Środowisko do identyfikowania wzorców zachowań w oparciu o podejście sieci społecznych

Wojciech Piekaj[1], **Grzegorz Skorek**[1], **Anna Zygmunt**[1], **Jarosław Koźlak**[1]

Streszczenie: Artykuł opisuje zastosowanie technik analizy sieci społecznych oraz analizy częstych sekwencji do identyfikacji wzorców zachowań w organizacji przestępczej na podstawie transakcyjnej bazy danych billingów telefonicznych. W artykule przedstawiono stan badań nad sieciami społecznymi w kryminalistyce ze szczególnym uwzględnieniem osiągnięć w zakresie identyfikacji ról. Na potrzeby badań wprowadzono kilka nowych ról. Zrealizowano i opisano system KASS do identyfikacji ról i sekwencji. Przeprowadzono wstępne testy potwierdzające poprawność działania.

Słowa kluczowe: SNA, sieci społeczne, sieci kryminalne, centralność, role, eksploracja danych

1. Wstęp

Zorganizowana przestępczość tworzy rozbudowane, silnie zhierarchizowane struktury. Aby z nią walczyć, konieczne jest zidentyfikowanie całej siatki przestępczej i rozpoznanie ról pełnionych w niej przez poszczególne osobniki. Metod do takiej analizy dostarcza prężnie się ostatnio rozwijająca dziedzina z pogranicza socjologii, psychologii i matematyki jaką są sieci społeczne (ang. Social Network Analysis - SNA). W niniejszej pracy zajmiemy się analizą sieci społecznych tworzonych przez przestępców wykorzystując informacje o charakterze prowadzonych przez nich rozmów telefonicznych oraz zaprezentujemy metody identyfikacji ról poszczególnych osobników na podstawie przejawianych przez nich zachowań.

2. Sieci społeczne i ich zastosowania

Charakterystyczną cechą SNA jest przywiązywanie mniejszej wagi do wewnętrznych cech danego osobnika - analizowane jest głównie jego położenie względem innych osobników w sieci społecznej. Analiza sieci społecznych próbuje odpowiedzieć na pytanie o wpływ struktury sieci na zachowanie jednostek (patrz Wellman (1999) - podstawowe paradygmaty sieci społecznych). Metody SNA są szeroko stosowane w wielu dyscyplinach, przy analizie bardzo różnych rodzajów zjawisk. Badając podstawowe parametry sieci, takie jak np. indeksy rozkładu, miary centralności, podobieństwa węzłów czy strukturę społeczności można opracować modele analityczne obrazujące rozprzestrzenianie się chorób, formowanie grup, rozprzestrzenianie się wirusów komputerowych czy przeglądania informacji na WWW. Traktowanie sieci społecznej jako grafu pozwala wykorzystać grafowe algorytmy eksploracji danych (ang. link mining) (Jensen (2002), Wolfe (2004)).

[1] Katedra Informatyki, Akademia Górniczo-Hutnicza, Al. Mickiewicza 30, 30-059 Kraków
e-mail: {piekaj,skorek}@student.agh.edu.pl, {azygmunt,kozlak}@agh.edu.pl

3. Sieci społeczne w badaniach kryminalnych

Wydarzenia z 11 września 2001 wpłynęły na wzrost zainteresowania analizą sieci społecznych. Okazało się bowiem, iż atakom można było zapobiec, poprzez dokładną analizę społecznych sieci terrorystycznych. W latach 2003-2005 na Uniwersytecie opracowano system Coplink (Xu (2004)), który miał pomagać analitykom policyjnym w ich pracy nad billingami telefonicznymi. System CrimeNet Explorer (Xu 2005a) wykrywa podgrupy w sieciach kryminalnych, a NETEST (Dombroski 2002) - ustala strukturę sieci w rzadkich kryminalnych i terrorystycznych sieciach. Z kolei John Arquilla i David Ronfeldt (Arquilla (2002)) zaproponowali kilka ról, jakie mogą odgrywać w sieci społecznej przestępcy. Są to:

