

Damian DZIDA, Adam MARKOWSKI
Wydział Cybernetyki Wojskowej Akademii Technicznej
ul. S. Kaliskiego 2, 00-908 Warszawa
E-mail: ddzida@wat.edu.pl, amarkowski@wat.edu.pl

Rozpowszechnianie opinii w sieciach złożonych. Model wirusowy i jego weryfikacja na podstawie danych z Twittera

1 Wprowadzenie

Przedmiotem zainteresowania w niniejszym artykule jest proces rozpowszechniania opinii w populacji. W artykule przedstawiono wirusowy model opinii bazujący na teorii wpływu społecznego [2, 3]. Wirusowość modelu polega na „zarażaniu” własną opinią innych agentów, którzy to cechują się pewną odpornością na zmianę opinii.

Portale społecznościowe cieszą się coraz większą popularnością wśród internautów. Niektóre z nich, takie jak Facebook czy Twitter, stały się jednym z podstawowych mediów do komunikacji między ludźmi. Na szczególną uwagę zasługuje portal Twitter, który pomimo swojej prostoty jest wykorzystywany przez najważniejsze osoby świata polityki, w tym prezydenta Stanów Zjednoczonych [15].

Dostęp do struktury sieci oraz treści wiadomości wysyłanych przez użytkowników pozwala nam na obserwowanie i analizę zachodzących zjawisk społecznych, takich jak powstawanie opinii i jej rozpowszechnianie w społeczeństwie. Na podstawie rzeczywistych danych możliwe jest tworzenie modeli opinii, a następnie ich weryfikowanie z wykorzystaniem symulacji.

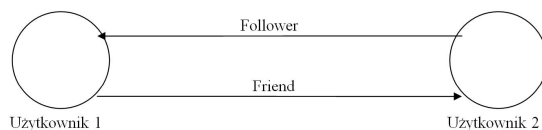
W niniejszym artykule przedstawiono modele wpływu społecznego oraz wyniki symulacji dotyczące zjawiska agitacji poprzez głoszenie poglądu zaobserwowanego w sieci Twitter. W punkcie drugim zaprezentowano strukturę sieci społecznej portalu Twitter oraz jej wpływ na obieg informacji. W punkcie trzecim opisano Twitterową akcję "Siema Tu Polska", a w punkcie czwartym przeanalizowano zaobserwowaną akcję społeczną. W punkcie piątym zaprezentowano dwa modele opinii oraz przeprowadzone eksperymenty. W punkcie szóstym przedstawiono analizę wyników symulacji. Punkt siódmy jest propozycją dalszych prac badawczych.

2 Twitter

Twitter jest rozbudowanym mikroblogiem pozwalającym na dzielenie się krótkimi tekstami z innymi użytkownikami. Obecnie posiada setki milionów użytkowników i odgrywa znaczącą rolę w obiegu informacji. Dla przykładu można przytoczyć akcje społeczne, o których wspominały media, m.in. dziennikarze relacjonujący przebieg „rewolucji egipskiej” w 2011 roku donosili, że jednym z głównych źródeł wymiany informacji był portal Twitter [9]. Warto również wspomnieć o atakach DDoS w sprawie

ACTA na strony rządowe w styczniu 2012 roku, koordynowanych poprzez Twittera [10].

Sieć społeczna w Twitterze składa się z jednostek połączonych między sobą związkami zainteresowania (przyjaźni, koleżeństwa czy obserwacji). Jednostką może być użytkownik, firma, portal lub grupa ludzi komunikująca się wzajemnie. Związek zainteresowania jest jednokierunkowy. Dla kierunku wychodzącego jest to „friend” (obserwuje, lubi), natomiast dla kierunku przychodzącego jest to „follower” (jest obserwowany, jest lubiony). W uproszczeniu można przyjąć, że występuje tylko jeden rodzaj związku, czyli ‘friend’, natomiast ‘follower’ jest przeciwieństwem skierowanym ‘friend’.



Rys. 1. Związki dla Użytkownik 1

Fig. 1. Relationship for Użytkownik 1

Związki między użytkownikami mogą być dwukierunkowe, jeżeli użytkownik obserwuje innego użytkownika i jest przez niego śledzony. Sytuacje takie nie są regułą, ponieważ znane lub wpływowe osoby są śledzone przez miliony użytkowników, natomiast same śledzą niewielu. Odpowiada to bardzo dobrze strukturze społecznej, w której tak zwane „autorytety” [6] są obserwowane przez wielu użytkowników, a wpływ innych na sieć jest ograniczony.

Portal Twitter pozwalał na niemalże swobodne przeglądanie i pobieranie wszystkich wiadomości. Na prośbę udostępniał konto z nieznacznymi limitami, co badaczom w 2010 roku pozwoliło na pobranie całej sieci społecznej, tj. 41,7 miliona profili użytkowników i 1,47 miliardów powiązań oraz dodatkowo 106 milionów tweetów [8].

Badacze zaobserwowali bardzo szybką propagację treści 4-5 kroków w głąb sieci niezależnie od liczby obserwujących źródło wiadomości. Ponowne przesłanie wiadomości przez innego użytkownika (retweet) pozwalało uzyskać średnio zasięg 1000 użytkowników w bardzo krótkim czasie. Obserwacja ta wskazuje, że sieć społeczna Twittera może cechować się wysoką zaraźliwością (wirusowością).

Twitter udostępnia dwie metody pobierania danych. Pierwszą jest subskrypcja na dane wysyłane w czasie rzeczywistym („stream”), bez dostępu do danych archiwalnych. Drugą metodą są zapytania („query”). Obecnie (rok 2012) portal Twitter wprowadził znaczne ograniczenia w dostępie do danych poprzez mechanizm „query”, szczególnie do danych archiwalnych. Wyszukiwanie wiadomości jest ograniczone czasowo oraz ilościowo, przeglądanie profili użytkowników również jest ograniczone ilościowo, a ponadto związki użytkowników (kogo obserwują) mogą być prywatne, dostępne dla grona znajomych. Do utrudnień należy również zaliczyć możliwość zmiany nazwy użytkownika, kasowania kont oraz ciągła zmiana sieci powiązań.

