studie och har valts ut med hänsyn till studiens frågeställning och omfång.

Teorin är uppdelad på fyra delar, där den första introducerar de grundläggande SNA-begrepp som behövs för resterande teori. Del 2.2 och 2.4 fokuserar på ett nätverks övergripande respektive strukturella egenskaper och ligger till grund för den första delfrågeställningen. Del 2.3 beskriver teorin som användas för besvarandet av andra delfrågan.

2.1 Grundläggande begrepp

Inom SNA modelleras individuella entiteter som *noder*, och relationer mellan dem som *kanter* (eller *länkar*) (Bichler 2019, s. 16) i strukturer som kallas för *grafer* som betecknas med G. För denna rapports syften tolkas *grafen* G som syftande till den matematiska strukturen eller illustrationen av *nätverket*, som är den faktiska konstellationen av individer. För att förtydliga resonemangen används därmed främst begreppet nätverk (eller forum) i denna studie, då det sällan är nödvändigt att hänvisa till endast den teoretiska representationen av nätverket och inte nätverket i sig. Samma distinktion gäller för skillnaden mellan användare eller individ respektive nod, samt länk eller koppling respektive kant. Se figur 1 för en väldigt enkel graf (en så kallad diad) med två noder, kallade *grannar* (Bichler 2019, s. 19), och en mellanliggande kant. I denna rapport kommer mängden av alla noder att refereras till som V, och en individuell som v_i (engelska; *vertex*), där i syftar till nodens index (till exempel ges den första noden beteckningen v_1). Motsvarade notation för kanter är E respektive e_{ij} (engelska; *edge*). En kant är alltid associerad med två noder som kanten sammanbinder, vilkas index ges av i och j.



Figur 1: Två noder (grannar) med en mellanliggande kant.

I denna rapport kommer inga rekursiva kanter (ibland kallat själv-loopar), kanter som börjar och slutar i samma nod (Bichler 2019, p. 28), att konstrueras då dessa inte anses beskriva något meningsfullt inom ramen för denna studie. Detta kommer att bli tydligare i del 4. I vissa fall kan kanter ges en riktning och en vikt. I denna studie kommer riktade kanter inte att användas av samma anledning som rekursiva kanter utesluts. Viktade grafer (grafer med viktade kanter) kan användas för att tillskriva kanter olika signifikans efter någon parameter, till exempel antalet interaktioner mellan två noder.

För vissa av beräkningarna är det fördelaktigt att uttrycka grafer i matrisform. För dessa syften används *viktmatrisen W* (engelska; *weight*) (NetworkX n.d.[a]), ibland kallad angränsningsmatrisen för grafer utan vikter (Bichler 2019, s. 28-29). Denna är kvadratisk med storleken $|V|$, antalet noder. $W_{ij} > 0$ endast om kanten e_{ij} existerar, och är lika med dess vikt om grafen är viktad (om grafen är oviktad kan W_{ij} endast anta värdet 0 eller någon konstant, typiskt 1). Denna matris kan endast anta positiva värden eftersom att riktade kanter inte behandlas i denna studie.

Noder, kanter, och vikter är tillsammans de tre elementen som definierar grafen G enligt: $G = (V, E, W)$. Notera att vikterna W inte behöver konstrueras som en matris för denna studies operationalisering då en enkel lista av vikter räcker, men används i den matematiska definitionen av vissa metoder. Vikterna kommer att refereras till som W även om de tar en matrisform eller inte för att hålla notationen simpel, då vikterna är samma oavsett form.

Vägar används inte direkt i denna studie. En väg γ är en sekvens av noder där en efterföljande nod har en kant som kopplar den till den tidigare:

$$\gamma = (v_1, v_2, ..., v_k), \quad e_{i(i-1)} \in E, \quad \forall 1 \leq i \leq k$$

Dessa anses inte representera något tillräckligt meningsfullt inom ramen för ett onlineforum, då individer inte behöver gå via sina existerande kontakter för att nå nya på samma sätt som i analoga nätverk. Det går däremot att tolka kanter mellan noder inte endast som kontaktvägar, utan även som potential för utövande av influens mellan noder. Därför har centralitetsmått som beräknas med hjälp av vägar fortfarande inkluderats, och en definition av vägar är värdefull för förståelsen av andra koncept. Däremot har resultat baserade på vägar som är av intresse i andra kontexter exkluderats, som att till exempel undersöka längsta möjliga vägen genom nätverket.

2.2 Allmän deskriptiv statistik

Det första som kan vara intressant när ett nätverk ska analyseras i sin helhet är det som Bichler (2019, s. 156-159) kallar allmän deskriptiv statistik, vilket innefattar grundläggande och övergripande egenskaper hos nätverket.

Först och främst inkluderar detta nätverkets totala antal noder $|V|$ och kanter $|E|$, för att ge en övergripande bild av nätverkets storlek. En annan egenskap som kan vara av intresse är nätverkets *komponenter*.

En komponent är en samling noder och kanter så att alla noderna i komponenten går att nå via vägar från alla andra noder i samma komponent (Bichler 2019, s. 158). Eftersom att denna rapport inte inkluderar riktade grafer är en komponent en samling noder där alla individuella noder har en kant till minst en annan nod i samma komponent. Mer formellt kan komponenterna delas upp enligt:

$$V = V_1 \cup V_2 \cup ... \cup V_k, \quad V_i \cap V_j = \emptyset, \quad i \neq j$$

Där k är antalet komponenter och där $\emptyset$ är den tomma mängden. Det är ofta intressant att studera ett nätverks största komponent, $V_{max} = \max_{1<i<k}(V_i)$, den komponent som innefattar flest noder. Andelen av de totala antalet noder som ingår i denna ges av $V_{\%} = \frac{|V_{max}|}{|V|}$.

2.2.1 Sammanfattning: Allmän deskriptiv statistik

Sammanfattningsvis inkluderar den allmänna deskriptiva statistiken:

- Totalt antal noder: $|V|$
- Totalt antal länkar: $|E|$
- Antal komponenter: k
- Andel noder i den största komponenten: $V_{\%}$

2.3 Noders positionering

Individuella noders centralitet är en fundamental aspekt av deras positionering i ett nätverk. Det simplaste måttet på hur central en nod är i ett nätverk är gradcentralieten (engelska; *degree centrality*); $d(v_i)$. Gradcentralitet och viktad gradcentralitet för nod i, $d(v_i)$ respektive $d_w(v_i)$ beräknas enligt (Bichler 2019, s. 163):

$$d_w(v_i) = \sum_{j=1}^{|V|} w_{ij} \qquad d(v_i) = \sum_{j=1}^{|V|} e_{ij}$$

Där vikten w_{ij} är lika med noll om kanten e_{ij} inte existerar och e_{ij} antar värdet 1 eller 0 beroende på om kanten existerar eller inte. Det är alltså antalet kanter en nod har, och dess viktsumma om grafen är viktad. Gradcentrlitet kan hjälpa med identifikationen av ledare, eller åtminstone viktiga noder i någon annan mening (Koschade 2006), denna tolkning är dock inte applicerad på ett forum. Inom ramen för en studie av ett hackernätverk (dock inte ett forum) beskriver Lu et al. (2010) en nod med hög gradcentralitet som en individ som är sannolik att känna till och sprida information, samt att isolering av noden kan försvaga nätverket. Gradcentralitet har även påståtts korrelera med spridningen av information i en studie av ett hackerforum (Grisham et al. 2017).

En potentiell begränsning av gradcentralitet är att alla grannar behandlas som lika (inte kanten som länkar till grannen då denna kan vara viktad). Om centralitet liknas med en aktörs inflytande på nätverket, räcker det inte att ta hänsyn till en enskild nods grannar, då grannarna själva är påverkade av sina grannar, och så vidare. Det är problemet som egenvektorcentralitet ämnar att lösa (Bichler 2019, s. 186). Egenvektorcentraliteten z uppfyller följande (NetworkX n.d.[c]):

$$\lambda z = zW'$$

där W' är den transponerade vikmatrisen, λ är dennas ledande egenvärde och z är dennas egenvektor, vilken innehåller samtliga noders egenvektorcentralitet. Därmed blir nod i:s egenvektorcentralitet:

$$z(v_i) = \frac{1}{\lambda} \sum_{j \to i} z(v_j)$$

Egenvektorcentraliteten påverkas av grannars egenvektorcentralitet, och dessa är i sin tur påverkade av sina grannars, och så vidare. Hög egenvektorcentralitet har inom ramen för ett hackernätverk tolkats som att noden i fråga helt enkelt är kopplad till många centrala individer, och att dess isolering inte sannolikt har stor effekt på nätverket (Lu et al. 2010). I en mer generell kontext har måttet beskrivits liknande, med tillägget att det kan bidra till identifieringen av delvis dolda, men fortfarande centrala och viktiga individer (Bichler 2019, s, 181). Annan forskning på ett hackerforum har dock tolkat det som ett mer direkt mått på nodens influens i nätverket (Grisham et al. 2017).

Mellanliggande centralitet (NetworkX n.d.[b]; Bichler 2019, s. 182-184) är ett annat mått av intresse (engelska; *betweenness centrality*. Mellanliggande centralitet för en nod definieras som andelen kortaste vägar mellan andra nodpar i grafen passerar genom noden:

$$b(v_i) = \sum_{v_j, v_k \in V} \frac{\sigma_{jk}(v_i)}{\sigma_{jk}}$$

där σ_{jk} är antalet kortaste vägar γ mellan v_j och v_k och $\sigma_{jk}(v_i)$ är antalet kortaste vägar mellan v_j och v_k som passerar genom v_i. Om ett högt värde erhålles kan detta mått indikera att noden i fråga har en mäklande eller medlande roll i nätverket, med viss kontroll över flöden av information eller resurser genom nätverket (Bichler 2019, s. 182; Koschade 2006). Detta är en tolkning som även gjorts specifikt för hackernätverk (Lu et al. 2010). Inom ramen för ett hackerforum består dessa flöden mest troligen av information, men potentiellt även hackerverktyg och andra resurser. På grund av hur ett forum är strukturerat är det inte möjligt för en nod att positionera sig som mäklare mellan två andra, utan att de har möjlighet att direkt kontakta varandra. I kontexten av ett forum kan därför hög mellanliggande centralitet fortfarande tolkas som medlande, men möjligen mer i termer av influens snarare än information eller resurser.

2.3.1 Sammanfattning: Noders positionering

Sammanfattningsvis inkluderar beskrivningen av nodernas positionering:

- Gradcentralitet: $d(v_i)$
- Egenvektorcentralitet: $z(v_i)$
- Mellanliggande centralitet: $b(v_i)$

2.4 Strukturell beskrivning

Bichler (2019, s. 159-168) samlar nätverkets sammanhållning och centralisering under begreppet strukturell beskrivning. För att undersöka nätverkets sammanhållning kommer dess densitet att beräknas. Densitet, D, definieras som andelen av nätverkets potentiella kanter som faktiskt har observerats (Bichler 2019, s. 161). Densitet beräknas enligt:

$$D = \frac{|E|}{\frac{1}{2}|V|(|V|-1)}$$

För bedöma nätverkets centralisering tar Bichler (2019, s. 163-166) upp genomsnittlig grad och gradcentralisering (som applicerat på hela nätverket och därmed distinkt från gradcentraliteten som den definierades för en enskild nod i del 2.3).

Genomsnittlig grad syftar till det genomsnittliga antalet kanter som en individuell nod ingår i (alternativt den genomsnittliga viktsumman av dessa kanter för viktade grafer). Genomsnittlig grad $\bar{d}$, beräknas enligt:

$$\bar{d} = \frac{\sum_{\forall i} d(v_i)}{|V|}$$

I en artikel av Freeman definieras denna gradcentralisering för hela nätverket tillsammans med motsvarande centraliseringsmått för mellanliggande centralitet vilka kommer att användas för att beräkna centraliseringen för nätverket i denna studie (Freeman 1978). Freeman (1978) definierar centraliseringsparametrarna generellt som:

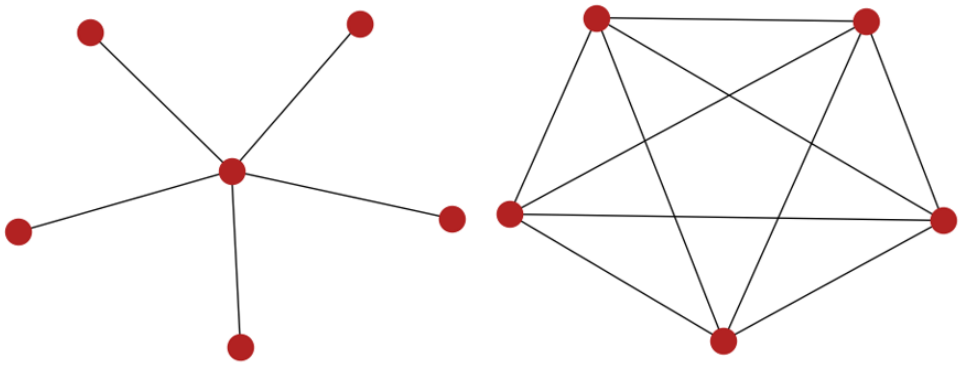
$$C_X = \frac{\sum(C_X(v*) - C_X(v_i))}{\max \sum(C_X(v^*) - C_X(v_i))}$$

där C_X är centraliseringen i fråga (grad eller mellanliggande), $C_X(v_i)$ är motsvarande centralitet för nod i och $C_X(v^*)$ är det största värdet på centralitetsmåttet i fråga. Uttrycket med max-funktionen i nämnaren syftar till det störta möjliga värdet på skillnaden i fråga, och beror endast av nätverkets storlek, $|V|$. För härledning av nämnaren uttryckt i termer av $|V|$ hänvisas läsaren till Freemans (1978) artikel. Detta är alltså en omvandling mellan ett centralitetsmått för noder till ett mått på hur centraliserat hela nätverket är, vilket kommer att kallas *centralisering* i denna studie. Med denna generella formel i åtanke definieras nu gradcentralisering; C_d, och mellanliggande centralisering; C_b för hela nätverket enligt (Freeman 1978):

$$C_d = \frac{\sum_{\forall i} d(v*) - d(v_i)}{|V|^2 - 3|V| + 2}$$

$$C_b = \frac{\sum_{\forall i} b(v*) - b(v_i)}{|V|^3 - 4|V|^2 + 5|V| - 2}$$

Dessa kan tyckas vara mer abstrakta, men har tydliga implikationer för hur centraliserat ett nätverk är. Gradcentralisering är högre när centrala noder har många kanter till andra noder i nätverket, medan perifera noder har få (Lu et al. 2010). Den är som lägst när alla noder har samma gradcentralitet. Mellanliggande centralisering indikerar till vilken utsträckning en eller ett fåtal noder är positionerade på kortaste vägen mellan andra oftare än resten. För att illustrera dessa värden har två grafer inkluderats i figur 2. En stjärngraf är så centraliserad som en graf kan vara, med samtliga centraliseringsvärden $C_d = C_b = 1$, medan en komplett graf är så decentraliserad som det går, eftersom att alla noder är kopplade till alla andra, och centraliseringsmåtten kommer att ha lägre värden (Freeman 1978). I stjärngrafen har den centrala noden $b(v_i) = 1$ och $d(v_i) = 5$, medan övriga har $b(v_i) = 0$ och $d(v_i) = 1$. I den kompletta grafen har samtliga noder samma centralitet i alla avseenden. I sammanfattning; ju mindre jämnt fördelat ett centralitetsmått är inom ett nätverk, desto mer närmar sig motsvarande centraliseringsmått det högsta värdet 1.



Figur 2: En centraliserad stjärngraf (vänster) och en decentraliserad komplett graf (höger).

2.4.1 Sammanfattning: Strukturell beskrivning

Sammanfattningsvis inkluderar den strukturella beskrivningen:

- Densitet: D
- Genomsnittlig grad: $\bar{d}$
- Gradcentralisering: C_d
- Mellanliggande centralisering: C_b

3 Bakgrund och tidigare forskning

3.1 Utgångspunkter och antaganden

SNA är förenklat uttryckt en metodik för att studera den sociala strukturen som individer ingår i. Det framstår som att det finns delade meningar inom forskningen om hur SNA betraktas och i vilken utsträckning denna är en självstående, formell samhällsvetenskaplig teori. Ett tidigt och till synes inflytelserikt verk (då det refereras till i forskningen som har lästs för denna studie) av Berkowitz (1982, s. 22, 161) beskriver det som en paradigm som inte har utvecklats till fullo och skriver att SNA inte är ett fält i konventionell mening. Det förekommer även senare forskning som refererar till denna (Berkowitz 1982) och vidhåller liknande ställningar, som till exempel att SNA är bred strategi (Otte & Rousseau 2002) eller ett tillvägagångsätt (Zhang 2010) för undersökningen av sociala strutkurer, snarare än en formell teori. Bichler (2019, s. 14) beskriver i sin bok om SNA det som en ”bona fide disciplin” med egna teorier, metoder och statistiska tillämpningar. Freeman (2004, s. 3) kallar det en ung, men organiserad paradigm i sin bok.

Hur SNA som område betraktas inom samhällsvetenskapen i stort anses inte ha en avgörande och direkt påverkan på dess merit som metod för denna studie. Det är dock en fråga som behöver nämnas då denna rapport, som egentligen inte ämnar att ta ställning i frågan, kommer att behandla, SNA som en egen teori och disciplin.

Bakgrunden till SNA kan spåras till sent 1800-tal, när förklaringar till sociala fenomen började forumelaras med hjälp av individers sociala grupper och kopplingar (Zhang 2010). Det skulle dock dröja innan sociala nätverk som de beskrivs idag började konceptualiseras. SNA utvecklades med tiden och i slutet av 1970-talet var det utbrett och erkänt inom samhällsvetenskapen. Det är ett tvärvetenskapligt angreppsätt som har utvecklats under influenser från sociologi, matematik, och många andra fält, med ett stort fokus på den sociala kontexten snarare än individuella egenskaper (Zhang 2010). Sammantaget med tidigare resonemang om den vetenskapliga synen på SNA framstår det som en rättvis bedömning att det successivt blir och har blivit formaliserat och accepterat som disciplin.

Oavsett om SNA betraktas som en självständig teori eller mer som ett angreppsätt eller metodik har den samma vetenskapsteoretiska utgångspunkter. Som nämnt i ovan stycke grundar sig SNA från början i en övertygelse om att individers sociala kontext, deras kopplingar och relationer till andra individer, är signifikanta (Zhang 2010). Relationerna och de mönster de bildar över större nätverk är viktiga egenskaper i sig och påverkar hur individerna agerar. Det är inte heller bara de direkta länkarna till andra individer som är relevanta, då en individs nära relationer i sin tur är influerade av sin omgivning i nätverket (Bichler 2019, s, 1). Alltså influeras en individ inte endast av dennas direkta relationer, utan även indirekt via dessa av deras direkta relationer, och så vidare. Bichler (2019, s. 4) beskriver det också som att ett grundläggande perspektiv inom SNA:n är att beroenden spelar roll, medan

många andra vetenskapliga metoder antar att faktorer är oberoende av varandra.

Freeman definierar i sin bok *The Development of Social Network Analysis* fyra egenskaper som återfinns inom den moderna SNA:n och tillsammans definierar fältet (Freeman 2004, s. 3):

1. SNA motiveras av en strukturell intuition baserat på band som länkar sociala aktörer.
2. SNA är grundat i systematisk och empirisk data.
3. SNA bygger på grafteoretisk visualisering.
4. SNA förlitar sig på matematiska beräkningsmodeller.

Detta stämmer väl överens med de utgångspunkter som just diskuterades. SNA utgår från att banden som länkar sociala aktörer är betydelsefulla, men det väsentliga som Freeman (2004) tillför här är att det också är meningsfullt att empiriskt och mattematiskt modellera och studera dessa.

Bichler definierar i sin bok (2019, s. 26-35) fem grundläggande antaganden inom SNA:n:

1. Ömsesidiga beroenden mellan aktörer beskriver deras beteende bättre än individuella egenskaper.
2. Konstellationen av relationer mellan aktörer influerar varje social enhet som ingår i nätverket.
3. Sociala nätverk är dynamiska och utvecklas kontinuerligt.
4. Övergripande sociala strukturer och transformationer uppkommer från de kombinerade beteenden och preferenser hos individer i nätverket, med begränsad kunskap om det större sociala nätverk de innefattas av.
5. Sociala nätverk består, även när individuella aktörer blir medlemmar av eller lämnar nätverket, eller när relationer skapas eller löses upp.

Återigen stämmer dessa antaganden väl med vad som har presenterats innan, och tillför ytterligare dimensioner. Relationer mellan individer har inte bara signifikant påverkan på individerna och deras beteende, utan är uttryckligen en bättre beskrivning av detta. Relationerna och den sociala struktur som de bildar är inte betydelsefulla för att individerna inte har egna preferenser och beteenden, utan för att de påverkar varandra och sin omgivning med dessa i en så hög grad. Slutligen bidrar Bichler's (2019, s. 26-35) perspektiv med den viktiga poängen att nätverk är dynamiska. En studie av det slag som görs här kräver av sin natur att ett relativt statiskt urval av data används, så det är viktigt att ha de sociala nätverkens föränderlighet i åtanke när resultaten ska granskas.

Systematisk och empirisk data, matematisk metodik, och fokus på beroenden och påverkan (orsak och verkan) mellan aktörer indikerar ett mer naturvetenskapligt eller positivistiskt perspektiv. Det bör dock nämnas att ingen av författarna som inkluderats i materialet för denna studie uttryckligen gör denna distinktion. Däremot skriver Freeman (2004, s. 125) om strukturalism i samband med utvecklingen av SNA. Strukturalismen är präglat av ett intresse av relationer snarare än individer alternativt definierar objekt i termer av deras relationer snarare än sina egenskaper i isolering (Barbosa de Almeida 2015), vilket definitivt är ett teoretiskt perspektiv som passar in med utgångspunkterna för SNA. Ett strukturalistiskt perspektiv tolkas inte heller som en motsättning mot ett positivistikt, enligt de faktorer som nämndes ovan.

3.2 SNA och kriminella nätverk

SNA-metodik har länge applicerats på olika typer av kriminella och dolda nätverk. De grundläggande metoderna och angreppsätten förblir väsentligen densamma som för öppna nätverk, men metodiken får andra problem jämfört med öppna nätverk. Kriminella nätverk har incitament att dölja sina relationer och aktiviteter för att undgå upptäckt och motåtgärder.

Detta incitament kan ge olika effekter för hur nätverket ser ut och hur de interagerar, exempelvis kan det leda till en preferens för decentraliserad och gles nätverksstruktur (Bichler 2019, s. 95). Ett tidigt exempel är en artikel från Baker och Faulkner (1993) , som kartlägger tre fall av konspirationer med prissättningar inom elektronikbranschen. De fann att nätverkstrukturen främst var driven av behovet av hållas dolt och att hög centralitet för en nod bidrog till sannolikheten för utfallet av dom och straff (Baker & Faulkner 1993). De illustrerade den fundamentala avvägningen mellan effektivitet och doldhet, då mer frekvent kontakt kan öka både nätverkets produktivitet och synlighet. Centrala individer (åtminstone i termer av gradcentralitet som diskuterat i del 2.3) har fler kontakter inom nätverket, vilket gör att de även har fler vittnen som kan identifiera dem.

Sparrow (1991) skrev en idag fortfarande relevant artikel som konceptualiserade problemen för SNA applicerat på kriminella nätverk. Enligt studien är datamaterialet vid studier av kriminella nätverk (till exempel kriminalunderrättelsedatabaser) ofta väldigt stort, inkomplett, dåligt avgränsat, och dynamiskt (Sparrow 1991). Dessa problem kan rimligen antas vara närvarande för studier av kriminella nätverk i stort, men också troligen vara synnerligen problematiska för SNA, då det är nätverket och den sociala kontexten i sig som ska studeras och inte individuella aspekter. Problematiken för SNA som metod kommer att diskuteras under rapportens gång, och mer specifikt i del 4.4, men dessa aspekter är värda att ha i åtanke under studiens gång.

Det har även gjorts bidrag till SNA-området inom studier av terrorism och terroristnätverk. Ett exempel är en studie av en del av terroristnätverket bakom terrorattackerna den 11:e september 2001, som gjordes baserat på offentlig information, främst hämtat från nyhetsartiklar (Krebs 2002). Förutom en illustration av och viss insikt i strukturen av nätverken understryker artikeln möjligheterna och problemen med SNA-metodik applicerat på dolda nätverk. Exempelvis belyser studien vikten av att samarbeta kring informationsinsamling mellan organisationer, då författaren hittade öppen information från vissa källor som hade varit av intresse för andra (Krebs 2002). Alltså styrker resultaten även de grundläggande problemen identifierade av Sparrow (1991). En annan artikel av Koschade (2006) studerade med hjälp av SNA cellen av Jemaah Islamiyah som låg bakom bombdåden på Bali 2002. Studien identifierade nyckelindivider och en nod som utgjorde en svag punkt, då dennas avlägsnande hade haft stor inverkan på nätverkets effektivitet (Koschade 2006). Det var också ett förhållandevis kompakt och sammankopplat nätverk, som i viss utsträckning uppoffrade potential att förbli dolda för högre effektivitet (Koschade 2006).

Dolda och kriminella nätverk, samt studierna av dessa, skiljer sig ofta markant från legala. Dolda nätverk, vare sig de är hackers, bedragare, eller terrorister har incitament för att dölja sin verksamhet eller behålla sin anonymitet, vilket kan lämna tydliga spår i nätverkets struktur. Det finns dock naturligen stora skillnader mellan ett hackerforum och en terroristcell, vilket kräver en anpassning hur det kan angripas med en SNA.

3.3 SNA, hackernätverk, och forum

Idén att applicera SNA på nätverk av cyberkriminella och hackerformum är inte ny för denna studie. Det finns ingen snäv och vedertagen definition av en kriminell hackare, men denna studie tar definitionen av en person med ett stort intresse för datorer och teknologi, som använder sina kunskaper för att illegalt få tillgång till digitala system (Holt et al. 2012). Kriminella aktörer i dessa nätverk har allt annat lika samma incitament att hålla sig själva, sina relationer, och aktiviteter dolda som andra kriminella. Den största väsentliga skillnaden är att denna typ av kriminell aktivitet till sin natur är digital och sker över internet. Annan kriminell aktivitet och kommunikation sker självklart också över internet och på forum, men det är troligen generellt sett inte fullt lika naturligt att organisera sig online. Kommunikationskanalen som används av ett nätverk kan antas ha en signifikant påverkan på hur nätverket ser ut och är strukturerat. Till exempel existerar det ett stort antal forum och marknadsplatser för hackare och cyberkriminella, vilket troligen tillåter snabbare och simplare anslutning för nya medlemmar än mer traditionellt organiserade kriminella och dolda nätverk.