- *Organizatorzy* – stanowią rdzeń całej organizacji. Czasami są to małe grupy sterujące działaniem całej sieci.
- *Izolatorzy* – pojedyncze osobniki lub małe grupy, których rolą jest zasadniczo izolować rdzeń organizacji (organizatorów) przed infiltracją. Jednostki te są również odpowiedzialne za transmisję informacji i zarządzeń od rdzenia do brzegów sieci.
- *Komunikatorzy* – pojedyncze osobniki odpowiadające za przepływ informacji pomiędzy dwoma węzłami w całej sieci. Są oni odpowiedzialni za transmisję rozporządzeń z rdzenia i powracających do niego odpowiedzi.
- *Strażnicy* – koncentrują się na bezpieczeństwie sieci i minimalizują podatność sieci na zewnętrzne ataki lub infiltrację. Wymuszają lojalność wszystkich osobników i członków ich rodzin. Zapobiegają również dezercji osobników z sieci, a gdy taka sytuacja nastąpi minimalizują ryzyko z tym związane - starają się naprawić powstałe w sieci uszkodzenia i nieprawidłowe działanie.
- *Rozszerzający* – zajmują się poszerzeniem sieci, poprzez rekrutowanie nowych członków oraz poprzez łączenie innych sieci. Zachęcają do współpracy osobniki z innych sieci (np. prawników, policjantów, polityków). Kiedy im się to udaje, sieć zyskuje nowe źródła informacji, co wpływa na poprawę jej działania.
- *Monitorujący* – są odpowiedzialni za wydajność sieci. Raportują słabości sieci i jej problemy do organizatorów, którzy dzięki temu mogą podjąć kroki naprawcze. Odpowiadają za poprawę działania sieci, dbają również o to, aby sieć była zdolna zareagować na nowe okoliczności.
- *Łącznicy* – osobniki, które zostały zrekrutowane do sieci, ale mimo wszystko kontynuują swoje działanie w innych sieciach (w legalnych instytucjach rządowych, politycznych, finansowych). Takie osobniki również mogą wprowadzać dodatkowe informacje i ochronę.

Powyższe role dotyczą jedynie takich osobników, którzy istnieją w sieci dłuższy okres. Pełnią role wymagające od nich zaufania ze strony pozostałych osobników. W tej charakterystyce brakuje jednak pozostałej części osobników, którzy

pojawiają się w sieci sporadycznie lub nie pełnią żadnej ważnej lub zaufanej funkcji. Zatem oprócz wspomnianych powyżej ról, w niniejszej pracy zaproponowano uwzględnienie jeszcze czterech innych ról. Są to:

- *Żołnierze* - osobniki nie pełniące w zasadzie żadnej ważnej funkcji w sieci. Wykonują oni jedynie rozkazy innych osobników w sieci, stojących wyżej w hierarchii. Ich zwierzchnikami mogą być komunikatorzy, strażnicy, monitorujący, łącznicy, a nawet izolatorzy. Jest to dominująca rola w sieci kryminalnej. Tych osobników zwykle jest najwięcej. Mogą się oni zajmować podstawowymi, najmniej zaawansowanymi działaniami grupy.

- *Rekrut* - jest to nowy osobnik w sieci. Może jeszcze nie być uznany przez sieć jako jej członek, ale zaczyna już działać w jej ramach. Nowy osobnik może być rekrutowany na różne pozycje, nie tylko te najniższe (żołnierz). Może się nawet okazać, że nowo rekrutowany osobnik stanie się strażnikiem. Najbardziej charakterystyczną cechą rekruta jest jego krótka obecność w sieci.

- *Postronny* - jest to osobnik, który pomimo faktu należenia do sieci społecznej, nie należy do sieci kryminalnej. Taki osobnik to np. rodzina przestępców - nie musi działać w grupie przestępczej, a mimo tego należy do sieci (z racji częstych kontaktów z jednym z osobników sieci). Postronny z czasem może stać się aktywnym członkiem grupy przestępczej, jednak równie dobrze może nie mieć w ogóle świadomości istnienia grupy przestępczej.

- *Przypadkowy* - osobnik, z którym sieć komunikuje się bardzo rzadko. Jest to osobnik nie należący do sieci. Może się również zdarzyć, że sieć często komunikuje się z takim osobnikiem, a mimo wszystko osobnik nie utożsamia się z siecią. Może to być np. numer pizzerii.

Niniejsza praca skupia się w głównej mierze na wyznaczeniu poszczególnych ról w danej kryminalnej sieci społecznej. W tym celu będą wykorzystywane zarówno charakterystyki związane z budową sieci SNA (np. centralność czy bliskość), jak również atrybuty opisujące samych rozmówców. W tradycyjnej analizie sieci społecznych pomija się cechy osobników nie wynikające bezpośrednio ze struktury grafowej sieci. Mając jednak takie cenne źródło informacji jak charakterystyki danych rozmówców, (długość rozmów, zasięg, ilość rozmów wychodzących i przychodzących) doszliśmy do wniosku, że powinno się z nich skorzystać. Oprócz charakterystyk wynikających z grafowej struktury sieci, oraz tych bezpośrednio opisujących rozmówców wykorzystano trzecie źródło informacji, jakim są sekwencje występowania rozmów telefonicznych pomiędzy osobnikami sieci społecznej. Mając te trzy źródła informacji zaproponowany system stara się rozpoznawać strukturę sieci społecznej i przypisywać poszczególnych osobników do najbardziej pasujących do nich ról.