Dane do niniejszego artykułu pobrano przypadkowo przy pomocy metody „stream” (wcześniejsza subskrypcja na słowo 'polska'), a sieć społeczna użytkowników została odtworzona przy wykorzystaniu mechanizmu zapytań („query”).

3 Akcja „Siema Tu Polska”

Na początku lutego 2012 roku jedna z internatek [11] zaproponowała akcję, by do każdej wiadomości dopisywać frazę „Siema Tu Polska”. Celem akcji miało być dostanie się do „Twitter Trends” (najpopularniejszych fraz), co miało skutkować konsternacją wśród międzynarodowej społeczności Twittera. Akcja przeszła bez większego echa, jednak pozostawiła po sobie ślad w tej społeczności. Dwa tygodnie później inna internautka [12] zaproponowała identyczną akcję, by tym razem osiągnąć cel, który został odnotowany w mediach tradycyjnych [13].

Dwie powyżej opisane akcje były spontaniczne i organizowane głównie poprzez wykorzystanie portalu Twitter. Po sukcesie powstały inicjatywy powtórzenia akcji. W kolejnych próbach wykorzystano do organizacji i koordynacji inne portale społecznościowe, m.in. Facebook. W niniejszym artykule wykorzystano próbki danych dotyczącą dwóch pierwszych akcji.

4 Analiza akcji "Siema Tu Polska"

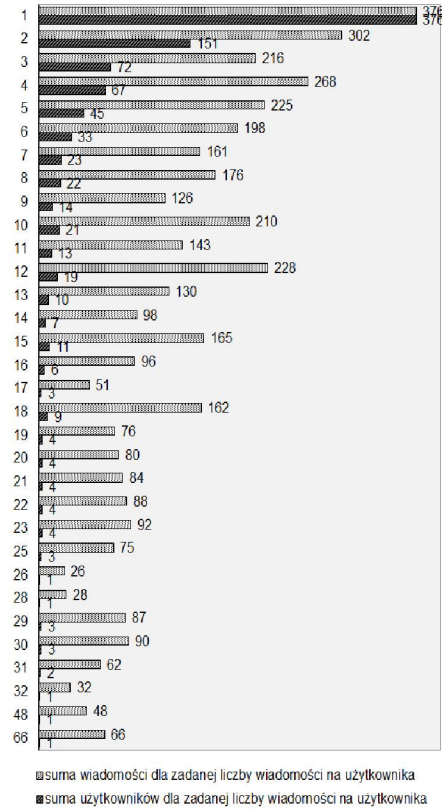
Do badań wykorzystano 4265 zebranych wiadomości. Pierwsza wiadomość została przesłana 2012-02-02 00:11, a ostatnia 2012-02-22 07:17.

Pobieżna analiza danych pozwala przypuszczać, że źródłem akcji "Siema Tu Polska" jest młodzież gimnazjalna. W początkowej fazie uczestniczyły głównie dziewczyny, które są fankami piosenkarza i aktora Justina Biebera. Aktor ten zaistniał na portalu społecznościowym YouTube [14], a w chwili obecnej ma oddaną sobie społeczność na wielu portalach, w tym na Twitterze, sięgającą kilkudziesięciu milionów [14]. W kolejnej fazie akcji brała udział dodatkowo młodzież gimnazjalna niezwiązana ze społecznością piosenkarza, by następnie wyjść poza kręgi gimnazjalne, czyli użytkowników postronnych.

Na wykresie przedstawiono udział w wiadomościach użytkowników z różnym stopniem aktywności. Przykładowo – tylko jeden użytkownik przesłał 66 wiadomości, a 376 użytkowników przesłało pojedynczą wiadomość. Z rysunku 2 widać, że w akcji większość użytkowników brała udział od kilku do kilkudziesięciu razy. Wielu użytkowników brało udział tylko raz, można więc przyjąć, że byli to mało aktywni lub przypadkowi uczestnicy, którzy zostali „zarażeni” akcją.

Ze względu na duży przedział czasowy akcji do analizy wykorzystano dane z początku drugiej udanej próby, do jej wygaśnięcia, tj. od 2012.02.19 17:35 do 2012.02.19 23:21.

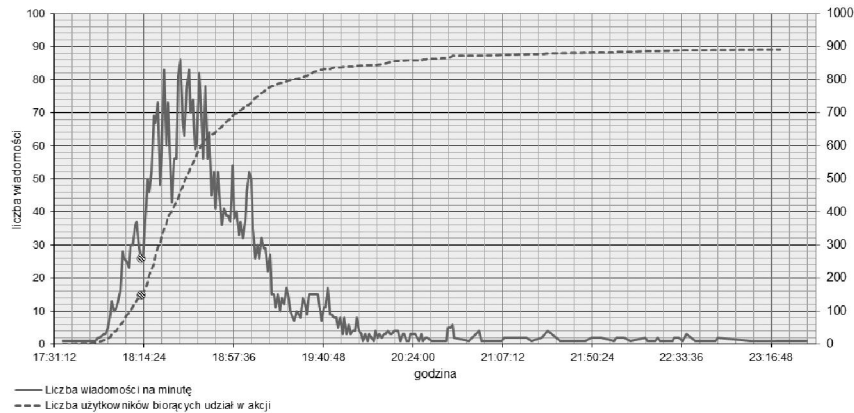
Pierwsze wzmianki o pojawieniu się frazy "Siema Tu Polska" w Twitter Trends odnotowano o godzinie 18:13, czyli na chwilę przed silnym wzrostem liczby wiadomości oraz użytkowników biorących udział w akcji (rys. 3). Powodem silnego wzrostu było dostanie się do Trends, co przyczyniło się do dyskusji o fakcie powodzenia akcji, oraz zdziwieniem wywołanym przez pojawienie się polskiej frazy w Trends. Z upływem czasu zauważalny jest powolny, w miarę jednostajny spadek zainteresowania akcją.



Rys. 2. Suma wiadomości i użytkowników w zależności od liczby wiadomości na użytkownika

Fig. 2. Sum of tweets and users

Do dalszej analizy oraz symulacji wykorzystano dane z początku drugiej akcji – do momentu dostania się do Twitter Trends.



