

Rozbijanie sieci terrorystycznych za pomocą teorii gier

Oskar SKIBSKI*

*Uniwersytet Warszawski, Instytut Informatyki, Wydział Matematyki, Informatyki i Mechaniki

Mimo licznych działań skierowanych na zwalczanie terroryzmu wiele organizacji terrorystycznych wciąż się powiększa. Aby poradzić sobie z tym problemem, agencje bezpieczeństwa poszukują nowych sposobów analizy pozwalających lepiej zrozumieć strukturę tych organizacji. Jednym z problemów jest zidentyfikowanie kluczowych członków organizacji terrorystycznej przy użyciu informacji jedynie o tym, jak wygląda sieć terrorystyczna – dzięki temu agencje bezpieczeństwa mogłyby skupić swoje ograniczone zasoby na tych jednostkach. W tym artykule omówimy nowe podejście do tego problemu oparte na teorii gier. Zaczniemy jednak od przeglądu standardowych metod, które niedawno wprowadzone zostały do IBM i2 *Analyst's Notebook* – oprogramowania używanego przez organy ścigania i agencje wywiadowcze na całym świecie.

Podstawowe miary centralności

Wyobraź sobie, Czytelniku, że pracujesz w Centrum Antyterrorystycznym ABW, a na Twojej korkowej tablicy znajdują się zdjęcia 19 terrorystów połączonych sznurkami tak jak na rysunku. Każdy sznurek określa połączenie pomiędzy dwoma terrorystami – takie połączenie może oznaczać kontakt (np. rozmowę telefoniczną) między nimi, o którym wiedzą służby, lub więzi rodzinne. Twoim zadaniem jest wybranie terrorysty, któremu agencja założy podsluch. Którego terrorystę wskażesz?

Trochę bardziej formalnie – nasza sieć to (nieskierowany) graf, czyli para (V, E) , w której V oznacza zbiór wierzchołków (terrorystów), a E to nasze połączenia (sznurki), czyli zbiór krawędzi. Krawędź pomiędzy wierzchołkami v i u oznaczamy przez $\{v, u\}$ (jest to zatem nieuporządkowana para). Naszym celem jest określenie, jak ważny jest każdy wierzchołek, czyli przypisanie każdemu v pewnej liczby rzeczywistej $F(v)$. Takie przypisanie, czyli formalnie funkcję, nazwiemy *miarą centralności*. Miary centralności mają szereg zastosowań od tak odległych jak wyznaczanie najbardziej wpływowych osób w sieciach społecznych, przez kluczowe węzły w infrastrukturze drogowej czy informatycznej, aż po analizę istotności genów w sieciach biologicznych.

Wróćmy jednak do naszego przykładu. Jak wybrać terrorystę, którego będziemy podsłuchiwać? Jeżeli zależałoby nam po prostu na zmaksymalizowaniu liczby podsłuchanych osób, powinniśmy wybrać osobę z największą liczbą połączeń. Miara oceniająca wierzchołek ze względu na ilość jego krawędzi nazywana jest miarą stopnia wierzchołka (*degree centrality*):

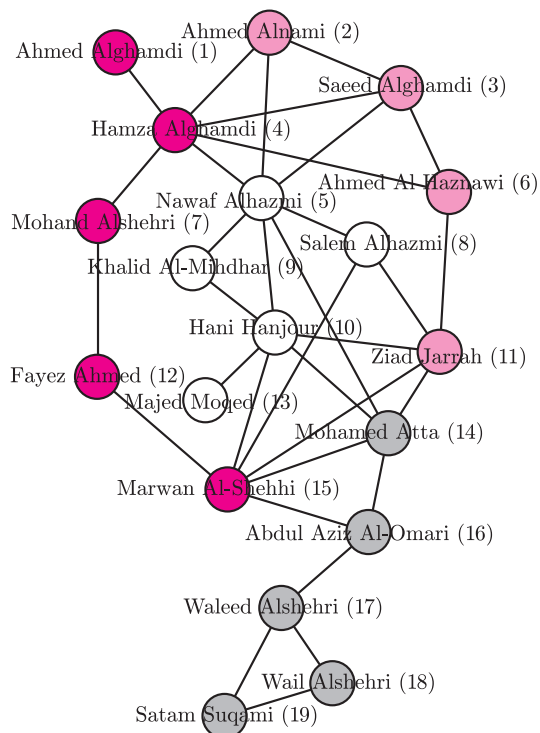
$$D(v) = |\{u \in V : \{v, u\} \in E\}| \quad (\text{miara stopnia}).$$