Dzięki tego typu badaniom, analityk policyjny jest w stanie nie tylko przyjrzeć się strukturze sieci i charakterystykom poszczególnych osobników, ale może zobaczyć jakie role w sieci dane osobniki odgrywają. Jest to niezmiernie istotna wiedza, z punktu widzenia działania zapobiegającego przestępczości zorganizowanej oraz prowadzącego do rozbijania silnie zhierarchizowanych struktur przestępczych.

Oczywistym jest, iż aby najlepiej rozbić strukturę mafijną należy uderzyć w najważniejsze ogniwa, a nie skupiać się na pojedynczych żołnierzach. Uniemożliwiając normalne funkcjonowanie jedynie kilku osobnikom w sieci, stosowne służby są w stanie trwale sparaliżować działanie całej struktury mafijnej.

4. Analiza billingów telefonicznych

W wykorzystywanych w pracy billingach telefonicznych występuje kilka istotnych z punktu widzenia badań pozycji. Są to: data i godzina połączenia, długość rozmowy (w przypadku SMS-a: -1), jaki numer wykonywał połączenie i z jakiego miejsca, z jakim numerem się łączył i gdzie wówczas przebywał dany rozmówca. W celu zbudowania charakterystyk grafowych opisujących rozmówców wykorzystywane są jedynie dwa pola. Są to numery rozmówcy i jego partnera. Na podstawie tych pól wyznacza się kilka rodzajów centralności opisujących rozmówców, które zostały omówione w rozdziale 5.1.

W celu zbudowania charakterystyk opisujących rozmówców zdecydowano się wykorzystać również dodatkowe atrybuty występujące w billingach, a więc mobilność, zasięg i ilość rozmów czy SMS-ów wchodzących i wychodzących oraz długości rozmów, ilość znajomych dzwoniących do danego osobnika oraz ilość tych, do których on dzwonił, stosunek ilości połączeń do SMS-ów wchodzących i wychodzących oraz długość przebywania w sieci.

Ostatnim, trzecim, ogniwem systemu wspomagającego pracę analityka policyjnego jest moduł operujący na sekwencjach występujących w rozmowach telefonicznych. Chodzi tutaj o zaobserwowanie takich wzorców zachowań jak np. fakt, iż zwykle organizatorzy występują na początku lub na końcu sekwencji, a następnie powinni po nich występować izolatorzy. Powinni oni również poprzedzać organizatorów, gdy występują oni na końcu sekwencji.

Scalając te trzy rodzaje informacji uzyskane z wykazów rozmów telefonicznych system stara się przewidywać role, jakie mogą pełnić poszczególne osobniki w sieci.

5. Zastosowane algorytmy

W celu identyfikacji powiązań i wzorców zachowań w sieci kryminalnej potrzebne było zastosowanie szeregu algorytmów obliczających podstawowe parametry na poziomie sieci społecznej oraz na poziomie danych z transakcyjnej bazy billingów telefonicznych. Algorytmy te podzielono na 4 grupy. Do pierwszej z nich należą metody obliczania centralności osobników w sieci społecznej. Druga grupa algorytmów to obliczanie parametrów na poziomie bazy billingów, takich jak średnia długość rozmowy, mobilność obliczana na podstawie współrzędnych stacji bazowych GSM, czy też średnia ilość połączeń w rozrachunku dziennym. Trzecia grupa dotyczy dzielenia billingów na sekwencje zdarzeń oraz wyszukiwania najczęstszych wzorców w tych sekwencjach. Ostatnia, czwarta opisuje połączenie rezultatów wszystkich poprzednich algorytmów i zastosowanie ich jako wyznacznik do stwierdzenia przynależności osobnika do odpowiedniej roli w sieci kryminalnej.

5.1. Centralność w sieci społecznej

Centralność (ang. network centrality) jest atrybutem każdego obiektu, opisującym jego pozycję w strukturze sieci. Stanowi miarę ważności, znaczenia i wpływu na inne obiekty w tej samej sieci społecznej (White (2003)).
Stworzony system KASS umożliwia wyznaczanie 8 rodzajów centralności w zagregowanej reprezentacji sieci, zbudowanej na podstawie billingów. Poniższe parametry obliczane są na poziomie biblioteki JUNG (Java Universal Network/Graph Framework) (O'Madadhain (2005)).