Rys. 3. Liczba wiadomości na minutę oraz liczba użytkowników biorących udział w akcji

Fig. 3. Number of tweets per minute and number of users taking part in action

5 Proponowany model opinii

W badaniach wykorzystano model, który jest rozwinięciem modelu zaproponowanego przez Latané i Nowaka [2] [3]. Jest to model bazujący na teorii wpływu społecznego, która została zaprezentowana przez Latana w pracy [1]. Wpływ społeczny definiowany jest następująco: „Jest to jakakolwiek zmiana uczuć, myśli i zachowania wywołana rzeczywistą bądź wyobraźną obecnością innych ludzi”. Model wpływu społecznego zakłada, że wielkość wpływu wywierana przez grupę na jednostkę zależy od:

- siły przekonywania osób wywierających wpływ (S),
- bliskości, w jakiej pozostają osoby wywierające wpływ od osoby poddanej oddziaływaniu (V),
- liczby osób wywierających wpływ (N).

Zależność powyższych czynników można wyrazić jako funkcje trzech wielkości S, V, N [2]:

$$W=f(S, V, N) . \tag{1}$$

Siła przekonywania S oznacza zbiorczo sumę wszystkich czynników decydujących o sile przekonywania pojedynczej osoby. Są to zarówno czynniki stałe, charakteryzujące daną osobę (status społeczny, umiejętność przekonywania) i charakterystyczne dla danej sytuacji, np. chęć do przekonywania innych. Bliskość według autorów [3, s.191] oznacza „intensywność kontaktu między źródłem i obiektem wpływu w przestrzeni społecznej”. Odległość fizyczna między źródłem a obiektem nie jest jednak tożsama z odległością w przestrzeni społecznej. Jest jedynie ważnym składnikiem odległości w przestrzeni społecznej. Innym elementem mającym wpływ na bliskość może być stopień pokrewieństwa między jednostkami.

Model Latané-Nowaka zakłada, że podstawą zmiany opinii jest wpływ społeczny. Jednostka w rozmowach z innymi sprawdza stopień poparcia, którym cieszą się różne opinie. Założono, że jednostka zmienia opinię wtedy, gdy inna opinia znajduje większe

poparcie niż jej własna. Formalnie można zdefiniować każdą niezależną jednostkę, jako inteligentnego agenta. Załóżmy, że populacja jest złożona z N agentów.

Każdy agent i posiada następujące cechy:

- o_i – aktualna opinia popierana przez agenta i – aktualny stan agenta, wyrażony poprzez popieraną opinię;
- s_i – siła wpływu agenta i – określająca zdolność jednostki do przekonania innych jednostek;
- ϕ_i – stopień sąsiedztwa agenta i – liczba naturalna określająca promień, w obrębie którego agent może komunikować się z innymi agentami; w przypadku sieci społecznych jest to długość minimalnego łańcucha między węzłami sieci, przy czym węzłom odpowiadają agenci; długość minimalnego łańcucha jest to najmniejsza liczba krawędzi łącząca dwa węzły; w przypadku grafu skierowanego jest to minimalna długość drogi łącząca dwa węzły [7].

Wielkość wpływu społecznego agentów o ustalonym poglądzie I_i , odbierany przez agenta i , jest równy [2]:

$$I_i = \sqrt{\left(\sum_{j=1}^{N_s} \left(\frac{s_j}{d_{ij}^2}\right)^2\right)}, \quad (2)$$

gdzie:

N_s – liczność agentów o danym/ustalonym poglądzie.

Współczynnik d_{ij} określa odległość między agentem i , a j . W oryginalnym modelu wykorzystano automaty komórkowe, które składają się z n -wymiarowej regularnej, dyskretnej siatki komórek. Każdej komórce był przyporządkowany dokładnie jeden agent. W takim przypadku odległość pomiędzy agentami była wyliczana jako odległość euklidesowa pomiędzy komórkami.

W proponowanym modelu wprowadzono następujące modyfikacje w celu lepszego odwzorowania rzeczywistego zjawiska rozpowszechniania plotki:

- zastąpienie sieci regularnej siecią złożoną jako modelu związków pomiędzy agentami,
- wprowadzenie modelu komunikacji między agentami wraz historią agenta dotyczącą kontaktów między agentami,
- wprowadzenie zachowania agitacji oraz modyfikowanie wartości poparcia opinii poprzez uwzględnienie wpływu agitacji,
- modyfikacja cech jednostki poprzez dodanie atrybutu „zdolność komunikacyjnej”, stopień sąsiedztwa,
- zaproponowanie algorytmu symulacji krokowej w celu weryfikacji modelu opinii.

W proponowanym modelu wykorzystywane są sieci złożone, w których każdemu węzłowi odpowiada dokładnie jeden agent. W tym przypadku odległość d_{ij} jest wyznaczana jako długość minimalnego łańcucha łączącego dwa węzły i i j . Formalnie model zachowania zmiany opinii przez agenta i można zdefiniować w następujący sposób:

$$A_t = \left\{ \begin{array}{l} a_{it} \in (s_i, \varphi_i, v_{it}, o_{it}, \mu_i, m_{it}): t \in \{0,1,2,3, \dots\}, \\ s_t \in (0,1), \varphi_i \in \mathbb{N}, v_{it} \in V_t, o_{it} \in P, \mu_i \in \langle 0,1 \rangle, m_{it} \in M \end{array} \right\};$$

$$t \in \{0,1,2,3, \dots\}, \quad (3)$$

$$|A_t| = N \text{ dla danego } t,$$

$$P = \{p_1, \dots, p_K\}, |P| = K,$$

$$G_t = \{V_t, E_t\}, v_{it} \in V_t, |V_t| = N, E_t \subset V_t \times V_t,$$

$$d_{ijt} = \min[l_1(v_{it}, v_{jt}), l_2(v_{it}, v_{jt}), \dots, l_h(v_{it}, v_{jt})];$$

$$d_{ijt} \in \mathbb{N},$$

$$a_{it} = (s_i, \varphi_i, v_{it}, o_{it}, \mu_i, m_{it}) \in A_t,$$

$$w_{it} = [w_{it1} w_{it2} \dots w_{itK}],$$

$$w_{itk} = \sqrt{\sum_{j=1}^N \left[\frac{s_j}{(d_{ijt})^2} f(p_k, o_{jt}, d_{ijt} - \varphi_{it}) \right]^2},$$

$$f: P \times P \times R \rightarrow \{0,1\}, f(p, o, d) = \begin{cases} 1 & \text{jeżeli } d \leq 0 \text{ i } p = o \\ 0 & \text{inaczej} \end{cases}$$

gdzie:

A_t – model wszystkich agentów w chwili t , a_{it} – agent i w chwili t ,

t – krok symulacji,

P – zbiór opinii,

G_t – graf w chwili t , stanowiący odzwierciedlenie związków (relacji) między agentami w chwili t ; V_t – zbiór węzłów grafu G_t ; E_t – zbiór krawędzi grafu G_t ,

$l_h(v_{it}, v_{jt})$ – liczba krawędzi łącząca węzły v_{it} i v_{jt} w łańcuchu h ,

μ_i – zdolność komunikacyjna agenta i – określenie procentowej liczby osób z otoczenia, z którymi agent może się komunikować,

M – zbiór typów agenta, m_{it} – typ agenta a_{it} w chwili t ,

w_{itk} – wartość poparcia k -tej opinii przez sąsiadów agenta i znajdujących się w odległości φ_i ,

f – funkcja definiująca możliwość komunikacji między agentami.

Algorytm zmiany opinii przez agenta jest następujący:

- Wyznacz w_{it} dla $i=1..N$ w danym czasie t .
- Znajdź taką opinię p_{u_i} dla każdego agenta i , $i=1..N$, której wartość poparcia w_{itk} jest maksymalna. Następnie dla każdego agenta dokonaj zmiany opinii na nowo wyznaczoną opinię p_{u_i} :

$$u = [u_1, \dots, u_N],$$

$$u_i^* = \arg \max_{k=1, \dots, K} w_{itk}, \quad (4)$$

$$a_{it+1} = (s_i, d_i, v_i, p_{u_i^*}) \text{ dla } i = 1 \dots N.$$

W powyższej definicji modelu przyjęto, że istnieje możliwość zmiany położenia agenta w grafie (G_t) oraz odległości pomiędzy poszczególnymi węzłami (d_{ijt}) w czasie symulacji. Jednak w przeprowadzonych badaniach wprowadzono ograniczenie na brak możliwości zmiany struktury grafu oraz brak możliwości zmiany położenia agenta w sieci w trakcie symulacji. Wprowadzone ograniczenia pozwalają na uproszczenie modelu, co ma istotny wpływ na możliwość jego analizy. Dodatkowo struktura sieci uczestników biorących udział w opisanej akcji jest stała i dlatego przyjęcie powyższego ograniczenia nie ma wpływu na możliwość zbadania adekwatności modelu. Powyższe ograniczenie można zdefiniować w następujący sposób:

$$\begin{aligned} G_t &= \{V_t, E_t\}, \\ G_0 &= G_1 = G_2 \dots = G, \\ d_{ij0} &= d_{ij1} = d_{ij2} \dots = d_{ij}. \end{aligned}$$

W modelu agenta zdefiniowano φ_i jako maksymalną liczbę krawędzi, w obrębie której agent może komunikować się z swoimi sąsiadami. Dla wartości $\varphi_i=1$ agent może komunikować się jedynie z najbliższymi swoimi sąsiadami. W takim przypadku można interpretować krawędź łączą dwóch agentów jako relację przyjaźni zdefiniowaną w rozdziale 2. W przypadku wartości $\varphi_i>1$ występuje możliwość kontaktowania się agentów z sąsiadami swoich najbliższych sąsiadów. Mogłoby to opisywać relację określaną jako poznawanie opinii osób za pośrednictwem innych osób. Takie możliwości udostępnia portal Twitter poprzez wykonywanie retweetów, czyli cytowania opinii innych osób w wysyłanych wiadomościach. W przeprowadzonych badaniach przyjęto, że wartość φ_i jest zawsze równa 1.

Cecha typ agenta m_i określa możliwe zachowania agenta i w ogólności może być zależna od czasu. W proponowanym modelu typ agenta jest stały i niezależny od czasu dlatego $m_{it}=m_i$.

W niniejszym artykule proponuje się wprowadzenie do powyższego modelu istotnych zmian w celu lepszego odzwierciedlenia procesu rozpowszechniania opinii w społeczeństwie. Pierwszą istotną zmianą jest wprowadzenie autorskiego modelu komunikacji między agentami ACM (ang. Advanced Communication Model). Pierwotny model nie uwzględnia problemu komunikacji między agentami. Zakłada, że w każdym kroku symulacji wszyscy agenci komunikują się ze swoim otoczeniem. W rzeczywistości proces komunikacji jest bardziej złożony i odgrywa istotną rolę w rozpowszechnianiu opinii. Proponowany model komunikacji ACM zakłada, że agenci komunikują się częściej z osobami z swojego otoczenia o podobnych opiniach. Każdy agent rejestruje historię swoich kontaktów z innymi agentami. Kontakt może być pozytywny lub negatywny. Pozytywny kontakt ma miejsce w przypadku, gdy agenci mają takie same opinie. W przeciwnym razie kontakt jest negatywny. Długość pamiętanej historii kontaktów jest nieograniczona, tzn. agent zapamiętuje całą historię kontaktów w trakcie symulacji. Na podstawie historii kontaktów oraz odległości między agentami wyznaczana jest dla każdego agenta ocena. Funkcja oceny agenta j znajdującego się w otoczeniu agenta i definiowana jest w następujący sposób:

$$q_{ijt} = \frac{1}{(d_{ij})^2} + \frac{pcon_{ijt}}{pcon_{ijt} + ncon_{ijt}} \quad (5)$$

gdzie:

q_{ijt} – ocena agenta j przez agenta i w chwili t ,
 d_{ij} – minimalna odległość między agentem i a j ; w przypadku sieci minimalna liczba krawędzi między węzłami v_i i v_j ,
 $pcon_{ijt}$ – liczba pozytywnych kontaktów agenta i i j do chwili t ,
 $ncon_{ijt}$ – liczba negatywnych kontaktów agenta i i j do chwili t .