Det är som tidigare nämnt i del 1.2 just ett forum som ska kartläggas och studeras i denna studie. Detta är viktigt att ha i åtanke när teorin och den tidigare forskningen diskuteras, då koncept kan ta en något annorlunda betydelse. Till exempel kan det antas att alla medlemmar i ett forum finns tillgängliga att kontakta för alla andra medlemmar, vilket inte är garanterat för nätverk i andra medium. Exempelvis leder detta till att om en individ i ett digitalt och en i ett fysiskt nätverk

jämförs på basis av sin förmåga att förmedla eller kontrollera kommunikationen inom nätverket kan resultaten bli missvisande. Det påstås inte att dessa positioner i ett nätverk som existerar på ett onlineforum är obetydliga, då det fortfarande kan vara en betydelsefull position i termer av hur influens sprids genom nätverket, men skillnaden är viktig när olika nätverk ska jämföras.

Denna del av studien kommer att presentera och diskutera ett urval av den tidigare SNA-forskningen på cyberkriminella nätverk för att möjliggöra en jämförelse och kontextualisering av denna studies resultat.

3.3.1 Tidigare fallstudier

Som ett första exempel har det gjorts en analys av strukturen av den ökända gruppen *ShadowCrew* (Lu et al. 2010). Studien studerade nätverkets centralisering för att identifiera ledare, deras potentiella influens, och undergrupperingar inom nätverket. De visade att nätverket var decentraliserat, vilket som tidigare nämnt är associerat med att hålla nätverket dolt (Bichler 2019, s. 95), och hade en utvecklad arbetsfördelning (Lu et al. 2010). Påståendet att *Shadowcrew* är decentraliserat baseras på att det hade låga värden på grad- och mellanliggande centralisering. Nätverket hade flera inflytelserika ledare och undergrupperingar existerade, med vissa medlemmar mer direkt involverade i dessa (Lu et al. 2010). De relevanta resultaten från studien är inkuderade i tabell 1.

Mått	*Värde*
Gradcentralisering, C_d	0,269
Betweenness-centralisering, C_b	0,041
Största gradcentralitet, $\max(d(v_i)$	0,977
Största egenvektorcentralitet, $\max(z(v_i))$	0,241
Största betweeness-centralitet, $\max(b(v_i))$	0,178

Tabell 1: Relevanta resultat från studien av *ShadowCrew* (Lu et al. 2010).

Det bör understrykas att nätverket som studerades i den ovan nämnda studien (Lu et al. 2010) är en förhållandevis liten och väldefinierad grupp, åtminstone i jämförelse med potentialen ett onlineforum har att tillåta anslutning av nya medlemmar. En annan studie betitlad *A Social Network Analysis and Comparison of Six Dark Web Forums* (Pete et al. 2020) studerade som titeln anger sex stycken forum på den så kallade *"mörka webben"*. Notera att denna studie inte är begränsad till hackerforum. Författarna skriver uttryckligen att fyra av forumen behandlar hackerrelaterad information, och endast resultaten från SNA av dessa kommer att diskuteras här. De studerade vilka strukturella egenkaper dessa nätverk har, samt hur inläggsaktiviteten ser ut för de mer centrala noderna (Pete et al. 2020). Deras resultat antyder bland annat ett inverterat samband mellan storleken på nätverket och dess densitet (Pete et al. 2020). De relevanta resultaten har inkluderats i tabell 2.

Mått	*Forum 1*	*Forum 2*	*Forum 3*	*Forum 4*
Antal noder, $\|V\|$	16401	1781	22	2887
Antal kanter, $\|E\|$	624926	19636	57	63688
Densitet, D	0.004	0.012	0.247	0.015
Genomsnittlig gradcentralitet, $\bar{d}$	76.2	22.05	5.18	44.12
Största grad, $\max(d(v_i))$	15617	628	14	1202

Tabell 2: Relevanta resultat från de fyra hackerrealterade forumen (Forum C - F i studien) (Pete et al. 2020).

En studie av Samtani & Chen (2016) studerade ett forum för att identifiera nyckelindivider och programvara, de relevanta resultaten presenteras i tabell 3. En intressant slutsats som de drog var att även fast det fanns relativt lätt tillgång till programvara och erfarna användare, tog majoriteten av medlemmarna endast nytta av ett fåtal kontakter (Samtani & Chen 2016). Detta motiverades med att vägarna genom nätverket var korta, men densiteten var ändå låg. Relaterat till detta så visade de att de mest centrala användarna är en liten grupp, och att de flesta av dessa är seniora hackare som under en längre tid har bidragit kunskap till forumet (Samtani & Chen 2016). Det finns stöd för denna observation i tidigare forskning (Holt et al. 2012), alltså uppdelningen mellan en liten grupp centrala och skickliga hackers, och en större grupp med mindre centrala och skickliga användare. Holt et al. (2012) skriver även att hackergemenskapen i stort generellt kan delas upp i tre grupper, med ökande storlek och minskande skicklighet.

Mått	*Värde*
Antal noder, $\|V\|$	63
Antal kanter, $\|E\|$	510
Densitet, D	0,131
Antal komponenter, k	5
Andel noder i största komponent, $V_{\%}$	28,78%
Genomsnittlig gradcentralitet, $\bar{d}$	16,19
Stösta gradcentralitet, $\max(d(v_i))$	36

Tabell 3: Relevanta resultat från studien av Samtani & Chen (2016).

4 Metod

Metoden för denna studie utgörs huvudsakligen av de metoder och mått som presenterades i teorin idel 2. I denna del presenteras dataubderlaget, operationaliseringen av metoderna inklusive några tillskott som inte kräver ytterligare teoretisk grund, och slutligen potentiella begränsningar med metodiken och operationaliseringen.

4.1 Val av studieobjekt

Syftet med denna studie är att göra en djup och detaljerad analys av ett hackernätverk. Det är en ansats att förstå det specifika med just det nätverk som har analyserats snarare än att försöka styrka eller underminera någon generell hypotes. Möjligen går det att se denna studie som en prövning av SNA-teorins underliggande antaganden och validitet som underlag för beslut, men eftersom det inte har funnits någon dokumentation från verkliga åtgärder mot (eller i skydd mot) forumet att jämföra med är detta inte ett direkt syfte med studien. Förhoppningen är, som nämnt i introduktionen (del 1.1), att studien kan bidra till en större förståelse för hur ett hackerforum kan se ut och till framtida forskning på detta och närliggande områden, som till exempel proaktiv CTI. Det är dock utanför denna studies omfång att göra ett direkt och signifikant bidrag till mer generaliserande forskning på dessa områden, och ett enstaka fall kan förhoppningsvis bidra som en byggsten för framtida forskning. Dataunderlag har valts för att det passar problemformuleringen, samt för att det är lättillgängligt både för detta arbete och framtida studier, vilket kan bidra till bättre potential för reproducerbarhet och framtida jämförelser.

Forumet innehåller information om illegalt hackande och annan cyberkriminalitet. Huruvida individuella aktörer på forumet är kriminella, oavsett om de har gjort inlägg som behandlar illegal aktivitet, är inte något som denna studie tar hänsyn till. Detta är en kartläggning av ett forum där information om illegal och potentiellt illegalt hackande sprids, och ett ställningstagande till huruvida forumet i praktiken konstituerar ett kriminellt nätverk är utanför dennas studies omfång. Det anses finnas tillräckligt med underlag för att forumet åtminstone har tillräckligt gemensamt med dolda och illegala hackerforum och nätverk för att resultaten ska kunna bidra till forskningen på området.

4.2 Empiri

Empirin för denna rapport består av data från ett hackerforum samlat av ett projekt hos University of Arizona (Robert, Ning, & Mohammadreza 2018). Vid tidpunkten för publicering av datan var det ett av de största hackerforumen. Det är det enda dataset projektet specifikt rekommenderar för en SNA. Det går att läsa mer om projektet på universitetets hemsida (University of Arizona n.d.).

Datan (Robert, Ning, & Mohammadreza 2018) är insamlad mellan 8 april 2013 och 24 februari 2018. Den är utformad som en tabell, mer specifikt en *MySQL-dump*, där varje rad beskriver ett inlägg och kolumnerna innehåller beskriver olika egenskaper hos inlägget, som till exempel associerad tråd och författare. Datan är hämtat med programmet Offline Explorer enligt författarna själva och har sedan behandlats med en syntaxanalyserare skriven i programmeringsspråket Java.

För den intresserade läsaren har bilder av forumet inkluderats i bilagorna i del A.1. Forumet är inte längre tillgängligt, men är arkiverat av Wayback Machine från Internet Archive. Det går att läsa mer om denna på deras hemsida (Archive n.d.).

Kolumnerna (med sitt namn i databasen inom parantes om applicerbart) som beskriver datan (Robert, Ning, & Mohammadreza 2018) är:

- Inläggets ID (postID),
- trådens ID (threadID),
- trådens titel (threadTitle),
- subforum,
- författarens namn (authorName),
- författarens medlemskap (postAuthorMembership),
- författarens datum för anslutning till forumet (postAuthorJoinDate),
- författarens rykte (authorReputation),
- innehåll (flatContent),
- inläggets datum (postDate),
- inläggssekvens (postSequence, stegar upp med 1 för varje inlägg som görs på samma tråd),
- antal gilla-markeringar (likes),
- namn på bifogad fil (attachmentName),
- URL,
- innehåll med HTML-etikett (contentWithHTMLTag),
- och slutligen författarens ICQ[1] (authorICQ).

[1] En tjänst för direkta medellanden.

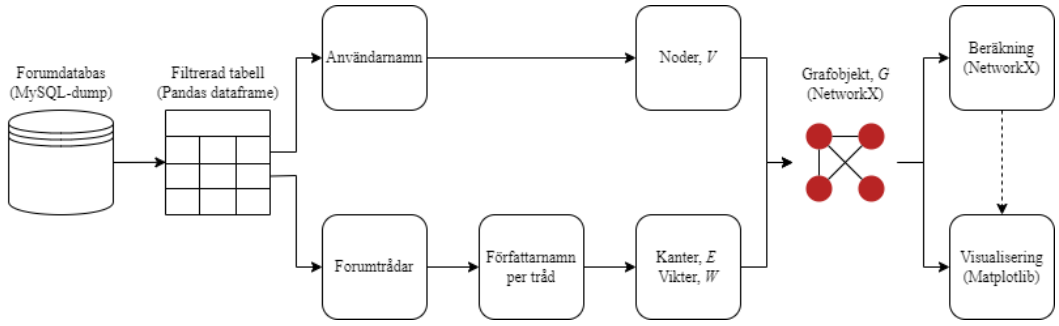
Tyvärr är ett flertal kolumnerna, vilka hade kunnat bidra till analysen, tomma. Detta gäller för kolumnerna postAuthorMembership, authorReputation, likes, och authorICQ. Det hade varit intressant att jämföra en författares rykte, medlemsstatus och antal gilla-markeringar med olika mått på deras centralitet.

Som det går att utlösa av listan ovan är forumet uppdelat i subforum. Mer specifikt, 32 subforum innehållandes allt från 1 till över 10 000 inlägg. Ett stort antal behandlar naturligen hackande-relaterade frågor, men vissa gör inte det (och vissa har ett svårtolkat tema). Ett antal av dessa var grupperade under namnet ”*hacking zone*” och när specifika subforum behandlats är det dessa och några intressanta tillskott som har inkluderats. Det ansågs inte vara nödvändigt att utesluta subforum som behandlar ämnen utanför hackingsfären, då det antas att alla som besöker forumet är intresserade av legalt tveksamt hackande. Andra ämnen kan lika lätt konsumeras på andra sidor, och underlaget för studien ville begränsas så lite som möjligt. Kopplingar mellan intressanta individer, även i mindre intressanta kontexter, anses alltså fortfarande vara lika betydelsefulla. Ett tomt användarnamn hittades och exkluderades från beräkningarna eftersom det var tvetydigt om detta faktiskt representerade en användare eller inte.

4.3 Operationalisering

Stommen av metodiken består av ett skript i programmeringsspråket python, med hjälp av externa bibliotek. Mer specifikt *NetworkX* (Hagberg, Schult, & Swart 2008) för SNA-metodik, *Pandas* (McKinney 2010) för datahantering, *MySQL* (Axmark & Widenius 2023) och tillhörande pythonbibliotek (Axmark & Widenius 2023, s. 440) för dataimportering, samt slutligten *Matplotlib* (Hunter 2007) för visualisering. De flesta av de beskrivna och definierade måtten i del 2 finns redan färdiga för användning i biblioteket NetworkX, med framförallt centraliseringsbegreppen som undantag. Det operationaliserade arbetsflödet illustreras i figur 3 nedan.

Denna del ska beskriva operationaliseringen och börjar med konstruktionen av grafen i del 4.3.1. Sedan följer operationaliseringen samma struktur som del 2: Allmänna deskriptiva statistiken i del 4.3.2 (teori från del 2.2), noders positionering i del 4.3.3 (teori från del 2.3), och strukturell beskrivning i del 4.3.4 (teori från del 2.4). Som tidigare nämnt är det den allmänna deskriptiva statistiken och den strukturella beskrivningen (del 4.3.2 och 4.3.4) som ska hjälpa besvara den första frågeställningen och nodernas positionering (del 4.3.3) den andra.