5.1.1. Środek ciężkości

Środek ciężkości (ang. bary center) to miara opisująca dany węzeł, obliczana na podstawie wszystkich najkrótszych ścieżek prowadzących do tego węzła. W sieci kryminalnej, w której połączenia między osobami stanowią billingi telefoniczne, współczynnik ten interpretuje się w następujący sposób: jeżeli rozmówca ma wysoką wartość Bary Center, to leży on w centrum sieci kryminalnej, czyli byłby w stanie szybko przesłać (lub odebrać) informacje od dowolnych członków organizacji. Osoba z największym środkiem ciężkości może w najszybszy sposób uzyskać informację zgromadzoną we wszystkich węzłach sieci.

5.1.2. Centralność „Betweenness"

Jest to miara obliczana na podstawie wszystkich najkrótszych ścieżek do wszystkich węzłów. Z punktu widzenia sieci kryminalnej, rozmówca o wysokiej wartości „betweenness" jest postrzegany jako broker przekazujący informacje. Bywa on też punktem krytycznym dla sieci, gdyż w momencie jego usunięcia, uzyskanie poprzedniej jakości przepływu informacji w siatce kryminalnej byłoby utrudnione.

5.1.3. Stopnie wierzchołków wchodzących i wychodzących

Stopnie wierzchołków wchodzących (ang. degree distribution in) oraz wychodzących (ang. degree distribution out) są to miary obliczane na podstawie ilości krawędzi wchodzących do danego węzła oraz z niego wychodzących. W sieci kryminalnej parametr dla danego rozmówcy jest tym większy, im więcej miał on rozmów przychodzących lub wychodzących.

5.1.4. Stopień koncentracji i autorytatywność

Stopien koncentracji (ang. hubness) i autorytatywność (ang. authoritativeness) są to miary obliczane na podstawie sposobu w jaki węzły wskazują na siebie. W sieci kryminalnej rozmówca, który jest hubem (węzłem o największym parametrze stopnia koncentracji) posiada dużą liczbę rozmów wychodzących do partnerów - autorytetów (węzłów o największej wartości parametru autorytatywności), którzy stanowią punkt centralny i odbierają telefony i sms'y od dużej ilości innych hubów.

Oczywiście nie zawsze odbieranie rozmów świadczy o dużej ilości posiadanej informacji - może się zdarzyć, że to właśnie hub dzieli się swoja informacja z autorytetem. Hub może zarówno wydawać polecenia jak i prosić o dyrektywy, przekazywać informacje jak i je otrzymywać.

5.1.5. Page Rank

Parametr Page Rank w przypadku sieci kryminalnej może być interpretowany jako wyznacznik posiadanej informacji. Można założyć, że jeśli dana osoba dzwoni do innej, to albo chce się czegoś dowiedzieć, albo ją o czymś poinformować. Jeżeli takich połączeń jest znacznie więcej, to można przyjąć, że taka osoba posiada dużo więcej informacji, niż osoba do niej dzwoniąca i przy tym posiadająca mniej kontaktów. Należy również zwrócić uwagę na przypadek, w którym osoba z niewielką liczbą kontaktów dzwoni do osoby z dużą liczbą kontaktów. Może ona wtedy potencjalnie zdobyć więcej informacji, niż dzwoniąc do kilku osób bez kontaktów, lub z małą ich liczbą.

5.1.6. Centralność Markowa

W analizie sieci społecznych parametr ten jest interpretowany następująco: skoro najczęściej odwiedzanym punktem jest punkt, do którego można dojść największą liczbą możliwych ścieżek, to możemy również przyjąć, że taki węzeł w sieci jest w stanie zgromadzić najwięcej informacji.

5.2. Parametry billingów telefonicznych

Na podstawie innych pól dostępnych w billingach budowane są dodatkowe parametry. I tak współrzędne stacji bazowej służą do wyliczenia zasięgu rozmów przychodzących i wychodzących, jak również do sprawdzenia, czy rozmówca wykonując rozmowy przemieszczał się, czy raczej był statyczny. Na podstawie pozostałych pól można wyliczyć m. in. średnie długości rozmów i ilości sms'ów oraz ilości rozmówców, jak również określić, od jakiego czasu dany rozmówca pojawia się w sieci. Dodatkowe parametry reprezentowane są również przez procentowe odchylenie standardowe od średniej. Parametry te będą służyły do końcowego dostrajania aplikacji i dodawania dodatkowych punktów za małe odchylenie standardowe, co można interpretować jako nagradzanie za stałość zachowań.

5.3. Wyszukiwanie częstych sekwencji

Wyszukiwanie częstych sekwencji składa się z dwóch etapów. W pierwszym z nich billingi rozdzielane są na sekwencje rozmów. W drugim etapie wyszukiwane są najczęściej powtarzające się wzorce w wyznaczonych sekwencjach za pomocą algorytmu PrefixSpan (Pei (2004)).