Na podstawie ocen tworzona jest funkcja prawdopodobieństwa wyboru agenta, która zdefiniowana jest w następujący sposób:

$$pr_{it} = \left[\frac{q_{i1t}}{\sum_{k=1}^{J_i} q_{ikt}}, \frac{q_{i2t}}{\sum_{k=1}^{J_i} q_{ikt}}, \dots, \frac{q_{iJ_it}}{\sum_{k=1}^{J_i} q_{ikt}} \right] \quad (6)$$

gdzie:

J_i – liczba agentów znajdujących w otoczeniu agenta i .

W celu określenia procentowej liczby osób z otoczenia, z którymi agent może się komunikować wprowadzono dodatkową cechę „zdolność komunikacyjną (μ)” do modelu wszystkich agentów A_i . Wartość cechy „zdolność komunikacyjna” jest stała w trakcie symulacji. Cały algorytm komunikacji ACM jest następujący:

Algorytm 1. Algorytm komunikacji ACM

1. Wyznacz wartość oceny $Q = [q_{i1t}, q_{i2t}, \dots, q_{iJ_it}]$
2. Wyznacz funkcję prawdopodobieństwa PR_i
3. Wylosuj $m = [\mu_i J_i]$ agentów zgodnie z funkcją prawdopodobieństwa PR_i
4. Zaktualizuj historię kontaktów

W danym kroku symulacji nie biorą udziału wszyscy agenci, ale pewna podgrupa losowo wybrana z rozkładem równomiernym. Są to „aktywni agenci”, mogący zmieniać swój pogląd. Liczność aktywnych agentów (podgrupy) w danym kroku symulacji jest ustalana losowo z rozkładem równomiernym z przedziału $\langle y, z \rangle$ gdzie $y, z \in \mathbb{N}$, $0 < y \leq z \leq N$. Wartości y, z są parametrami globalnym symulacji i nie ulegają zmianie podczas symulacji.

Wprowadzenie zachowania agitacji

Początkowe uzyskane wyniki symulacyjne wykazały, że w opisywanym modelu brak jest możliwości symulacji zjawiska agitacji, podejmowanych przez jednostki w celu przekonania innych do swojej opinii. Brak aktywności agentów powoduje, że zastosowanie modelu w zjawiskach o charakterze wirusowym jest praktycznie niemożliwe. W celu odwzorowania zjawiska agitacji wprowadzono do modelu możliwość przesyłania komunikatów między agentami. Komunikat zawiera informację o popieranej opinii oraz osobie, która go wysłała. Jednostki mogą wysłać komunikaty lub odbierać je. W przypadku wysyłania komunikatów agent-agitator z pewną intensywnością wysyła komunikaty do agentów ze swojego otoczenia. Strategia wyboru agentów, do których zostanie przesłany komunikat, została ustalona w oparciu o siłę przekonywania (s) oraz aktualną opinię agenta, do którego zostanie wysłana wiadomość. W przypadku agitacji lepiej jest przyjąć, że agitowani agenci posiadają siłę przekonania (s) do własnego poglądu. Tym samym, im mniejsza jest siła przekonania, tym bardziej prawdopodobne jest przekonanie jednostki, dlatego agent będzie preferował jednostki o mniejszej sile

przekonania do aktualnej opinii. Dodatkowo agent wybiera z większym prawdopodobieństwem jednostki, które popierają opinię inną niż agitowana.

Prawdopodobieństwo wysłania wiadomości agentowi j znajdującemu się w otoczeniu agenta i ($pr_{s_{ijt}}$) w chwili t jest równe:

$$pr_{s_{ijt}} = \frac{ps_{ijt}}{\sum_{k=1}^{J_i} ps_{ikt}}$$

$$ps_{ijt} = \frac{1}{s_j} + e(o_{it}, o_{jt})$$

$$e: PxP \rightarrow \mathbb{R} \quad (7)$$

$$e(o_i, o_j) = \begin{cases} 0 & \text{gdy } o_i = o_j \\ 0.5 & \text{gdy } o_i \neq o_j \end{cases}$$

$$Pr_{s_{it}} = \left[\frac{ps_{i1t}}{\sum_{k=1}^{J_i} ps_{ikt}}, \frac{ps_{i2t}}{\sum_{k=1}^{J_i} ps_{ikt}}, \dots, \frac{ps_{iJ_it}}{\sum_{k=1}^{J_i} ps_{ikt}} \right]$$

gdzie:

$pr_{s_{ijt}}$ – prawdopodobieństwo wysłania wiadomości do agenta j przez agenta i w chwili t ,

ps_{ijt} – ocena agenta j jako celu agitacji dla agenta i w chwili t ,

s_j – siła przekonywania (przekonania) agenta j ,

o_{it} – opinia agenta i w chwili t .

Kolejnym krokiem jest określenie współczynnika intensywności wysyłania wiadomości (α), czyli liczby wiadomości wysyłanych w danym kroku symulacji. W przeprowadzonych eksperymentach przyjęto, że wartość α rośnie liniowo o zadaną wartość δ , gdy liczba popierających agitowaną opinię maleje. W przypadku, gdy liczba popierających jest stała lub rośnie, wartość nie ulega zmianie. Wprowadzono ograniczenia, na minimalną i maksymalną wartość współczynnika intensywności wysyłania wiadomości. W przypadku, gdy liczba popierających agitowaną opinię rośnie przez pewien określony czas, agent próbuje zmniejszyć intensywność o wartość δ . Algorytm zmiany współczynnika intensywności jest następujący:

Algorytm 2. Algorytm zmiany współczynnika intensywności wysyłania wiadomości

1. Wyznacz liczbę agentów popierających agitowaną opinię w kroku poprzednim Na_{t-1}
2. Wyznacz liczbę agentów aktualnie popierających agitowaną opinię Na_t
3. Wyznacz różnicę $\Delta Na = Na_t - Na_{t-1}$
4. Jeżeli $\Delta Na > 0$ to:
 - a. Jeżeli wartość $\Delta Na > 0$ utrzymuje się przez Δt kroków to:

$$\alpha = \alpha + \Delta \alpha$$
 Jeżeli $\alpha < \alpha_{\min}$ to

$$\alpha = \alpha_{\min}$$
5. Jeżeli $\Delta Na < 0$ to:

- a. Jeżeli $\alpha = \alpha_{\max}$ i wartość $\Delta Na < 0$ utrzymuje się przez Δt kroków to $\alpha = \alpha_{\min}$. W przeciwnym przypadku przejdź do 5.B
 - b. Jeżeli $\alpha < \alpha_{\max}$ to
 $\alpha = \alpha + \Delta\alpha$
 Jeżeli $\alpha > \alpha_{\max}$ to:
 $\alpha = \alpha_{\max}$
 6. Jeżeli $\Delta Na = 0$ to nie zmieniaj wartość intensywności α
- Cały algorytm agitacji agenta jest następujący:

Algorytm 3. Algorytm agitacji agenta

1. Wyznacz prawdopodobieństwo (Prs_i) wysłania komunikatu dla wszystkich agentów znajdujących się w otoczeniu danego agent i .
2. Wyznacz wartość intensywności α
3. Dla $k = 0$ do $k < \alpha$ wykonaj:
 - a. Wylosuj agenta j zgodnie z Prs_i
 - b. Wyślij wiadomość do agenta j

Natomiast interpretacja przesyłanych komunikatów jest dokonywana niezależnie przez każdego agenta. Odebrane komunikaty są kolejgowane przez agenta, przy czym przyjęto założenie o nieograniczonej pojemności kolejki. Wpływ odebranych komunikatów na wybór opinii jest wyznaczany zgodnie z poniższym wzorem:

$$Q_i = \{(t, j, o) \in \mathbb{N} \times \mathbb{N} \times P \mid a_j \in A_t\} \quad (8)$$

$$Q_{ik} = \{(t, j, o) \in Q_i \mid o = o_k\}$$

$$Q_{ik} \subset Q_i$$

$$T_i = \left\{ t \in \mathbb{N} \mid \begin{array}{l} \text{istnieje } j \in \mathbb{N} \text{ i } o \in P, \\ \text{że } (t, j, o) \in Q_i \end{array} \right\}$$

$$ag_{ik} = \frac{\sum_{q \in Q_{ik}} g(q)}{\sum_{t \in T_i} t} \quad (8)$$

$$g: Q \rightarrow \mathbb{R}, g(t, j, o) = t \cdot s_j$$

$$w'_{ikt} = \gamma \cdot w_{ikt} + \beta \cdot ag_{ik}$$

$$\gamma, \beta \in (0, 1), \gamma + \beta = 1$$

gdzie:

Q_i – kolejka komunikatów agenta i , przy czym t – położenie w kolejce, j – numer agenta, który wysłał komunikat, o – agitowana opinia

Q_{ik} – kolejka komunikatów z wiadomościami o_k agenta i

ag_{ik} – wartość wpływu wiadomości o_k z kolejki na agenta i

w'_{ikt} – zmodyfikowana wartość poparcia opinii k (o_k) przez agenta i w chwili t

γ – wpływ opinii z otoczenia agentów

β – wpływ przesłanych komunikatów.

Agent na podstawie wektora w'_{ik} wybiera opinię, przy czym kryterium wyboru nie ulega zmianie tzn. wybiera opinię, dla której wartość w'_{ik} jest maksymalna. Komunikaty po odczytaniu są usuwane z kolejki. W eksperymentach zdefiniowano trzy rodzaje typów agenta:

$$M = \{m1, m2, m3\} ,$$

gdzie:

$m1$ – zwykły agent bez możliwości interpretacji komunikatów – działający z algorytmem komunikacji ACM [Algorytm 1] bez możliwości wysyłania komunikatu i interpretacji otrzymanych komunikatów,

$m2$ – zwykły agent interpretujący komunikaty – posiadający algorytm komunikacji ACM [Algorytm 1] i interpretujący otrzymane komunikaty [Algorytm 2],

$m3$ – agitator – agent posiadający tylko zdolność wysyłania komunikatów, bez możliwości zmiany swojej opinii [Algorytm 3].

Pełny algorytm symulacji jest następujący:

Algorytm 4. Algorytm symulacji

1. Określ maksymalną liczbę kroków symulacji t_{max}
2. Zainicjuj krok symulacji $t=0$
3. Wylosuj liczbę n z przedziału $\langle y, z \rangle$ w każdym kroku symulacji t , określając liczbę agentów uczestniczących w wymianie opinii
4. Zainicjuj zbiór wylosowanych agentów S jako zbiór pusty.
5. Dla $j=1$ do n wykonaj:
 - a. Wylosuj agenta i z populacji
 - b. Jeżeli agent i nie należy do S to:
 - i. Dodaj agenta i do S
 - ii. W zależności od cechy m_i agenta wykonaj:

Dla $m_i = m1$:

 - a. Wykonaj [Algorytm 1]
 - b. Wyznacz opinię zgodnie z wzorem nr (4)

Dla $m_i = m2$:

 - a. Wykonaj [Algorytm 1]
 - b. Wyznacz opinię zgodnie z wzorem nr (8)

Dla $m_i = m3$:

 - a. Wykonaj [Algorytm 3]
 - c. W przeciwnym przypadku wróć do pkt. 5.A
6. $t=t+1$
7. Jeżeli $t < t_{max}$ to wróć do pkt. 3.
8. Koniec symulacji

S – zbiór wylosowanych agentów w danym kroku symulacji.

Wykorzystane modele agentów w eksperymentach

W eksperymentach wykorzystano dwa modele agentów. W pierwszym eksperymencie wykorzystano model agenta z ACM, bez możliwości wysyłania komunikatów i odbierania komunikatów. W drugim eksperymencie wykorzystano dwa modele agentów. Pierwszy z

nich to model z ACM z możliwością interpretacji odbieranych komunikatów zgodnie z wzorem 8, ale pozbawionych zachowania agitacji. Drugi to model agenta mającego jedynie zachowanie agitacji. W eksperymencie agenci z modelem agitacji stanowią mniejszość i są odpowiednikami inicjatorów badanej akcji. Wykorzystując przedstawione modele, przeprowadzono szereg eksperymentów symulacyjnych, których wyniki są przedstawione w kolejnym rozdziale. Do symulacji wykorzystano fragment sieci społecznej Twitter obejmujący użytkowników biorących udział w akcji do momentu osiągnięcia Twitter Trends. Za agitatorów wybrano kilku użytkowników (tab. 1), którzy jako pierwsi brali udział w akcji. Parametr siły wpływu poszczególnych agentów był losowany ze zdefiniowanego zakresu (tab. 1).