Figur 3: Arbetsflödet för konstruktionen, analysen, och visualiseringen av grafen.

4.3.1 Konstruktion av graf

Datan som har beskrivits ovan i del 4.2 är som nämnt i formatet av en MySQL-dump. För att kunna läsa in denna behövdes en MySQL-databas skapas. MySQL-dumpfilen lästes sedan in i denna databas, som kan kopplas till python genom MySQLs pythonbibliotek. Datan lästes in från databasen till en Pandas dataram (engelska *dataframe*) där den behandlades och filtrerades.

För att nyttja funktionaliteten i NetworkX krävs konstruktionen av ett grafobjekt från detta bibliotek. För att konstruera en oriktad, viktad graf krävs tre element som presenterade i del 2: En mängd noder; V, en mängd kanter; E, och vikter för dessa kanter, enligt $G = (V, E, W)$.

Noderna representerar unika användarnamn i databasen. För att generera noderna till V isolerades därför alla unika användarnamn från den dataramen. Att generera viktade kanter mellan nodpar i V är något mer komplicerat. För detta sorterades först alla inlägg efter vilken tråd de tillhör (genom värdet på threadID), alltså efter attributet threadID. För varje tråd sammanställdes en lista över alla användare som har gjort ett inlägg på tråden, samt en ytterligare lista med alla parvisa kombinationer av dessa användarnamn. Dessa parvisa kombinationer är vad som ska representeras av kanterna i grafen. Slutligen gås denna lista med parvisa kombinationer av noder v_i, v_j igenom med frågan; finns det redan en kant e_{ij} eller e_{ji}? Om ja; öka dess vikt med 1, om nej; skapa kanten.

En tråd i datan (med threadID 0) verkade till synes bestå av flertalet trådar. Den hade inlägg med olika trådtitlar (threadTitle) och inläggssekvensen (postSequence) var inte sekvensiell. Om en tråd tillhör tråd 0 i datan utgår konstruktionen från titeln av tråden istället för trådens id. Det har verifierats att resterande trådar är sekvensiella och korrekta.

Därmed har det nu skapats en lista av användarnamn som ska representeras av en mängd av noder V och en lista med par av användarnamn och vikter som ska representeras av mängderna kanter E och vikter W. Med dessa element är det väldigt simpelt att låta NetworkX konstruera ett grafobjekt som kan representera nätverket i den SNA som ska genomföras. Det är sedan detta grafobjekt som är utgångspunkten för en betydande majoritet av SNA:n.

De kvarvarande delarna av operationaliseringen är som kan läsas i figur 3 beräkningar och visualiseringar. Beroende på vilket specifikt resultat som åsyftas så ritades grafen med hjälp av Matplotlib, eller användes som underlag för någon beräkning, helt eller till övervägande del genomförd med NetworkX. Slutligen visualiseras även vissa av de beräknade resultaten också med Matplotlib. En sammanställning av beräkningar och metoder finns nedan i del 4.3.2 - 4.3.4 och en komplett sammanfattning återges i tabell 4.

4.3.2 Allmän deskriptiv statistik

Från teorin för allmän deskriptiv statistik i del 2.2 inkluderas samtliga metoder. Även visualiseringar av hela nätverket G samt den största komponenten V_{max} är inkluderade, samt en figur för storleksjämförelse med utvalda forum från den tidigare forskningen.

4.3.3 Noders positionering

Från teorin för noders positionering i del 2.3 inkluderas samtliga metoder och de tio mest centrala noderna för varje centralitetsmått kommer att presenteras.

För kontextualisera jämförelsen av noderna har även en tabell som visar de användare som har gjort flest inlägg inkluderats, samt hur många trådar dessa användare har startat.

Till beräkningarna av gradcentralitet lades även normaliserad gradcentralitet till för att öka jämförbarheten med tidigare forskning. Detta definieras som gradcentralitet delat med det teoretiskt största möjliga antalet kopplingar inom nätverket och kommer att refereras till som $d'(v_i) = \frac{d(v_i)}{|V|-1}$. Största normaliserade gradcentralitet beräknades även för hackerforumen från den tidigare forskningen i jämförande syfte. Viktad gradcentralitet presenteras också på normaliserad form enligt $d'_w(v_i) = \frac{d_w(v_i)}{|V|-1}$.

Vikterna anses väsentliga för beräkningarna av egenvektorcentralitet, eftersom att denna ska ta hänsyn till grannar. Eftersom att alla noder i någon mån påverkar beräkningarna av egenvektorcentralitet ansågs ingen normalisering mot storlek vara nödvändig, och inga andra nätverk skulle jämföras med.

Det är standard för NetworkX att beräkna mellanliggande centralitet (NetworkX n.d.[b]) på oviktade grafer, där eventuella vikter försummas. Som skrivet i del 4.3.1 representerar vikten antalet trådar som två användare båda har gjort inlägg på. Det hade gått att invertera vikterna för att representera distanser, men detta anses inte uttrycka något tillräckligt meningsfullt inom kontexten av ett forum. Därmed har standardläget för NetworkX bibehållits, och kanterna betraktas som oviktade för dessa beräkningar. Mellanliggande centralitet är redan en andel och ska inte jämföras med andra nätverk, så normalisering ansågs inte vara nödvändig.

4.3.4 Strukturell beskrivning

Från teorin för den strukturella beskrivningen i del 2.4 inkluderas samtliga metoder. Notera att densitet är det enda måttet som direkt representerar sammanhållning, men resultaten i tidigare delar kan bidra med indikationer om detta. Även en graf över fördelningen av gradcentralitet och en graf som illustrerar gradcentralitetsfördelningen på utvalda subforum har inkluderats. Dessa har delat upp noderna i grupper efter hur stor deras gradcentralitet är i procent jämfört med den största uppmätta värde, för att undersöka hur vanliga värden på gradcentralitet i olika storleksordningar är.

Centraliseringsbegreppen viktades inte, eftersom att detta inte anses vara representativt. Till exempel ska inte en väldigt stark relation mellan två individer kunna påverka hur centraliserat hela nätverket anses vara.

Kategori	*Resultat*
Allmän deskriptiv statistik	Antal noder, $\|V\|$
	Antal kanter, $\|E\|$
	Antal komponenter, k
	Andel noder i största komponent, $V_{\%}$
	Storleksjämförelse med tidigare forskning
	Visualisering: Hela grafen, G
	Visualisering: Största komponent V_{max}
Noders positionering	Gradcentralitet, $d(v_i)$
	Normaliserad gradcentralitet*, $d'(v_i)$
	Egenvektorcentralitet, $z(v_i)$
	Mellanliggande centralitet $b(v_i)$
	Antal inlägg och antal startade trådar
Strukturell beskrivning	Densitet, D
	Genomsnittlig grad, $\bar{d}$
	Gradcentralisering, C_d
	Mellanliggande centralisering, C_b
	Gradcentralitetsfördelningar

Tabell 4: Sammanfattning av vilka resultat som kommer att presenteras, fördelat på kategori.
*Även för forumen presenterade i den tidigare forskningen

4.4 Begränsningar

En SNA är ofrånkomligen en förenkling av verkligheten. Denna del kommer att diskutera några mer specifika problem med datamaterialet och operationaliseringen i denna studie. I del 3.2 nämndes det fyra aspekter som gör SNA av kriminella nätverk problematiskt, datamaterialet är ofta; stort, inkomplett, dåligt avgränsat, och dynamiskt (Sparrow 1991).

Underlaget för denna studie är stort, vilket presenterar vissa svårigheter för potentiellt intressanta beräkningar. Exempelvis hade så kallad *closeness-centralitet* (Bichler 2019, s. 187) i kunnat vara av intresse, men beräkningarna blev för tunga. Det gör det även svårare att överblicka nätverket. I ett mindre kriminellt nätverk är det troligen lättare att identifiera strukturella mönster, eftersom att det finns mindre utrymme och möjlighet för större och mer komplexa strukturer.

Däremot är datamaterialet väl avgränsat, då ett specifikt forum under en specifik tidsperiod studeras. Det finns dock andra perspektiv, enligt vilka datamaterialet är att betrakta som sämre avgränsat. Det finns ingen garanti för att en kriminell gruppering på forumet endast kommunicerar där, eller att denna inte överlappar med andra forum och användare i andra medium. Det är snarare osannolikt att forumet som studeras motsvarar ett väldefinierat och organiserat kriminellt nätverk i den traditionella bemärkelsen. Med andra ord, eftersom att det är forumet i sig som är studieobjektet för denna studie betraktas materialet som väl avgränsat, men betraktas potentiella kriminella grupperingar på forumet i en bredare bemärkelse är det mycket sämre avgränsat. Ett forum där illegal hackerrelaterad information utbytes anses vara ett intressant studieobjekt för CTI-fältet, även om det inte överensstämmer med mer typiska kriminella nätverk.

Samma diskussion som gjorts ovan är också applicerbar på huruvida materialet för denna studie är att betrakta som inkomplett. Om forumet i sig betraktas som nätverket som ska studeras, är datan ytterst komplett och endast begränsat över en tidsperiod. Det finns dock som sagt inga garantier för att användarna inte kommunicerar på andra sätt, eller hackare i andra medium.

Slutligen bör ett nätverks dynamik tas i åtanke. Analyserna som görs inom denna studie är statiska. Hela tidsperioden som datamaterialet betraktas, men inte hur det har förändrats med tiden. Resultatens mer direkta tillämpningar är därmed nästintill utdaterade redan när de producerats, men målet var aldrig att resultaten skulle kunna användas i några prakiska åtgärder.

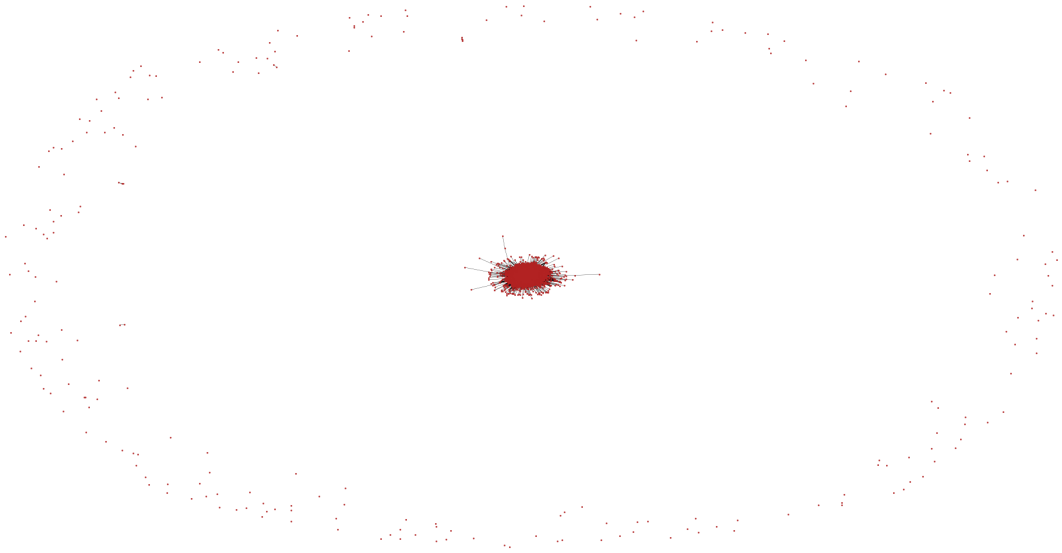
Det finns även dynamik i hur forumet fungerar som denna studie inte tar hänsyn till. Som diskuterat i del 4.3.1 skapas en kant mellan två användare om de har gjort inlägg på samma tråd. Detta var den bästa operationaliseringen som kunde identifieras, och har använts i tidigare forskning (Pete et al. 2020), men det är inte egentligen inte en optimal representation av hur användarna interagerar. Samma tråd kan innehålla inlägg från en stor mängd användare, och det har inte gjorts något försök att identifiera om alla av dessa faktiskt kommunicerar med alla andra. Diskussionen på en tråd kan utvecklas med tiden, och kopplingen mellan ett tidigt och ett sent inlägg kan bli successivt tunnare. Det är möjligt att denna problematik kan motverkas med någon form av innehållsanalys eller manuell kodning, men detta ansågs vara orimligt inom denna studies omfång. En i verkligheten central individ är troligen central även i resultaten för denna studie, men mer exakta strukturer missas av denna operationalisering.

Det har också antagits att varje användarkonto representerar en, och endast en, unik individ. Detta är inte garanterat, men det är svårt (om inte omöjligt) att bekräfta.