Na rysunku jest to Nawaf Alhazmi (5), który ma połączenie z aż siedmioma innymi osobami. Zwróćmy jednak uwagę, że wybierając tego terrorystę, rzeczywiście nagramy dużo osób, ale uzyskane informacje mogą mało znaczyć. Prezydent ma bezpośredni kontakt z mniejszą liczbą osób niż sprzedawca w McDonalddie.

Zastanówmy się teraz, jak przechwycić losową informację krążącą po sieci terrorystycznej. Najlepsza pod tym względem jest osoba, która jest w centralnym punkcie sieci. Jak bardzo dany wierzchołek jest w centrum, możemy zmierzyć, patrząc na to, jak daleko jest od innych wierzchołków. Odległość wierzchołka v od u definiujemy w grafie, mierząc najkrótszą ścieżkę między nimi, czyli licząc, ilu krawędzi musimy użyć, aby dotrzeć z v do u . Miara oparta na tej idei nazywa się miarą bliskości (*closeness centrality*):

$$C(v) = \frac{1}{\sum_{u \in V \setminus \{v\}} \text{dist}(v, u)} \quad (\text{miara bliskości}),$$

gdzie $\text{dist}(v, u)$ to odległość wierzchołka v od u . Powyżej wzięliśmy odwrotność sumy odległości, a zatem wierzchołek, który jest blisko innych, będzie miał tę wartość wysoką, a ten położony na obrzeżach – niską. Miara bliskości zastosowana do naszej sieci z rysunku po raz kolejny wskazuje Nawafa Alhazmiego (5), ale tym razem jest on *ex aequo* z Mohamedem Attą (14) (obaj uzyskali wartość $1/35$).



Sieć terrorystów bezpośrednio odpowiedzialnych za ataki na World Trade Center w 2001 roku. Kolorami wyróżnione są cztery samoloty, którymi lecieli zamachowcy-samobójcy.

Tak naprawdę szukamy ogólnej metody określania ważności wierzchołka w sieci (sieci terrorystycznych jest przecież bardzo dużo!). *Miara centralności* to zatem funkcja, która dla każdego grafu daje wektor wag:

$$F : \mathcal{G}^V \rightarrow \mathbb{R}^V,$$

gdzie \mathcal{G}^V to zbiór wszystkich grafów z wierzchołkami V , a \mathbb{R} to zbiór liczb rzeczywistych.

Formalnie ścieżka to ciąg wierzchołków $p = (v_0, v_1, \dots, v_n)$, w którym każde dwa kolejne wierzchołki są połączone krawędzią, czyli $\{v_i, v_{i+1}\} \in E$ dla każdego $i \in \{0, \dots, n-1\}$.

Warto zwrócić uwagę na to, że sposób, w jaki narysowany jest graf, nie wpływa na odległości między wierzchołkami i może mocno mylić przy szukaniu centralnego wierzchołka „na oko”.



Mówimy, że graf jest *spójny*, kiedy z każdego wierzchołka do każdego innego istnieje ścieżka. *Wierzchołek rozcinający* czy inaczej *punkt artykulacji* to wierzchołek, którego usunięcie powoduje, że graf przestaje być spójny.

Powyższy sposób jest skuteczny, jeżeli informacje krążą losowo. Zastanówmy się jednak, kto najczęściej będzie na szlaku informacji. Załóżmy, że wierzchołek s chce przekazać pewną informację wierzchołkowi t . Najlepiej, żeby informacja dostała się do t jak najszybciej i jak najbezpieczniej (lepiej nie wtajemniczać osób, których nie trzeba), w związku z tym założymy, że porusza się ona po najkrótszej ścieżce. Najkrótszych ścieżek może być jednak kilka. Np. na rysunku Mohamed Atta (14) może dostarczyć informację do Salema Alhazmiego (8) przez Ziada Jarraha (11) albo przez Marwana Al-Shehiego (15). Możemy jednak powiedzieć, że skoro Ziad Jarrah (11) należy do połowy najkrótszych ścieżek pomiędzy (14) i (8), to przechwyci połowę informacji wysyłanych między nimi. Rozważając wszystkie pary, które mogą wysyłać do siebie informacje i dodając te liczby, otrzymujemy miarę pośrednictwa (po angielsku *betweenness centrality*):