6 Wyniki symulacji

Przyjęto, że Tweet ze zwrotem „Siema Tu Polska” reprezentuje pogłoskę, która może zostać przekazana dalej. Aby wykorzystać zaobserwowane zjawisko, w modelu opinii zdefiniowano dwie przeciwstawne opinie. Pierwsza opinia „siema” oznacza zaciekawienie akcją, natomiast opinia „brak” odpowiada brakowi zainteresowania. Celem symulacji było odtworzenie opisywanej akcji. Przeprowadzono dwa eksperymenty: eksperyment z użyciem modelu bez możliwości wysyłania komunikatów i eksperyment z użyciem modelu z możliwością wysyłania komunikatów.

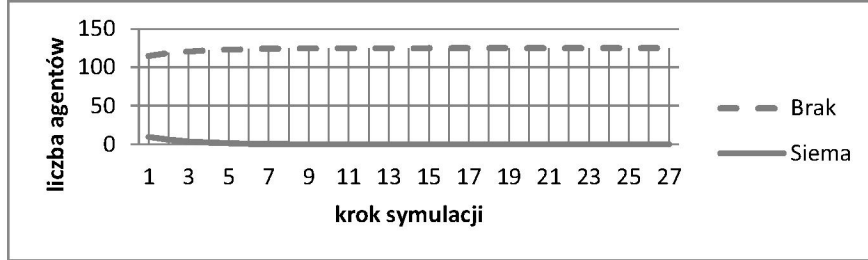
W eksperymentach wykorzystano dwa rodzaje agentów, przy czym dobór parametrów poszczególnych modeli wyznaczono metodą eksperymentalną. Poniższa tabela przedstawia wartości parametrów symulacji oraz wykorzystanych modeli agentów:

Tab. 1. Parametry symulacji dla eksperymentów

Tab. 1. The setting of simulation for experiments

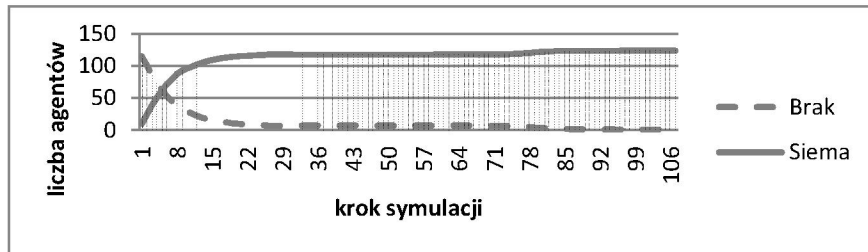
	Eksperyment dla zaproponowanego modelu							
	Bez wykorzystania agitacji				Z wykorzystaniem agitacji			
Liczba możliwych opinii	2 (S - siema, B - brak siema)							
Liczba powtórzeń	30							
Liczba kroków symulacji	1000							
Liczba agentów ogółem	125							
Wpływ opinii otoczenia i komunikatów	-				$\beta=0.7$			
	-				$\gamma=0.3$			
Zestaw	Zestaw I		Zestaw II		Zestaw III		Zestaw IV	
Opinia	S	B	S	B	S	B	S	B
Liczba agentów z początkową opinią	10	115	10	115	10	115	10	115
Przedział siły wpływu	[0.1,0.9]	[0.1,0.9]	[0.8,0.9]	[0.05,0.1]	[0.7,0.9]	[0.1,0.9]	[0.8,0.9]	[0.1,0.9]
Stopień sąsiedztwa	1	1	4	1	4	4	4	4
Komunikacja	-	-	-	-	0.4	0.8	0.4	0.8

Uzyskano następujące wyniki:



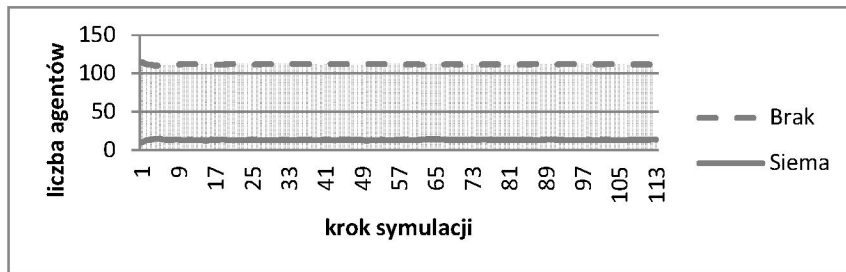
Rys. 4 Eksperyment bez wysyłania komunikatów Zestaw I

Fig. 4 Simulation result for model I with no sending messages



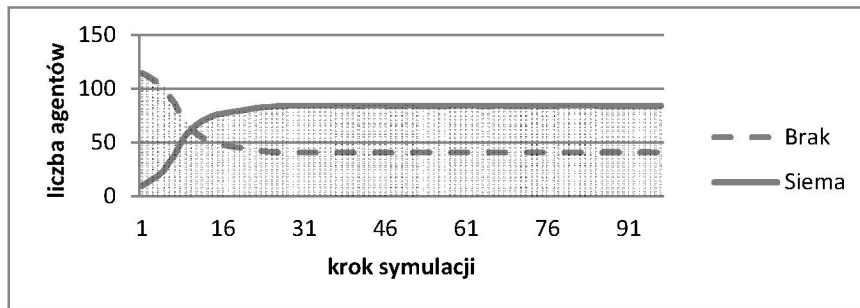
Rys. 5 Eksperyment bez wysyłania komunikatów Zestaw II

Fig. 5 Simulation result for model II with no sending messages



Rys. 6 Eksperyment z wysyłaniem komunikatów Zestaw III

Fig. 6 Simulation result for model III with sending messages



Rys. 7 Eksperyment z wysyłaniem komunikatów model IV

Fig. 7 Simulation result for model IV with sending messages

W eksperymencie bez wysyłania komunikatów udało się uzyskać dominację poglądu „Siema”, jednak w modelu przy znacznym ograniczeniu wpływu poglądu „Brak” (Zestaw II). W eksperymencie z wysyłaniem komunikatów udało się uzyskać dominację poglądu „Siema” już przy zwiększeniu siły wpływu z 0.7-0.9 (Zestaw III) na 0.8-0.9 (Zestaw IV).