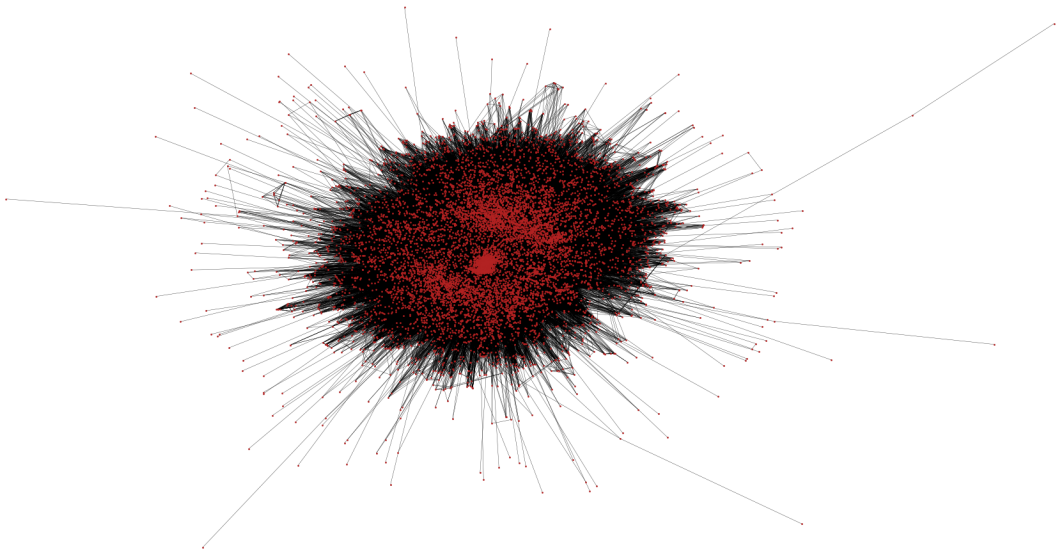
5 Resultat & Analys

5.1 Allmän deskriptiv statistik

Denna del kommer att presentera resultaten för den allmänna deskriptiva statistiken, enligt teorin i del 2.2 och metodiken i del 4.3.2. En illustration av hela nätverket och den största komponenten presenteras nedan i figur 4 och 5.



Figur 4: Graf som illusterar hela nätverket.



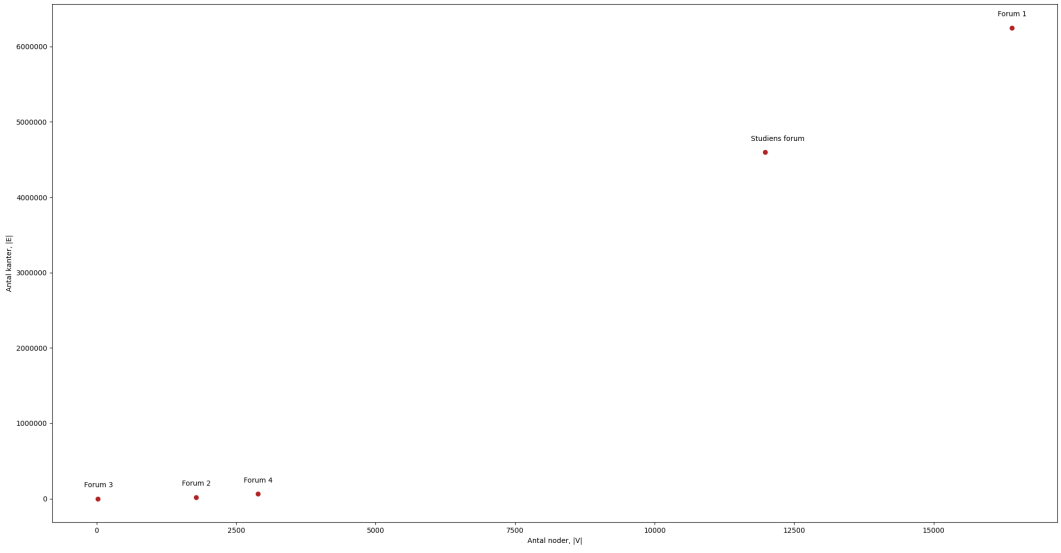
Figur 5: Graf som illusterar den största komponenten i nätverket.

Illustrationerna i figurerna ovan (4 och 5) är inkluderade för läsaren som är intresserad av hur nätverket ser ut i sin helhet. På grund av nätverkets storlek är det svårt att föra några resonemang eller analyser utifrån dessa. Istället fortsätter analysen nu direkt med den allmänna deskriptiva statistiken, resultaten av vilken presenteras i tabell 5.

Mått	*Värde*
Antal noder, $\|V\|$	11977
Antal kanter, $\|E\|$	4596702
Antal komponenter, k	241
Andel noder i största komponent, $V_{\%}$	0.9793

Tabell 5: Allmän deskriptiv statistik för nätverket; Antal noder, antal kanter, antal komponenter och andel noder i den största komponenten.

För att kontextualisera storleken på nätverket har en jämförelse med forumen (kallade 1-4) från studien av Pete et al. (2010) presenterade i tabell 2 inkluderats, se figur 6. Formumet i denna studie är stort, men inte det största av forumen inkluderade i figuren. Sammanhållning och densitet kommer att analyseras senare, men förhållandet mellan antal noder och kanter ser vid första anblick jämförbart ut med forumen från den tidigare forskningen.



Figur 6: Jämförelse av storlek mellan forumet i denna studie och de presenterade i tabell 2 (Pete et al. 2020).

Det går inte att direkt jämföra nätverkets 241 komponenter med de 5 komponenter i hackerforumet som studerades av Samtani & Chen (2016) (tabell 3), på grund av en markant storleksskillnad. Hade dock proportionerna mellan båda forumen varit samma (i förhållande till storleken) hade det förväntats ett mycket större antal komponenter än 241 inom forumet i denna studie. Detta antyder återigen en hög sammanhållning, eftersom att noderna verkar vara fördelade över ett i jämförelse litet antal komponenter.

Detta kan antingen bero på att forumet i denna studie har hög sammanhållning, eller för att större nätverk är mer sannolika att ta denna struktur. Den tidigare forskningen har dock observerat ett inverterat förhållande mellan sammanhållning och storlek (Pete et al. 2020), så i detta skede framstår en spekulation om att nätverket har en relativt hög sammanhållning rättfärdigad.

Det är även svårt att göra en rättvis jämförelse mellan de två nätverkens andel noder i största komponenten på grund av den markanta storleksskillnaden, men ca. 29% jämfört med ca. 98% talar ändå för spekulationen att denna studies forum har hög sammanhållning. 98% är en stor andel i egen rätt också, och behöver kanske egentligen inte jämföras för att bidra till (men inte fastställa) resonemanget om hög sammanhållning.

5.2 Noders positionering

I detta avsnitt presenteras resultaten av centralitetsberäkningarna för noderna i nätverket. Se teorin som beskriver noders positionering i del 2.3 och metoden i 4.3.3 för hur dessa beräknas.

nr.	*Användarnamn*	$d(v_i)$	$d'(v_i)$	$d'_w(v_i)$
1	unrated	5047	0.421	0.025
2	amirk2	4005	0.334	0.015
3	star41	3909	0.326	0.001
4	westy27	3890	0.325	0.992
5	harryj	3875	0.324	0.246
6	solocum	3826	0.319	0.451
7	persian	3770	0.315	0.001
8	lindes11	3746	0.313	0.352
9	mathsraper	3584	0.299	0.026
10	haloangel882	3539	0.296	0.041

Tabell 6: Användarnamn, gradcentralitet ($d(v_i)$), normaliserad gradcentralitet ($d'(v_i)$), och normaliserad viktad gradcentralitet ($d'_w(v_i)$) för de mest centrala användarna i termer av gradcentralitet. I fallande ordning av gradcentralitet.

För att möjliggöra en jämförelse av gradcentralitet med den tidigare forskningen har den största gradcentraliteten från dessa översatts till sitt centraliserade värde enligt del 4.3.3, se tabell 7. Notera att gradcentraliteten redan var i en normaliserad form i studien av Lu et al. (2010) och att detta värde har tagits direkt från tabell 1.

Som tidigare nämnt har hög gradcentralitet inom hackernätverk och forum påstått korrelera med sannolikheten att invididen känner till och sprider information (Grisham et al. 2017; Lu et al. 2010).

Forum	*Största normaliserade gradcentralitet,* $\max(d'(v_i))$
Denna studies forum	0.421
Lu et al. (2010)	0.977
Pete et al. (2020): Forum 1	0.952
Pete et al. (2020): Forum 2	0.353
Pete et al. (2020): Forum 3	0.667
Pete et al. (2020): Forum 4	0.416
Samtani & Chen (2016)	0.581

Tabell 7: Största normaliserade gradcentralitet för nätverken.

Den högsta normaliserade gradcentraliteten inom forumet är 0.421, vilket i jämförelse med den tidigare forskningen framstår som ett normalt och inte särskilt anmärkningsvärt resultat, då det förekommer både betydligt större och mindre värden. Det största värdet på normaliserad gradcentralitet är från nätverket *ShadowCrew* (Lu et al. 2010). Detta är som tidigare diskuterat ett relativt litet, väldefinierat, och strukturerat hackernätverk och en direkt jämförelse bör göras med försiktighet. Värdet är som nämnt beräknat av författarna och inte exakt redovisat, men det är utifrån deras beskrivningar osannolikt att det åsyftar något annat än normaliserad gradcentralitet som definierat i denna studie. En rimlig spekulation är att ett mindre och mer organiserat nätverk är mer sannolikt att ha noder som sticker ut i jämförelse med forum, trots författarnas slutsatser att nätverket är decentraliserat (Lu et al. 2010).

En jämförelse med forum 1 från studien av Pete et al. (2020) är däremot mer intressant. Detta är ett likt denna studies studieobjekt ett hackerforum och är av mer jämförbar storlek (se tabell 5 eller figur 6). Dess största normaliserade gradcentralitet är 0.952, vilket är ett stort värde nästan i linje med det från studien av Lu et al. (2022), och innebär att noden i fråga är kopplad till en övervägande majoritet av resterande noder. Detta anses något uppseendeväckande på grund av forumets storlek . Denna nod är nästintill extremt välkopplad inom forumet (med kanter till över 15000 noder), och sticker troligen ut som en tydlig ledare (ingen annan gradcentralitet redovisas i artikeln, men den genomsnittliga gradcentraliteten i forumet är endast 76.2). Poängen är att en största normaliserad gradcentralitet på 0.421 inte framstår som särskilt uppseendeväckande på ett forum av denna storlek, då det finns dokumenterade forum som både är större, och innehar mer centrala noder. Det ger dock en första inblick i vilka noder som kan vara centrala eller i någon mån viktiga eller inflytelserika på forumet.

De normaliserade viktade gradcentraliteterna i tabell 6 skiljer sig från de oviktade. De tre noder med högst gradcentralitet har en förhållandevis liten viktad gradcentralitet. Detta indikerar att medan de har haft kontakt med många andra noder på nätverket, är dessa kontakter inte lika frekventa eller upprepade.

Framförallt sticker noden på position 3 markant, med synnerligen små värden på viktad gradcentralitet i förhållande till oviktad. Användaren på position 4 i tabellen har tydligt störst viktad gradcentralitet, vilket antyder att denna användare inte bara

har haft kontakt med ett förhållandevist stort antal noder, utan att kontakten har varit återupprepad. I en bredare analys av forumet hade detta varit en intressant brytpunkt i mönstret och kunnat utgöra underlag för inriktning av analysen. Exempelvis hade en innehållsanalys av vad dessa återupprepade kontakter innehöll kunnat vara intressant. Det är dock tänkbart att detta värde endast kommer av att användaren i fråga har bredare intressen än övriga och är aktiv på en större variation av subforum, möjligen även de som är mindre relaterade till hackande.

Troligen ger den oviktade gradcentraliteten en bättre bild av nodens förmåga för informationsspridande, eftersom att denna representerar antalet andra användare noden har kontakt med. Däremot är det viktigt att inte försumma vikter då strukturen av nätverket ska kartläggas och ledande eller inflytelserika noder ska studeras, då denna bevisligen kan ge en signifikant förändrad bild av en nods centralitet.

nr.	*Användarnamn*	*Egenvektorcentralitet,* $z(v_i)$
1	unrated	0.0325
2	star41	0.0278
3	persian	0.0273
4	harryj	0.0258
5	lindes11	0.0255
6	sir	0.0252
7	westy27	0.0246
8	clarion55	0.0241
9	solocum	0.0240
10	pepelepe	0.0236

Tabell 8: Användarnamn och egenvektorcentralitet (viktad) för de mest centrala användarna i termer av egenvektorcentralitet.

I tabell 8 presenteras resultaten av beräkningarna av egenvektorcentralitet, och alla användarnamn utom tre är återkommande från tabell 6. Det som ger upphov till att nya noder inkluderas i denna lista (samt att ordningen skiljer sig något) är att värdet beror på vilka kontakterna skedde med, och hur centrala de är. Denna studie har inte genomfört någon innehållsanalys, men i den tidigare forskningen har det inom hackernätverk identifierats små grupper av seniora och skickliga hackers med högst centralitet (Holt et al. 2012; Samtani & Chen 2016). Om detta är ett mönster som återfinns i andra forum och nätverk är det viktigt att inkludera egenvektorcentralitet i analysen, för att kontakt med mer centrala individer i sådana fall sannolikt innebär större tillgång till mer sofistikerad information och programvara. Om denna mindre grupp är medvetna om sin status i dessa avseenden och är avsiktligt organiserade som någon form av ”inre krets” kan egenvektorcentraliteten även att bidra till att identifiera dem.