$$B(v) = \sum_{s,t \in V \setminus \{v\}} \frac{|\{p \in \Pi(s,t) : v \in p\}|}{|\Pi(s,t)|} \quad (\text{miara bycia pomiędzy}),$$

gdzie $\Pi(s, t)$ oznacza zbiór najkrótszych ścieżek pomiędzy s a t . A zatem w mianowniku mamy liczbę wszystkich najkrótszych ścieżek pomiędzy s a t , a w liczniku – liczbę takich ścieżek z wierzchołkiem v . Obliczenie miary pośrednictwa dla naszego grafu wymaga już trochę pracy. Kiedy to zrobimy, po raz kolejny okaże się, że wygrał Nawaf Alhazmi (5). Na drugim miejscu znajduje się jednak Abdul Aziz Al-Omari (16), który klasyfikowany jest nisko według innych miar – nie ma ani dużo połączeń, ani nie jest w centrum sieci. Czemu znalazł się on tak wysoko?

Wierzchołek ten jest wierzchołkiem rozcinającym, czyli ma taką wyjątkową cechę, że jego usunięcie powoduje, iż graf przestaje być połączony (spójny) i sieć traci możliwość koordynacji. Analitycy sieci terrorystycznych wskazują, że osoby takie odgrywają kluczową rolę w sieci, będąc na złączeniu wielu grup potrzebnych do przeprowadzenia ataku. Wylimitowanie takich osób może najbardziej przyczynić się do rozbitcia sieci terrorystycznej.

Miara łączenia grafu

Oprócz Abdula Aziza Al-Omariego (15) istnieją tylko trzy inne wierzchołki rozcinające: są to Hamza Alghamdi (4), Hani Hanjour (10) oraz Waleed Alshehri (17). Jeśli przyjmiemy to kryterium, reszta wierzchołków pozostaje nierozpoznawalna. W gęstszych grafach, tzn. grafach z większą ilością krawędzi przy tej samej ilości wierzchołków, często nie ma żadnego rozcinającego wierzchołka, tzn. usunięcie żadnego pojedynczego wierzchołka nie prowadzi do rozspójnienia grafu. Jak wówczas znaleźć wierzchołek najważniejszy w utrzymywaniu jego spójności?

W tym celu zamiast rozpatrywać pojedyncze wierzchołki musimy patrzeć na całe ich grupy. Grupy będziemy oceniać według tego, czy potrafią komunikować się bez pozostałych wierzchołków, czy nie, tzn. czy po usunięciu pozostałych wierzchołków z grafu będzie on spójny. W najprostszej wersji moglibyśmy przypisać grupie wartość 1, jeżeli jest ona spójna i 0 w przeciwnym przypadku. My postąpimy trochę mądrzej i weźmiemy pod uwagę, jak bardzo niespójna jest grupa: wartość 0 dostanie, jeżeli wszyscy jej członkowie są osobno, 1 – jeżeli dokładnie jedna para jest połączona itd. W rezultacie spójna grupa k wierzchołków otrzyma wartość $k - 1$. Bardziej formalnie – wartość grupy wierzchołków $S \subseteq V$ określamy równaniem

$$f^*(S) = |S| - |K(S)|,$$

gdzie $|K(S)|$ to liczba jej komponentów, czyli spójnych części. Przykładowo na rysunku trzysobowa grupa osób o nazwisku Alshehri (7, 17, 18) otrzyma wartość 1, a równie liczna rodzina Alghamdich (1, 3, 4) wartość 2. Jak wyznaczyć jednak teraz, kto najbardziej przykłada się do łączenia grafu? Z pomocą przychodzi nam teoria gier, a konkretniej gry koalicyjne. Jak się zaraz okaże, powyższy problem oceny ważności w grafie zamieniliśmy na problem oceny ważności graczy w grze koalicyjnej...

Gry koalicyjne na ratunek!