Wyniki przeprowadzonych eksperymentów potwierdzają, że do symulacji zjawisk agitacji bardziej adekwatny jest model z wysyłaniem komunikatów.

7 Dalsze prace

Symulacje przedstawione w niniejszym artykule ukazują lepszą skuteczność w zarażeniu sieci przy wykorzystaniu agentów z wysyłaniem komunikatów. Model wymaga dalszych prac, szczególnie związanych z wyznaczaniem współczynnika intensywności wysyłania komunikatów oraz sposobem wyznaczania wartości poparcia. Zaprezentowany w artykule algorytm zmiany wartości intensywności wysyłania wiadomości (α) wykorzystuje funkcję liniową. Wydaje się, że na podstawie samego zjawiska agitacji, które jest intensywne na początku akcji i wraz z upływającym czasem maleje, lepszą funkcją byłaby pewnego rodzaju funkcja nieliniowa. Wymaga to dodatkowych badań, które być może pozwolą na wyznaczenie adekwatnej funkcji.

Dodatkowo należy zbadać sam sposób wyznaczania poparcia opinii. Wydaje się, że w obecnym modelu zbyt duży wpływ na wartość poparcia opinii ma parametr siły przekonywania. Obecny model nie pozwala na odzwierciedlenie jednostek pasywnych, które mają małą siłę przekonywania, ale są trudne do przekonania. Być może należałoby wprowadzić do modelu dodatkową cechę określającą „odporność” na wpływ innych jednostek.

Innym interesującym zagadnieniem, wartym dalszych badań, jest zmiana zachowania agentów z popierających daną opinię na agitatorów danej opinii.

W przyszłości można przeprowadzić dokładniejszą analizę akcji „Siema Tu Polska”, z uwzględnieniem pozostałych danych.

Literatura

1. Latane B., „*The psychology of social impact*”, American Psychologist
2. Latane B, Nowak A, Szamrej J.: From Private Attitude to Public Opinion: A Dynamic Theory of Social Impact. *Psychological Review*, 1990

3. Nowak A, et a.: *Układy złożone w naukach społecznych*. Wydawnictwo Naukowe Scholar, Warszawa, 2009
4. Boccaletti S., Latora V., Moreno Y., Chavez M., Hwang D.: Complex networks: Structure and dynamics, *Phys Rep*, 424: 175–308, 2006
5. Fronczak A. Fronczak P.: *Świat sieci złożonych*. Wydawnictwo Naukowe PWN, Warszawa, 2009
6. Kleinberg J.: Authoritative Sources in a Hyperlinked Environment, *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, 1998
7. Korzan B.: *Elementy teorii grafów i sieci: metody i zastosowania*, Wydawnictwo Naukowo-Techniczne Warszawa, 1978
8. Kwak H., Lee C., Park H., Moon S.: What is Twitter, a Social Network or a News Media?“, *WWW '10 Proceedings of the 19th international conference on World wide web*, 591-600 (2010)
9. Hudson J.: The 'Twitter Revolution' Debate: The Egyptian Test Case, *The Atlantic Wire*, <http://www.theatlanticwire.com/global/2011/01/the-twitter-revolution-debate-the-egyptian-test-case/21296/>
10. Atak za ACTA? Anonimous blokuje strony rządowe, *Rzeczpospolita*, 2012, <http://www.rp.pl/artykul/795844.html>
11. *Twitter*, http://www.twitter.com/bs_natja/status/164848378348449794, <http://www.webcitation.org/6NHfRBAvw>
12. *Twitter*, <http://www.twitter.com/AwwBieeber/status/171271697113169922>, <http://www.webcitation.org/6NHfWCRap>
13. "Siema Tu Polska". Polacy podbijają Twittera!“, *Polskie Radio*, 2012, <http://www.polskieradio.pl/5/3/Artykul/572847,Siema-Tu-Polska-Polacy-podbijaja-Twittera>, <http://www.webcitation.org/6NHfavrFv>
14. Biography for Justin Bieber, *IMDB*, <http://www.imdb.com/name/nm3595501/bio>, <http://www.webcitation.org/6NHhbyN7L>
15. Barack Obama victory tweet becomes most retweeted ever, *The Guardian*, 2012, <http://www.guardian.co.uk/world/2012/nov/07/how-barack-obama-celebrated-twitter>, <http://www.webcitation.org/6NHhhJr6d>

Streszczenie

Przedmiotem zainteresowania w niniejszym artykule jest proces rozpowszechniania opinii w populacji. W artykule przedstawiono wirusowy model opinii bazujący na teorii wpływu społecznego. Wraz ze wzrostem popularności portali społecznościowych możliwe stało się zbieranie danych rzeczywistych dotyczących zjawisk społecznych. Do budowy oraz kalibracji zaprezentowanego modelu wykorzystano dane z portalu Twitter, które były powiązane z akcją „Siema tu Polska”. Wykonano analizę jakościową oraz ilościową zaobserwowanej akcji, a następnie na jej podstawie zaproponowano model opinii. Weryfikacja modelu została dokonana poprzez symulację wieloagentową. Przedstawiono wyniki symulacji oraz ich analizę, a następnie zaproponowano dalszy kierunek prac.

Słowa kluczowe: sieci złożone, wirusowy model rozpowszechniania opinii, dyskretna symulacja wieloagentowa

Viral model of spreading opinions in complex network

With the increasing popularity of social network portals it is easier to collect real data on social phenomena. This data can be applied to build and verify opinion models. In this article authors presents viral opinion model based on social impact theory. Construction and calibration of the model is based on data from Twitter, which have been collected during action "Siema Tu Polska". The article analyses carefully this social action. Based on this analyses authors propose new opinion model. Verification of proposed model was made through multi-agent simulation. The results of simulation were described and further work was proposed.

Keywords: complex networks, viral model of opinion propagation, multi-agent discrete simulation