Egenvektorcentralitet har påstås vara ett mer direkt mått på influens inom ett forum (Grisham et al. 2017). Det är författarens åsikt att denna tolkning bör göras försiktigt. Centrala individer behöver per definition vara aktiva och produktiva forumet för att uppnå sin centralitet. De har sedan ingen kontroll för vilka användare som länkas till dem, eftersom att det antas att alla användare är fria att göra inlägg på vilka trådar de vill. En ny användare kan alltså koppla sig till ett stort antal inflytelserika användare snabbt och signifikant öka sin egenvektorcentralitet, utan att de mer seniora användarna ens har noterat dennes existens. Den nya noden i detta hypotetiska scenario är signifikant för att den (speciellt om centrala användare antas vara skickligare och mer seniora) kan tillgå information och resurser, men den kan inte per automatik påstås vara inflytelserik.

Det finns alltså något varierande tolkningar av vad egenvektorcentraliteten innebär inom detta forum, men ett inkluderande i tabell 8 bidrar till att identifiera noden i fråga som en intressant användare inom forumet. Däremot är det tveksamt om detta ensamt indikerar att noden i fråga är inflytelserik, utan snarare att de är intressanta för att de har interagerat med individer som är centrala, och potentiellt inflytelserika och seniora.

nr.	*Användarnamn*	*Mellanliggande centralitet,* $b(v_i)$
1	unrated	0.0489
2	_2540echo_257Fuw8ecr	0.0311
3	amirk2	0.0180
4	star41	0.0109
5	persian	0.0101
6	westy27	0.0097
7	hujoczita	0.0095
8	multicracker	0.0091
9	harryj	0.0080
10	solocum	0.0071

Tabell 9: Användarnamn och mellanliggande centralitet för de mest centrala användarna i termer av betweenness-centralitet.

Resultaten för beräkningarna av mellanlligande centralitet är angivna i tabell 9 ovan. Här påminns att detta represnterar hur stor andel av gångerna noden ligger på den kortaste vägen mellan andra noder. Återigen är det lätt att konstatera att många av namnen som är listade i de tidigare tabellerna även återkommer här och att samma användare återigen är den mest centrala.

Generellt sett kan en hög mellanliggande centralitet som tidigare nämnt indikera en medlande roll i ett nätverk, men det är något som bör tolkas mer försiktigt i en analys av ett forum. Det finns mindre begränsningar för medlemmar på ett forum att kontakta varandra än det kan göra i ett mer traditionellt (analogt) organiserat nätverk. Däremot är det inte säkert att denna möjlighet utnyttjas bara för att den existerar, vilket till exempel fanns av Samtani & Chen (2016) som drog slutsatsen att tillgången till expertis och verktyg framstod som underutnyttjad inom forumet de studerade. Det är därför inte otänkbart att en hög mellanliggande centralitet även på ett forum kan indikera en medlande roll. Det begränsade besökandet som kunde göras av forumet antydde att en användare i vissa fall behöver svara med ett inlägg på en tråd för att kunna tillgå allt bifogat material (går att observera i bilaga A.1). Alltså; om en användare vill ha fullständig tillgång till allt som ett inlägg innehåller krävdes det att användaren själv gjorde ett inlägg på samma tråd, vilket hade dykt upp som en kant i SNA:n gjord av denna studie. Detta styrker i viss mån en mellanliggande central nods möjligheter att agera som en förmedlare av information och resurser inom nätverket.

Det finns som tidigare diskuterat en tolkning av ett nätverks kanter som inte endast kanaler för spridningen av information och resurser, utan som ömsesidigt utövande av influens. Det går inte att veta i vilken utsträckning användarna läser varandras inlägg utan att kommentera, och en grundlig kartläggning av influens inom forumet är omöjlig, för att inte nämna svårigheterna med att mäta influens från början. Det bör dock poängteras att det är möjligt att de mellanliggande-centrala individerna kan ha en medlande, eller möjligen brobyggande och sammanhållande funktion i forumet i termer av influens, idéer och synpunkter. Detta hade varit en intressant och signifikant position i nätverket, men sådana slutsatser är spekulativa.

Innan denna analys fortskrider till den strukturella beskrivningen av nätverket framstår en jämförelse med användarnas grundläggande statistik som nödvändig. Delvis för att placera analysen i perspektiv, och delvis för att i någon liten utsträckning validera SNA-metodiken som använts. Se tabell 10 för en sammanställning av användarnas produktivitet på forumet.

Inte överraskande så återfinns återigen samma användarnamn längst upp i tabell 10 som toppar alla resultat förutom normaliserad viktad gradcentralitet. Detta är naturligt eftersom att det redan konstaterats att denna, och andra som återfinns här, behöver vara produktiva på forumet för att erhålla sin centrala position. Det som är av intresse är att illustrera att tabell 10 inte berättar hela historien, och att det en enkel räkning av inlägg missar många av de viktiga strukturella aspekterna hos noderna, även om båda baseras på undersökning av inlägg på trådar. Det finns på ena sidan ett flertal användare som gör många inlägg utan att vara centrala användare ur ett SNA-perspektiv, och på andra användare som har tagit centrala positioner i nätverket utan att vara bland de mest produktiva i termer av antal inlägg.

nr.	*Användarnamn*	*Antal inlägg*	*Antal startade trådar*
1	unrated	1040	862
2	_2540echo_257Fuw8ecr	422	279
3	amirk2	289	239
4	persian	217	77
5	multicracker	207	143
6	slip_knot	187	39
7	star41	174	160
8	tresplot	156	154
9	rare	139	38
10	thelightishere	137	62

Tabell 10: Användarnamn, antal inlägg, och antal startade trådar för de användare med flest inlägg på forumet. Sorterat efter antal inlägg.

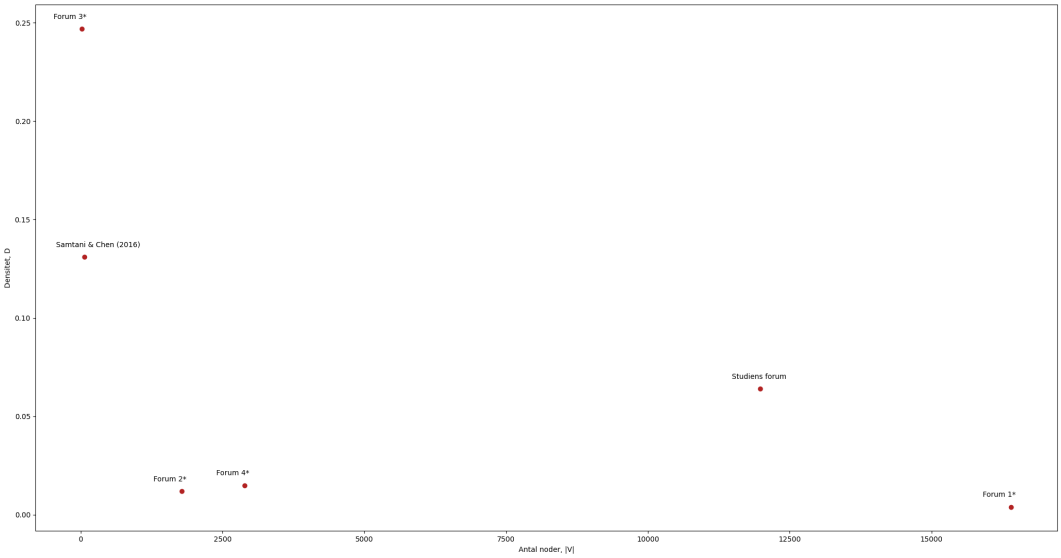
5.3 Strukturell beskrivning

I denna del presenteras resultaten för den strukturella beskrivningen av nätverket. Se del 2.4 och 4.3.4 för beskrivningar av hur dessa beräknas och framställs.

Mått	*Värde*
Densitet, D	0.0641
Genomsnittlig grad $\bar{d}$	767.588
Gradcentralisering, C_d	0.357
Betweenness-centralisering, C_b	$3.405 * 10^{-10}$

Tabell 11: Resultaten för strukturell beskrivning av hela nätverket; Densitet, genomsnittling grad, gradcentralisering och mellanligande centralisering.

Som tidigare beskrivit kretsar den strukturella beskrivningen av nätverket främst kring dess sammanhållning och centralisering, men andra undersökningar kommer också att vara relevanta. För att beskriva nätverkets sammanhållning har dess densitet beräknats, se tabell 11. En jämförelse av nätverkets densitet och storlek (i termer av antal noder) har inkluderats nedan i figur 7.



Figur 7: Jämförelse av densitet i förhållande till storlek mellan forumet i denna studie och de presenterade i tabell 2 och 3.
*Forum 1-4 refererar till de forum som inkluderades från studien av Pete et al. (2020) i tabell 2.

Densiteten för forumet är 0.0641, vilket är vid första anblick framstår som lågt, och antyder att spekulationen i del 5.2 var missvisande. När detta värde jämförs med de andra forumen från den tidigare forskningen framstår det dock som en rimlig och normal densitet, varken största eller minst. Det som är något uppseendeväckande är att densiteten är relativt hög med forumets storlek i åtanke och det är endast de betydligt mindre nätverken som har en högre densitet. Det verkar därför som den tidigare spekulation baserat på den allmänna deskriptiva statistiken hade en viss poäng, och att forumet är sammanhållet i förhållande till sin storlek. Detta stämmer speciellt om observationen som Pete et al. (2020) gjorde om densiteten och storlekens inverterade förhållande stämmer i en bredare kontext.

Det första och enklaste måttet av ett nätverks centralisering som används i denna studie är genomsnittlig gradcentralitet, som av tabell 11 kan utläsas vara ca. 768. Det framstår direkt som ett synnerligt högt värde, då noderna i de forum som presenterats från den tidigare forskningen (tabell 2-3) i genomsnitt är länkade till mellan ca. 5 och 76 andra noder. Det är dock inte ett otänkbart värde, då det har demonstrerats i tabell 6 att den högsta gradcentraliteten är på över 5000, och samtliga värden i tabellen är i tusental.

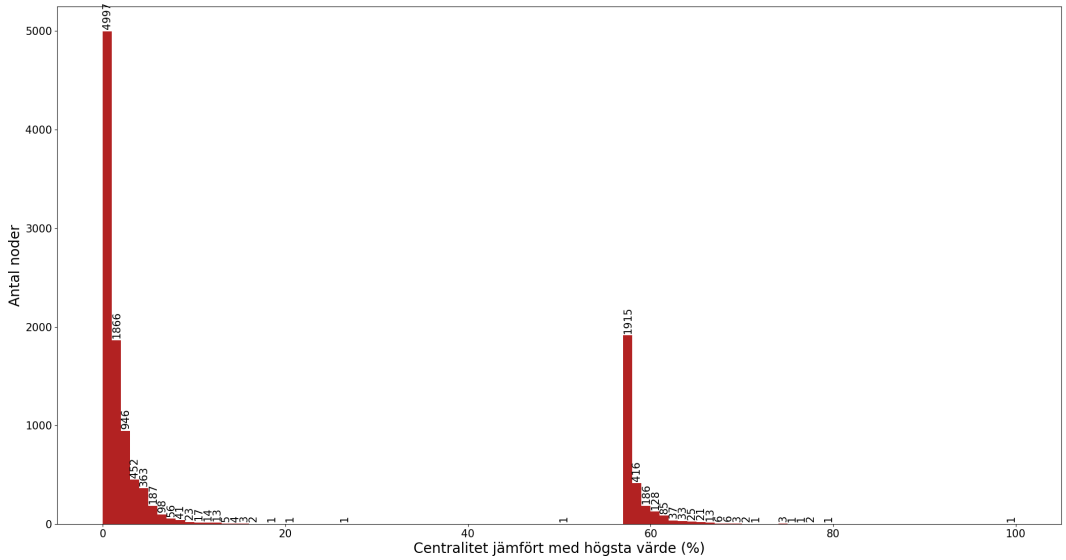
Förutom att detta återigen kan tolkas som ett styrkande av påståendet att nätverket har förhållandevis hög sammanhållning är detta det första konkreta antydandet att nätverket även är decentraliserat. En högre genomsnittlig grad (högre i förhållande till storleken) är associerad med en decentraliserad struktur, eftersom att ju fler andra noder en genomsnittlig nod är länkade med, desto mindre sannolikt är det att centraliteten i nätverket vara koncentrerad kring ett fåtal noder. Om den genomsnittliga gradcentraliteten ställs i proportion till nätverkets storlek är det endast de två betydligt mindre nätverken som har en i förhållande större genomsnittlig gradcentralitet. Forumet i denna studie har en genomsnittlig gradcentralitet som är större i förhållande till sin storlek än samtliga forum med en storlek i tusental.

När det kommer till en resultatet för beräkningen av forumets gradcentralisering finns det tyvärr inte lika mycket att jämföra med, då den enda inkluderade studien som innehåller detta resultat är studien av hackernätverket *ShadowCrew* (Lu et al. 2010). Som tidigare nämnt drogs slutsatserna att detta nätverk var decentraliserat, men i vissa avseenden välorganiserat. Forumet i denna studie har en gradcentralisering på 0.357 och är, trots den potentiella friheten för spontana kopplingar som ett forum medför, faktiskt i termer av gradcentralisering tydligt mer centraliserat än *ShadowCrew* som har en gradcentralisering på 0.269. Detta är intressant eftersom att *ShadowCrew* hade en betydligt större största normaliserad gradcentralitet. Dock är skillnaden inte alltför stor och resultatet är långt från 1, så forumet bör antagligen betraktas som mer decentraliserat i än centraliserat termer av gradcentralisering. Möjligen är större nätverk mer sannolika att vara centraliserade på grund av utrymmet som ges för bildandet av inre grupperingar, vilket kan bidra till en mindre jämn fördelning av centralitet. Detta är en tänkbar förklaring till skillnaderna i gradcentralisering och genomsnittlig gradcentralitet som framtsår i jämförelserna. Det är även möjligt att en decentraliserad struktur var ett medvetet val hos *ShadowCrew*, samt att det är lättare att direkt kontrollera strukturen av ett mindre, icke-forumsbaserat nätverk.