Gra koalicyjna to para (N, f) , gdzie N to zbiór graczy, a f to tak zwana funkcja charakterystyczna, która każdej niepustej grupie graczy przypisuje pewną liczbę rzeczywistą, będącą ich wypłatą (formalnie $f : 2^N \rightarrow \mathbb{R}$ przy czym zbiór pusty ma wartość 0: $f(\emptyset) = 0$). W grze koalicyjnej grupy nazywamy koalicjami. Gry koalicyjne



Rozwiązanie zadania F 915.

W obu przypadkach możemy posłużyć się analizą wymiarową. W przypadku a) zakładamy zależność postaci:

$$c \propto \sigma^x \rho^y \lambda^z.$$

Wymiarem prędkości c jest m/s, napięcia powierzchniowego $\sigma - \text{N/m} = \text{kg/s}^2$, gęstości $\rho - \text{kg/m}^3$, a długości fali $\lambda - \text{m}$. Po porównaniu wymiarów lewej i prawej strony otrzymujemy: $x = 1/2$, $y = -1/2$, $z = -1/2$, czyli:

$$c^2 = A \frac{\sigma}{\rho \lambda},$$

gdzie A jest pewną bezwymiarową stałą – ściśle rozwiązanie zagadnienia dla fal kapilarnych prowadzi do wartości $A = 2\pi$. Znalezione zależności poprawnie opisuje zachowanie bardzo krótkich fal ($\lambda < 1/3 \text{ cm}$) na powierzchni wody. W przypadku b) zakładamy:

$$c \propto \rho^x g^y \lambda^z$$

Analiza wymiarowa prowadzi do wartości: $x = 0$, $y = 1/2$, $z = 1/2$, czyli:

$$c^2 = B g \lambda,$$

gdzie B jest bezwymiarową stałą – ściśle rozwiązanie zagadnienia dla fal grawitacyjnych prowadzi do $B = (2\pi)^{-1}$. Znalezione zależności poprawnie opisuje fale na wodzie, gdy $\lambda > 8 \text{ cm}$.

używane są w ekonomii np. do modelowania rynków, a w informatyce stanowią popularny model kooperacji agentów w systemach wieloagentowych.

Podstawowym problemem w grze koalicyjnej jest następujące pytanie. Załóżmy, że wszyscy gracze będą razem współpracować i utworzą koalicję N . Jak teraz powinni podzielić się wspólną wypłatą? Przykładowo, niech firma A sprzedaje lody po 4 złote, a firma B sprzedaje patyki po 2 złote. Jeżeli firmy te się połączą, będą sprzedawać lody na patyku za 12 złotych. Jak powinny podzielić się tymi 12 złotymi?

Opisana gra to zatem para $(\{A, B\}, f)$, gdzie funkcja f jest określona następująco: $f(\{A\}) = 4$, $f(\{B\}) = 2$ i $f(\{A, B\}) = 12$.

Najsłynniejszą metodę rozwiązania tego problemu podał Lloyd Shapley, amerykański uczony i laureat Nagrody Nobla z ekonomii. Shapley zaproponował, aby oceniać gracza, patrząc na to, jaki jest jego wkład do różnych koalicji. Tak zwany *wkład marginalny* gracza i do koalicji S definiujemy jako różnicę pomiędzy wartością koalicji S z graczem i i bez niego: $f(S \cup \{i\}) - f(S)$. W oparciu o to pojęcie Shapley zaproponował następującą procedurę obliczania uczciwego udziału gracza i we wspólnej wypłacie. Ustalmy pewną permutację ω , czyli kolejność, w jakiej gracze dołączają do gry. Każdy kolejny gracz zwiększa wartość już istniejących graczy o swój wkład marginalny. W szczególności jeżeli przez S_i^ω oznaczymy zbiór graczy, którzy występują w permutacji ω przed graczem i , gracz i wnosi wartość $f(S_i^\omega \cup \{i\}) - f(S_i^\omega)$. Teraz uczciwa wypłata gracza i to średni wkład marginalny po wszystkich możliwych permutacjach:

$$SV_i(N, f) = \frac{1}{|N|!} \sum_{\omega \in \Omega(N)} f(S_i^\omega \cup \{i\}) - f(S_i^\omega) \quad (\text{wartość Shapleya})$$