Algorytm wydzielania sekwencji tworzy grupy połączeń pomiędzy rozmówcami i nadaje im oddzielny identyfikator. Operuje on na wszystkich billingach wczytanych do bazy danych i stara się powiązać w sekwencje poszczególne rozmowy.

Na początku podawany jest parametr określający długość okna czasowego pomiędzy rozmowami. Algorytm biorąc pierwszą rozmowę stara się odnaleźć kolejną, taką która nastąpiła nie później niż przyjęte okno czasowe od poprzedniej. Ponadto w następnej rozmowie musi uczestniczyć numer występujący w poprzedniej. W ten sposób powstają sekwencje następujących po sobie rozmów. Może to być sekwencja o charakterze kolejki tzn. A dzwonił do B, B dzwonił do C, C dzwoni do D, jak również struktura drzewiasta lub grafowa: A dzwonił do B, A dzwonił do C, A dzwonił do D, C dzwonił do E, B dzwoni do D.

Istotny jest jedynie fakt, aby czas pomiędzy zakończeniem się poprzedzającej rozmowy a rozpoczęciem bieżącej nie był dłuższy niż przyjęte okno czasowe.

Każda wydzielona sekwencja jest wektorem dwójek (rozmówca, numer powiązany). Wektor ten jest posortowany według dat nawiązywania połączenia (zapewnia to algorytm wydzielania sekwencji).

Algorytm PrefixSpan przeszukując cały zbiór sekwencji generuje zbiór wzorców i do każdego przypisuje jego częstość występowania (ang. support), czyli ilość sekwencji, w których on występuje. Parametrami wejściowymi do algorytmu są minimalna częstość występowania oraz minimalna i maksymalna długość wzorca.

Wyniki częstości pozwalają na zastosowanie pewnych reguł dotyczących komunikacji pomiędzy osobnikami, spełniającymi określone role. Przykładem może być następująca reguła: rozkaz od organizatora trafia najpierw do izolatora, a potem rozsyłany jest do innych osobników w sieci. A więc najczęściej organizator jest inicjatorem sekwencji, a izolator jest w sekwencji tuż za nim.

5.4. Identyfikacja ról

Algorytm wyznaczania ról w sieci kryminalnej polega na przyznawaniu punktów dla poszczególnych osobników za spełnienie określonych kryteriów. Przyznawanie punktów składa się z trzech głównych etapów. Pierwszy, to ocena węzłów w sieci na podstawie charakterystyk związanych z SNA, drugi to przyznawanie punktów ze względu na charakter zachowań danego osobnika, trzeci - za specyficzne zachowania wynikające z częstych sekwencji i występowania określonych ustawień osobników w tych sekwencjach.

6. Opis zrealizowanego systemu

System KASS składa się z trzech części:

1. Pakietów integracji danych (SSIS);
2. Relacyjnej bazy danych z billingami telefonicznymi wraz z procedurami składowanymi;
3. Aplikacji z graficznym interfejsem użytkownika do analizy i wizualizacji danych.

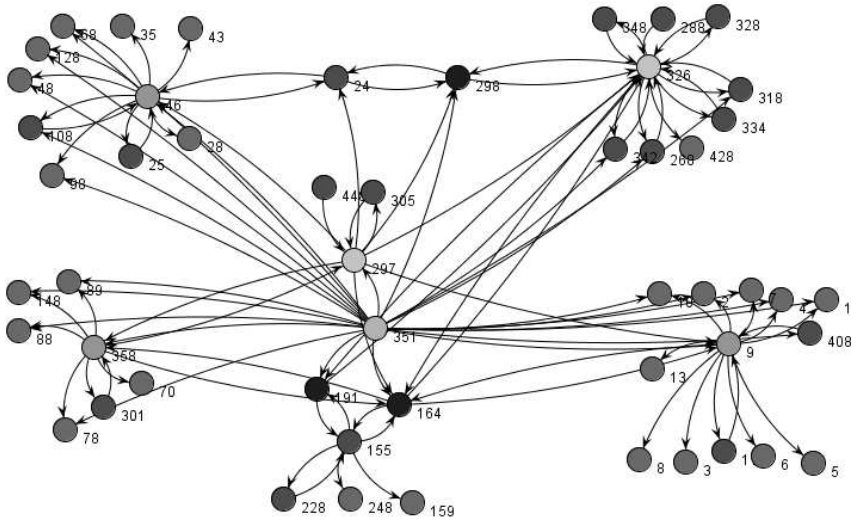
System wczytuje dwa rodzaje plików wejściowych. Pierwszy z nich to plik billingów w formacie CSV, drugi zaś to plik stacji bazowych GSM w formacie XLS.