Den mellanliggande centraliseringen är extremt liten och kan sägas närma sig noll, vilket antyder att nätverket är synnerligen decentraliserat efter denna måttstock. I detta hänseende är forumet tydligt mindre centraliserat än *ShadowCrew*. Det bör dock understrykas att även *ShadowCrew* fortfarande är att betraktas som tydligt decentraliserat i detta hänseende. Vad kan skillnaden i centraliseringsbegreppen bero på? En synnerligen liten mellanliggande centralisering och en betydligt större gradcentralisering innebär att det inom nätverket är någorlunda vanligt med noder som sticker ut i termer av gradcentralitet, men inte i termer av mellanliggande centralitet.

Med andra ord är kan det sägas att det är ganska vanligt med noder som är mer välkopplade inom nätverket än andra, men inte vanligt med noder som är positionerade mellan andra noder i en större utsträckning än resten. En möjlig förklaring är att det är svårt att positionera sig mellan andra noder i ett forum, eftersom att alla användare antagligen är helt fria att kontakta vem de vill och har möjlighet att göra inlägg på vilken tråd som helst. Det är därför svårt för en nod att medvetet positionera sig som en medlare, och det är troligen mindre sannolikt att mellanliggandecentrala positioner naturligt uppstår. Däremot är en enskild användare fri att bilda hur många kanter den vill med andra användare och som demonstrerat finns det tydliga skillnader i användarnas produktivetet på forumet (se tabell 10). Detta betyder att de mer aktiva användarna naturligt är mer sannolika att höja sin egen gradcentralitet i förhållande till övriga användare, och att gradcentraliseringen kommer att höjas.

Den låga mellanliggande centraliteten underminerar i viss utsträckningen analysen baserat på skillnader i mellanliggande centralitet som gjordes i föregående delen av analysen. Eftersom att nätverket är ytterst decentraliserat i detta hänseende bör skillnaderna i centraliteten betraktas som mindre signifikanta.



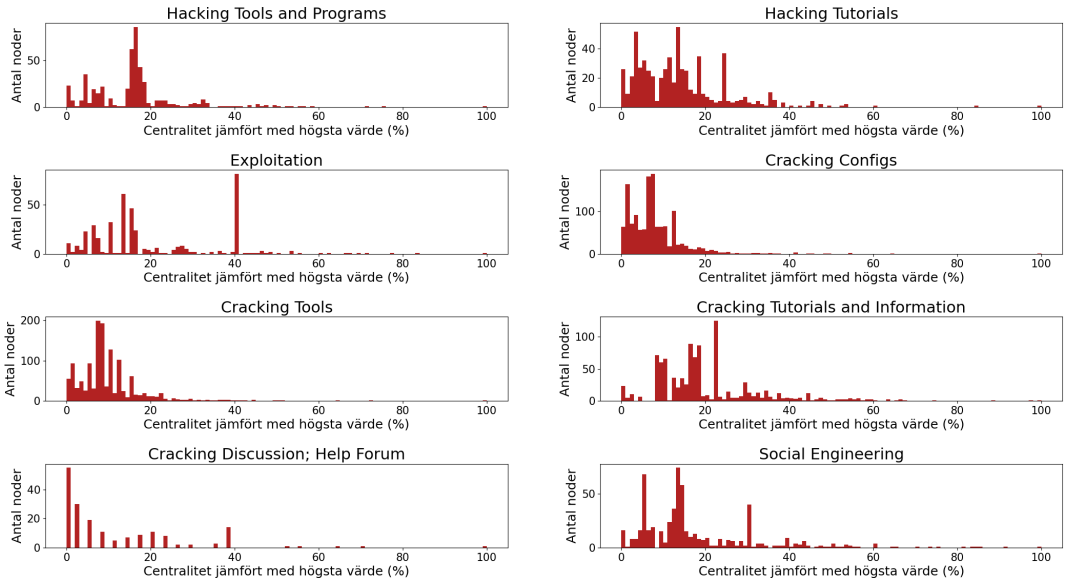
Figur 8: Gradcentralitetsfördelning för nätverket, med antal användare märkt på applicerbara percentiler.

Figur 8 ovan illustrerar gradcentralitetsfördelningen över hela forumet. Varje stapel illustrerar alltså antalet noder som har en gradcentralitet som faller inom varje procentuell andel av det högsta observerade värde (i hela procent). Markeringarna ovanför varje stapel visar det exakta antalet noder, men dessa är inte väsentliga och inkluderades till stor del för att tydligare markera de percentiler där stapeln är för liten för att synas.

Det är en vid första anblick nästintill extremt uppdelad fördelning. Det är två väldigt tydliga grupper med varsin specifik, väldigt tydlig percentil med en stor mängd användare. Det har tidigare diskuterats möjligheten att forumet är indelat i grupper av ökande storlek men minskande centralitet, senioritet, och skicklighet. Uppdelning i tre sådana grupper har observerats på hackerforum och i hackergemenskapen i bredare bemärkelse i den tidigare forskningen (Holt et al. 2012; Samtani & Chen 2016). Det finns dock inte tydliga belägg för att ett liknande fenomen kan förklara denna väldigt tydliga uppdelning på detta forum eftersom att det hade krävt en bredare analys, men det kan möjligen ha en bidragande effekt. Det är dock mer sannolikt, och det finns mer underlag för, att detta beror på forumets uppdelning i subforum. Som tidigare nämnt är forumet uppdelat på 32 subforum med över 10000 inlägg i det största. Fyra subforum har fler än 5000 inlägg och tio av dem har över 1000 inlägg. Samtliga av dessa är relaterade till hackande eller gratis tillgång till annars kostande programvara.

Med största sannolikhet skapar subforumen den stora klyftan i fördelningen av gradcentralitet, eftersom att de kan bidra till att en användare väldigt snabbt kopplas till en stor mängd andra användare ur ett SNA perspektiv. Om en ny användare ansluter sig till forumet och blir någorlunda aktiv på ett eller flera subforum kommer denna sannolikt göra inlägg på ett antal stora trådar och snabbt ansluta sig till hundratals användare. Till exempel har den (med god marginal) största tråden som hittades i datan har över 1000 inlägg, och genom att göra ett inlägg på denna och ett par andra har en användare direkt positionerat sig till höger i fördelningen. Ett par trådar har ca. 200 inlägg och många har över 100. Med andra ord är gradfördelningen anmärkningsvärd men inte fullt så extrem som den först kan framstå som.

Till sist gjordes liknande undersökningar av gradcentralitetsfördelningar hos några utvalda subforum. Dels de som ingår i det så kallade ”*hacking zone*”, samt ett par andra som framstod som relevanta. De har alltså inte valts med hänsyn till sin storlek, men de tillhör de större subforumen. Här observeras inte samma mönster som i figur 8, utan ser generellt mer ut som väntat. Detta styrker i någon mån förklaringen om att subforumsstrukturen har bidragit till fördelningen av centraliteten av hela forumet, snarare än att en tydlig hierarki existerar. Det är sannolikt att någon form av meritokratisk hierarki existerar på forumet, även om det som nämnt är svårt att spåra den i fördelningen i figur 8, och det är fullt möjligt att effekten av denna bara inte syns på individuella subforum. Sammantaget är det dock mer sannolikt att uppdelningen i subforum är en starkare bidragande faktor till hur fördelningen ser ut, och varför samma mönster inte återfinns på enskilda forum.



Figur 9: Gradcentralitetsfördelning utvalda subforum.

6 Slutsatser

Denna studie har ämnat att besvara vilka allmänna och strukturella egenskaper som forumet har, samt vilka noder som sticker ut som signifikanta och vad som karaktäriserar dem.

Det är ett förhållandevis men inte exceptionellt stort forum i förhållande till den tidigare forskningen. Det består av få komponenter i förhållande till sin storlek och en övervägande majoritet av noderna är inkluderade i den största komponenten. Båda dessa faktorer ger en tidig antydan att forumet är sammanhållet. Nätverkets densitet bekräftar detta endast delvist. Det framstår som att nätverket har hög sammanhållning men endast i förhållande till sin storlek, inte i absoluta termer. Forumet har en synnerligen hög genomsnittlig gradcentralitet, vilket i en liten utsträckning styrker resonemanget om relativ sammanhållning och framförallt en indikation på att nätverket är decentraliserat. Gradcentraliseringen antyder att nätverket är decentraliserat snarare än centraliserat, men med sämre marginal än för övriga mått. Slutligen är nätverket att betrakta som ytterst decentraliserat i termer av mellanliggande centralisering. Sammantaget är nätverket att betraktas som stort, decentraliserat, och sammanhållet.

En nod har högst värde för nästan samtliga värden som har beräknats, användaren *unrated*. Det råder inga tvivel om att detta var en ytterst signifikant användare på forumet under tidsperioden som har studerats, då denna var mest central i alla avseenden utom ett och var även mest produktiv. Andra noder återkom bland de tio högsta värdena för flertalet eller alla beräkningar, men med olika inbördes ordningar. En användare sticker ut signifikant i termer av viktad gradcenrtalitet, och hade också stor oviktad gradcentralitet, vilket indikerar att denna inte bara hade många kopplingar, utan även starkare sådana med återkommande kontakter. De noder som inte var lika centrala i andra avseenden men framträdde i beräkningarna av egenvektorcentralitet är intressanta för att interagerar med mer centrala och potentiellt seniora användare, vilket är ett viktigt kompletterande perspektiv vid analyser i CTI-syften. Slutligen indikerar mellanliggande centraliteten en grupp användare som potentiellt har en medlande roll inom nätverket, men vikten av denna slutsats undermineras av den låga mellanliggande centraliseringen.

7 Referenser

Archive, Internet (n.d.). *About the Internet Archive*. URL: https://archive.org/about/.

Axmark, David & Michael Widenius (2023). *MySQL 5.7 Reference Manual*. URL: https://dev.mysql.com/doc/refman/5.7/en/.

Baker, Wayne E. & Robert R. Faulkner (1993). "The Social Organization of Conspiracy: Illegal Networks in the Heavy Electrical Equipment Industry". In: *American Sociological Review* 58.6, pp. 837–860. ISSN: 00031224. URL: http://www.jstor.org/stable/2095954 (visited on 11/02/2023).

Barbosa de Almeida, Mauro W. (2015). "Structuralism". In: ed. by James D. Wright, pp. 626–631. DOI: https://doi.org/10.1016/B978-0-08-097086-8.12225-1.

Berkowitz, Stephen D (1982). *An introduction to structural analysis: The network approach to social research*. Butterworth & Co. (Canada) Ltd. ISBN: 0-409-81362-1.

Bichler, Gisela (2019). *Understanding Criminal Networks*. Oakland, California: University of California Press. ISBN: 978-0-520-29705-0.

Freeman, Linton (1978). "Centrality in social networks: Conceptual clarification". In: *Social network: critical concepts in sociology. Londres: Routledge* 1, pp. 238–263.

Freeman, Linton (Jan. 2004). *The Development of Social Network Analysis*. BookSurge, LLC. ISBN: 1-59457-714-5.

Grisham, John et al. (2017). "Identifying mobile malware and key threat actors in online hacker forums for proactive cyber threat intelligence". In: *2017 IEEE international conference on intelligence and security informatics (ISI)*. IEEE, pp. 13–18.

Hagberg, Aric A., Daniel A. Schult, & Pieter J. Swart (2008). "Exploring Network Structure, Dynamics, and Function using NetworkX". In: *Proceedings of the 7th Python in Science Conference*. Ed. by Gaël Varoquaux, Travis Vaught, & Jarrod Millman. Pasadena, CA USA, pp. 11–15.

Holt, Thomas J et al. (2012). "Examining the social networks of malware writers and hackers." In: *International Journal of Cyber Criminology* 6.1.

Hunter, J. D. (2007). "Matplotlib: A 2D graphics environment". In: *Computing in Science & Engineering* 9.3, pp. 90–95. DOI: 10.1109/MCSE.2007.55.

Jordan, Tim & Paul Taylor (1998). "A sociology of hackers". In: *The Sociological Review* 46.4, pp. 757–780.

Koschade, Stuart (2006). "A social network analysis of Jemaah Islamiyah: The applications to counterterrorism and intelligence". In: *Studies in Conflict & Terrorism* 29.6, pp. 559–575.

Krebs, Valdis E (2002). "Mapping networks of terrorist cells". In: *Connections* 24.3, pp. 43–52.

Lu, Yong et al. (2010). "Social network analysis of a criminal hacker community". In: *Journal of Computer Information Systems* 51.2, pp. 31–41.

McKinney, Wes (2010). "Data Structures for Statistical Computing in Python". In: *Proceedings of the 9th Python in Science Conference*. Ed. by Stéfan van der Walt & Jarrod Millman, pp. 56–61. DOI: 10.25080/Majora-92bf1922-00a.