Czytelnik Wnikliwy zauważy, że wielokrotnie będziemy rozpatrywać wkład gracza i do takiej samej koalicji S przy różnych permutacjach. Bazując na tej obserwacji, wartość Shapleya można zapisać następująco:

$$\sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!} (f(S \cup \{i\}) - f(S)).$$

Uzasadnienie tej formuły pozostawiamy Czytelnikowi.

gdzie $\Omega(N)$ to zbiór wszystkich permutacji zbioru N . Ta metoda podziału nazywana jest *wartością Shapleya*. Mimo że może się ona wydawać arbitralna, Shapley pokazał, że jest to jedyny liniowy sposób podziału całej wypłaty w sposób, który symetrycznym graczom przypisuje taką samą wartość, a graczom, którzy nie wnoszą nic do żadnej koalicji, nie daje nic. Wartość Shapleya w przykładzie powyżej da firmie A (od lodów) 7 złotych, a firmie B (od patyków) 5 złotych.

Użyjemy teraz wartości Shapleya do znalezienia wierzchołka najistotniejszego z punktu widzenia utrzymywania spójności grafu. Naszym zbiorem graczy są teraz wierzchołki grafu, a funkcją oceniającą koalicje f^* . Nasza gra to zatem (V, f^*) . Zastosowanie wzoru na wartość Shapleya da nam następujący wynik:

$$A(v) = \frac{1}{|V|!} \sum_{\omega \in \Omega(V)} |K(S_v^\pi)| - |K(S_v^\pi \cup \{v\})| + 1 \quad (\text{miara łączenia}).$$

Wkład marginalny wierzchołka v do koalicji S_v^π jest zatem równy różnicy ilości połączonych grup (komponentów), która była bez v , a będzie z graczem v plus jeden. Jest to zatem ilość komponentów z S_v^π , z którymi v jest połączony: jeżeli nie będzie połączony, zwiększy liczbę komponentów o 1 i otrzyma 0. Jeżeli złączy trzy grupy w jedną – otrzyma wartość 3. Tak jak chcieliśmy: gracz jest rzeczywiście oceniany w zależności od tego, jak dobry jest w łączeniu grafu!

Ogólny mechanizm użyty przez nas w tym artykule wygląda zatem następująco:

1. dla każdej grupy elementów określ jej wartość oraz
2. zaaplikuj wartość Shapleya.

W ten sposób dostajemy oszacowanie ważności każdego elementu, biorąc pod uwagę jego interakcje z innymi.

Wyniki, jakie otrzymujemy według tej miary, są zupełnie inne niż z podstawowych miar, a niechlubny zwycięzca poprzednich miar – Nawaf Alhazmi (5) – jest u nas dopiero szósty. Na pierwszym miejscu znajduje się Hani Hanjour (10), którego usunięcie odłącza Majeda Moqed (13) z sieci. Widzimy także, że jego sąsiedzi są dość słabo połączeni, dlatego często jest on łącznikiem pomiędzy nimi. Drugie i trzecie miejsce zajmują Hamza Alghamdi (4) oraz Marwan Al-Shehhi (15) – dwaj terroryści w całości odpowiedzialni za jeden z samolotów – bez nich pozostali trzej terroryści z tego samolotu nie byłiby w ogóle połączeni z siecią terrorystyczną. Czwartą wartość otrzymał Abdul Aziz Al-Omari (16) – wspomniany już terrorysta, który kontroluje „najniższy” na rysunku samolot. Piąty z kolei jest jego łącznik z kolejną dwójką – Waleed Alshehri (17). Wszystkie cztery wierzchołki rozcinające znalazły się zatem w pierwszej piątce naszego nowego rankingu – o to nam właśnie chodziło.

Sprytne użycie gier koalicyjnych pozwoliło nam zatem uzyskać pewien ranking oceniający wierzchołki według tego, jak ważne są w utrzymywaniu spójności grafu. Nie w każdej sieci ta miara będzie wskazywać najważniejsze wierzchołki i wyników tych nie można traktować jako twardych dowodów. Jest to jednak kolejna poszlaka, która w połączeniu z innymi może kiedyś pomóc agencjom bezpieczeństwa w udaremnieniu kolejnego ataku terrorystycznego.