Pliki te wczytywane są poprzez pakiety SSIS, następnie dane z tych plików, odpowiednio sformatowane, trafiają do tabel schematu bazy billingów. Jest to pierwszy etap wstępnego przetwarzania, który ma na celu wyczyszczenie danych i przygotowanie ich do następnego poziomu przetwarzania.
Drugi etap wstępnego przetwarzania, to uruchomienie procedur składowanych języka SQL, które zasilają tabele aplikacji (nazwano je tabelami EXT). Procedury na tym etapie identyfikują rozłączne sieci, agregują dane oraz wyliczają charakterystyki opisane w punkcie 5.2.
Trzeci etap, to analiza sieci kryminalnej w aplikacji z graficznym interfejsem użytkownika. Zapewniona jest tu pełna wizualizacja sieci w postaci graficznej oraz w postaci danych tabelarycznych. Z poziomu aplikacji wyznaczane są parametry SNA oraz uruchamiany jest szereg procedur składowanych SQL, takich jak dzielenie billingów na sekwencje, wyliczanie ról.

7. Opis przeprowadzonych testów

Przeprowadzone eksperymenty miały na celu sprawdzenie poprawności działania systemu, głównie w kwestii poprawnego typowania ról. Jako błędne wytypowanie ról przyjmuje się ewidentnie widoczne w strukturze sieci nieprawidłowe przypisanie roli do osobnika, np. żołnierz będący w centrum sieci.

Po zastosowaniu podstawowych mechanizmów systemu KASS do wyznaczania ról kryminalnych w sieci społecznej otrzymano przedstawiony poniżej wynik (rysunek 1). Analizę sieci przeprowadzono wyłącznie w oparciu o charakterystyki SNA.



Rysunek 1. Testowana mała sieć kryminalna

Otrzymane wyniki (biorąc pod uwagę topologię sieci) wydają się być właściwe. W związku, z tym iż analizowana sieć nie była duża, otrzymano jednego organiza-

tora (oznaczonego numerem 155), oddzielonego od reszty sieci przez dwóch izolatorów (oznaczonych numerami 191 i 164). Otrzymano również trzeciego izolatora (298). Jego rola nie jest do końca poprawnie określona, gdyż można by przypisać mu rolę łącznika. Łącznik ten łączyłby dwie podsieci skupione wokół osobników (46) i (326).

W przedstawionej sieci jest możliwe zidentyfikowanie pięciu podsieci. Pierwsza skupia się wokół organizatora, który ma swoich lokalnych współpracowników (jeden żołnierz i dwóch izolatorów) oraz dwie osoby postronne - może to być np. jego rodzina nie zaangażowana w działalność przestępczą. Pozostałe cztery podsieci to powinni być komunikatorzy otoczeni podporządkowanymi sobie żołnierzami i osobami postronnymi. Tak też w większości przypadków jest. Wyjątkiem jest osobnik (326), który został błędnie rozpoznany jako monitorujący. Stało się tak dlatego, iż ma on niespotykanie wysoki betweenness (rysunek 2) oraz dużą liczbę połączeń, co jest cechą właśnie monitorującego. Ma on również stosunkowo niski autorytet, co również jest cechą monitorującego (przede wszystkim ze względu na dominację połączeń wychodzących - zwykle wzrasta hubness i spada authoritativeness).

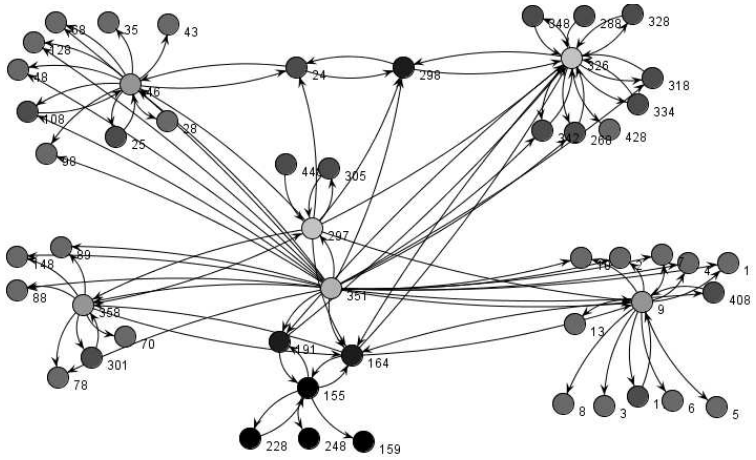
Kolejną rolą jest strażnik (297). Osobników tego typu w sieci nie powinno być dużo. W tej pracy przyjmuje się, iż w sieci strażników powinno być do 5% całej populacji. Wynika to z tego, iż jest to osoba zaufana i bardzo respektowana w sieci. Pełni ona specyficzne (nie wymagające wielu osobników) funkcje opisane już wcześniej w tym artykule. System odpowiednio rozpoznał strażnika, gdyż taki osobnik charakteryzuje się wieloma połączeniami wychodzącymi głównie do komunikatorów, żołnierzy, postronnych. Również może się komunikować z osobnikami o innych rolach. Rolą podobną do strażnika jest monitorujący, z tą różnicą, że nie komunikuje się on z żołnierzami, czy postronnymi, ale stara się monitorować całą sieć poprzez komunikowanie się z ważniejszymi rolami w sieci, takimi jak komunikatorzy, łącznicy, izolatorzy. Obie te role zostały odkryte przez system KASS. Jedynie niepotrzebnie został wyznaczony drugi monitorujący (326). Błąd ten został już wyżej omówiony. Właściwy monitorujący (297) jest również lokalnie powiązany z dwoma żołnierzami.