NetworkX (n.d.[a]). *adjecency_matrix*. URL: https://networkx.org/documentation/stable/reference/generated/networkx.linalg.graphmatrix.adjacency_matrix.html.

NetworkX (n.d.[b]). *betweeness_centrality*. URL: https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.centrality.betweenness_centrality.html.

NetworkX (n.d.[c]). *eigenvector_centrality*. URL: https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.centrality.eigenvector_centrality.html.

Otte, Evelien & Ronald Rousseau (2002). "Social network analysis: a powerful strategy, also for the information sciences". In: *Journal of information Science* 28.6, pp. 441–453.

Pete, Ildiko et al. (2020). "A social network analysis and comparison of six dark web forums". In: *2020 IEEE European symposium on security and privacy workshops (EuroS&PW)*. IEEE, pp. 484–493.

Robert, Schweitzer, Zhang Ning, & Ebrahimi Mohammadreza (2018). *Hacker Web Forum Collection: CrackingArena Forum Dataset*. 8 april 2013 – 24 februari 2018. URL: https://www.azsecure-data.org/other-forums.html.

Samtani, Sagar & Hsinchun Chen (2016). "Using social network analysis to identify key hackers for keylogging tools in hacker forums". In: *2016 IEEE conference on intelligence and security informatics (ISI)*. IEEE, pp. 319–321.

Sparrow, Malcolm K (1991). "The application of network analysis to criminal intelligence: An assessment of the prospects". In: *Social networks* 13.3, pp. 251–274.

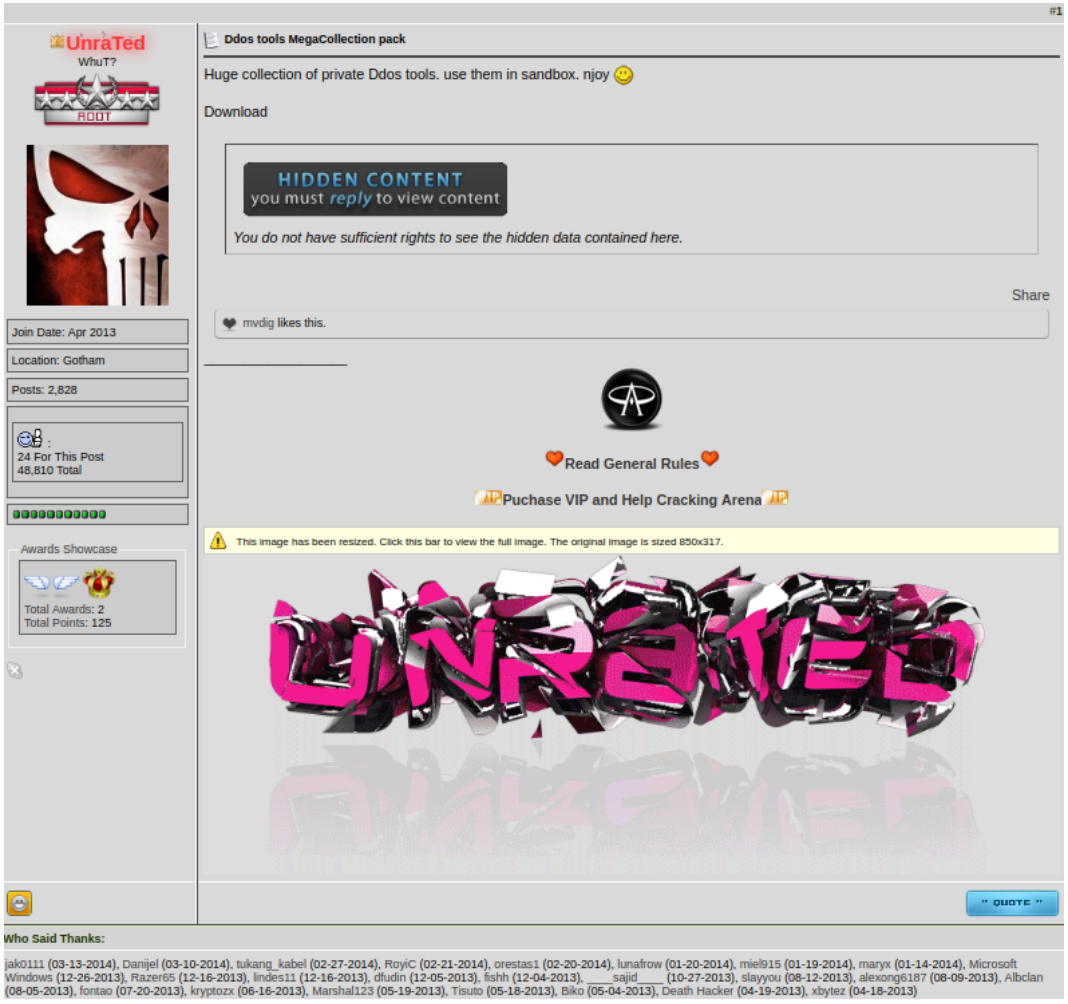
University of Arizona (n.d.). *SaTC: Hacker Web*. URL: https://eller.arizona.edu/departments-research/centers-labs/artificial-intelligence/research/previous/cyber.

Zhang, Mingxin (2010). "Social network analysis: History, concepts, and research". In: *Handbook of social network technologies and applications*, pp. 3–21.

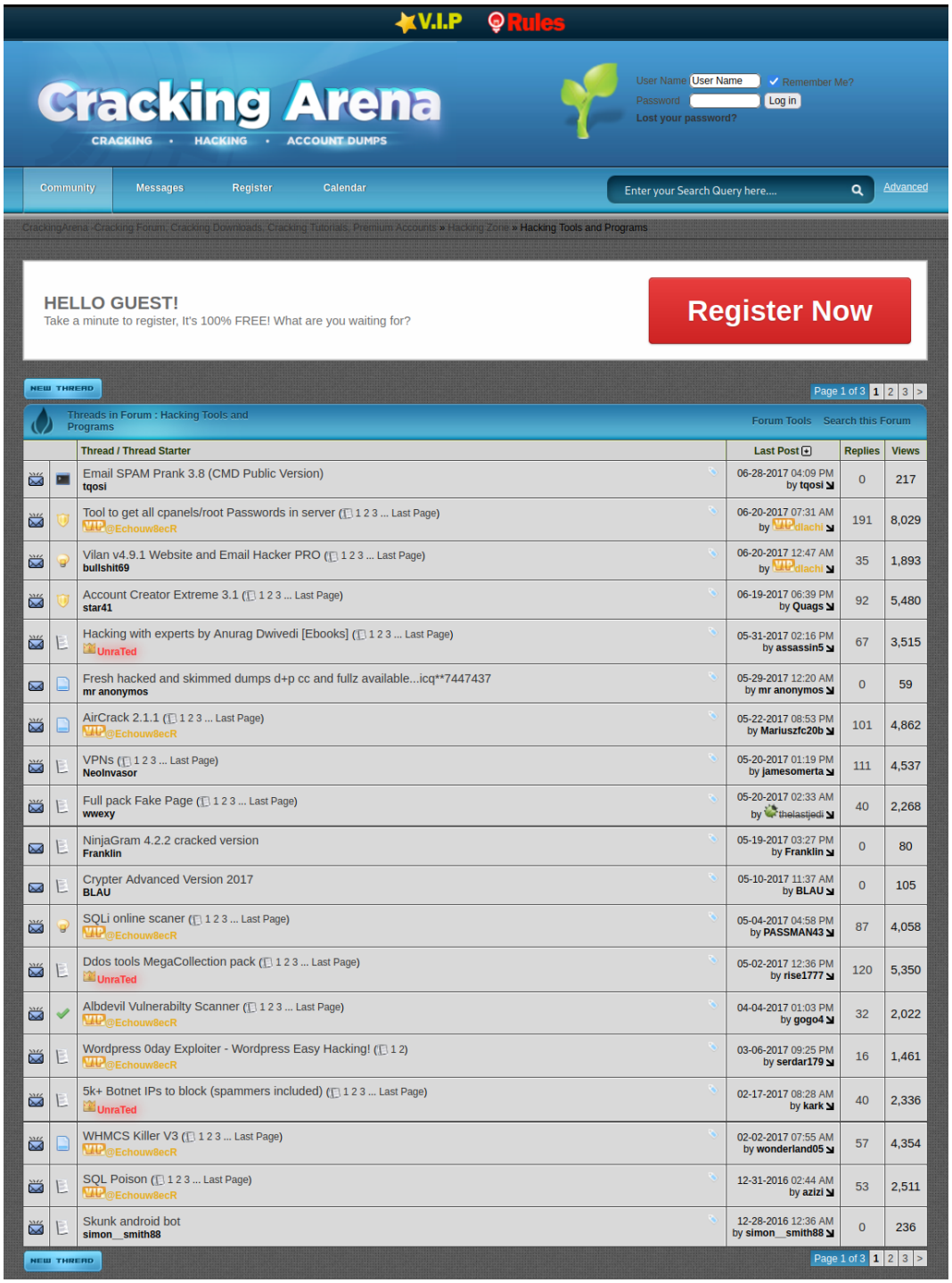
A Bilagor

A.1 Bilder av forumet

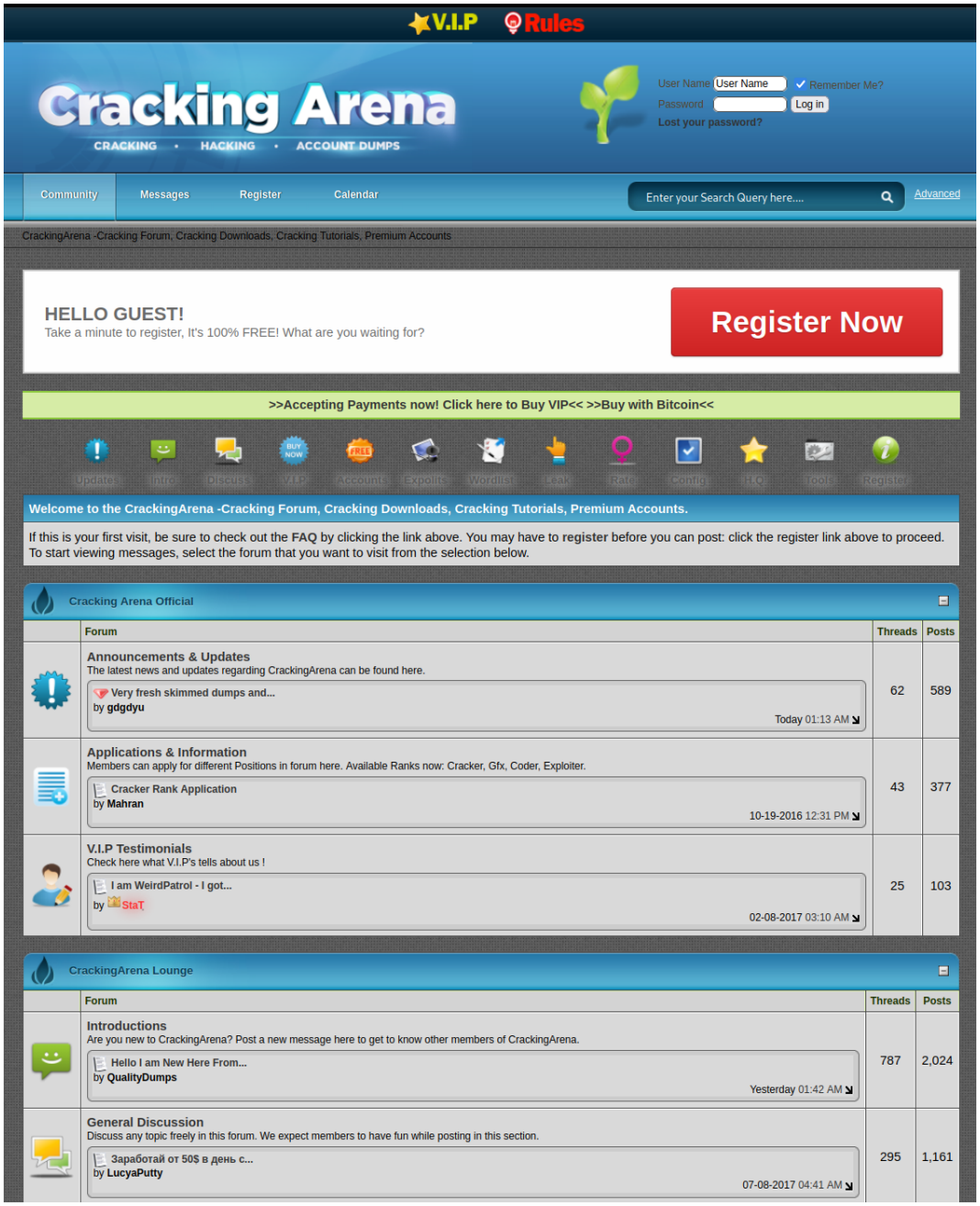
Forumet är inte längre åtkomligt från de URL:er som fanns bland datamaterialet. Bilderna av forumet är skärmdumpar tagna med hjälp av Wayback Machine från Internet Archive (Archive n.d.).



Figur A.1: Skärmdump ett inlägg från forumet, tyvärr är delar av innehållet blockat tills användaren har svarat, vilket idag är omöjligt.



Figur A.2: Skärmdump av första sidan för subforumet *Hacking Tools and Programs*.



Figur A.3: Skärmdump av forumets startsida.