Pozostałe wyznaczone role to żołnierze i postronni. Są to elementy w gruncie rzeczy podobne, niewiele się różniące pod względem topologii sieci. Na rysunku 2 przedstawiono zestawienie poszczególnych parametrów dla bardziej znaczących osobników opisanych powyżej.

ID BD	BaryC	BetweennessC	DegreeDistIn	DegreeDistOut	Hubness	Authoritativeness	PageRank	MarkovC
46	0,81557	0,47715	0,45455	0,40000	0,22231	0,63373	0,20870	0,20290
24	0,70492	0,33950	0,27273	0,08000	0,09324	0,23464	0,14165	0,15119
298	0,60246	0,39725	0,36364	0,08000	0,03267	0,73172	0,19172	0,22616
326	0,50820	1,00000	1,00000	0,40000	0,31361	0,24384	1,00000	1,00000
297	0,35656	0,36297	0,36364	0,32000	0,28875	0,58449	0,10242	0,09064
351	0,31967	0,00000	0,00000	1,00000	1,00000	0,00000	0,00000	0,00000
191	0,65984	0,06265	0,27273	0,08000	0,02051	0,61922	0,13847	0,19919
164	0,44672	0,87875	0,54545	0,16000	0,10656	1,00000	0,18614	0,26959
155	0,59836	0,21328	0,27273	0,20000	0,12128	0,05650	0,20398	0,21223
358	0,47131	0,46953	0,36364	0,32000	0,27461	0,62042	0,11819	0,12668
9	0,54098	0,55332	0,45455	0,48000	0,31902	0,63984	0,15094	0,16813

Rysunek 2. Parametry SNA

W dalszej części testów próbowano zaobserwować zachowanie się sieci po usunięciu wybranego osobnika symulując zmianę roli tego osobnika na przypadkowego. Jest to osobnik nie związany z siecią, zatem doskonale symuluje on oderwanie się od sieci osobnika ważnego. Wyznaczony wcześniej organizator został ustalony jako przypadkowy - od tego momentu również połączone z nim osobniki także stają się przypadkowe. Wykorzystując system KASS możliwe jest sprawdzenie jak ta rola zostanie w sieci przeorganizowana. Wyniki przedstawiono na rysunku 3.

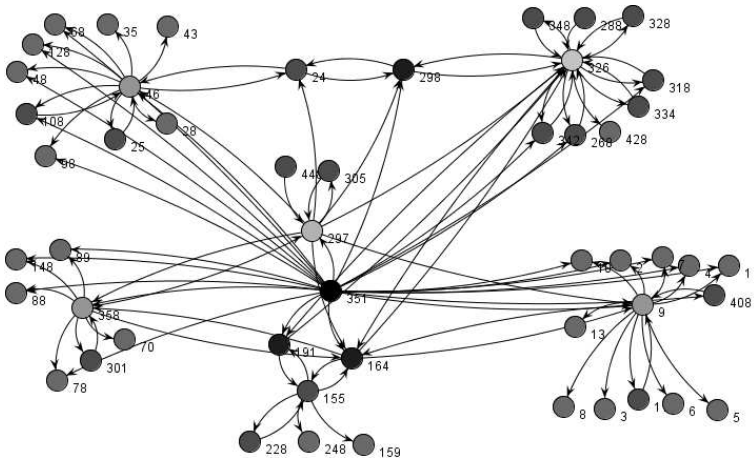


Rysunek 3. Migracja roli organizatora

Organizatorem został inny osobnik (24). Można również zaobserwować iż umiejscowił się on w pobliżu „niechcianego" wcześniej izolatora. Jednak w momencie zmiany roli organizatora takie ułożenie zaczęło mieć sens z punktu widzenia topologii sieci i zachowania się poszczególnych osobników realizujących określone role. Widać tutaj iż poprzednio niechciany izolator słusznie został wyznaczony jako pretendent do bycia izolatorem.

Kolejnym eksperymentem było wyłączenie osobnika będącego strażnikiem(351). Nowym strażnikiem stał się osobnik poprzednio będący monitorującym(297). Jest to sytuacja bardzo możliwa, gdyż oba te osobniki mają bardzo podobny profil. Zachował on wiele z połączeń występujących wcześniej u oryginalnego strażnika. Dzięki temu posiada on część możliwości strażnika. Sytuacja ta została pokazana na rysunku 4.

System na domyślnych ustawieniach popełnia niewielkie błędy. Jako niewielkie błędy przyjmuje się błędne wytypowanie ról dla maksymalnie 10% populacji. Sterowanie parametrami (w KASS ponad 40 dla każdej roli) można realizować poprzez modyfikacje współczynników w matrycy opisanej w 5.4 wprowadzając większe rozróżnienie w nagradzaniu osobników za posiadane cechy pretendujące do określonej roli.



Rysunek 4. Migracja roli strażnika

8. Podsumowanie

System KASS umożliwia automatyczne wykorzystywanie wszystkich wyliczanych atrybutów do odkrywania ról, jakie mogą pełnić w sieci poszczególne osobniki. Możliwe staje się sparaliżowanie działalności całej grupy przestępczej poprzez wyłączenie z jej szeregów pojedynczych osobników. Niekoniecznie muszą to być organizatorzy, którzy zwykle bardzo dobrze dbają o swoje alibi. Równie wielkie spustoszenie w sieci poczyni wyłączenie z niej strażnika, komunikatora, monitorującego czy łącznika. Aplikacja po odpowiednim dostrojeniu dopasowanym do charakterystyki badanej grupy przestępczej, może dostarczać bardzo cennych informacji służących do zwalczania zorganizowanych grup przestępczych. Dalsze prace badawcze idą w kierunku uwzględnienia odkrytych sekwencji w algorytmie wyliczającym role.

Literatura

Arquilla, J. and Ronfeldt, D. (2002) *Networks and Netwars : The Future of Terror, Crime, and Militancy (Consumer One-Off)*. RAND Corporation.

Dombroski, M. J. and Carley, K. M (2002) *Estimating a Terrorist networks Structure*. Proc. of CASOS 2002 Conference, Computational and Mathematical Organization Theory, 235-241.

Jensen, D. and Neville, J. (2002) *Data Mining in Social Networks*. Proceedings of National Academy of Sciences Symposium on Dynamic Social Network Analysis.

O'Madadhain, J., Fisher, D.,Smyth, P.,White, S. and Yan-Biao, B. (2005) *Analysis and Visualization of Network Data using JUNG*. Journal of Statistical Software.

Pei, J. and Han, J. and Mortazavi-Asl, B. and Wang, J. and Pinto, (2004) *Mining Sequential Patterns by Pattern-Growth: The PrefixSpan Approach.* IEEE Trans. Knowl. Data Eng. **16**, 11, 1424-1440.

Xu, J. and Marshall, B. and Kaza, S. and Chen, H. (2004) *Analyzing and Visualizing Criminal Network Dynamics: A Case Study.* IEEEConference on Intelligence and Security Informatics (ISI), Tucson, AZ.

Xu, J.J and Chen, H. (2005a) *CrimeNet Explorer: a Framework for Criminal Network Knowledge Discovery.* ACM Trans. Inf. Syst. **23**, 2, 201-226.

Wellman, B. (1999) *From Little Boxes To Loosely Bounded Networks: The Privatization and Domestication of Community.*

White, S. and Smyth, P. (2003) *Algorithms for estimating relative importance in networks.* ACM Press. KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining., 266–275.

Wolfe, A.P. and Jensen, D. (2004) *Playing Multiple Roles: Discovering Overlapping Roles in Social Networks.* Proceedings of the ICML-04 Workshop on Statistical Relational Learning and its Connection to Other Fields.

Environment for identifying behavioural patterns using social networks

The article describes how social network methods and frequent sequences analysis are used to identify behavioural patterns based on a database of phone calls made. We developed algorithms for specifying roles as well as carrying out verification tests. The roles will be determined using the rules containing standard parameters to specify the social network structure (for example: betweeness centrality, hubness, authoritativeness, page rank or Markov centrality) and some specific parameters characterising the network of telephone calls: for example, time of telephone conversation, mobility of interlocutors, distance of calls, daily average of incoming and outgoing calls and SMSes sent.