

STRESZCZENIE ROZPRAWY DOKTORSKIEJ

mgr inż. Artur Karczmarczyk

**Heterogeniczne oddziaływanie na procesy rozprzestrzeniania informacji
w sieciach społecznych**

Promotor
dr hab. inż. Jarosław Jankowski, prof. ZUT

Promotor Pomocniczy
dr hab. inż. Jarosław Wątróbski, prof. UŚ

Spis treści

1. Problem badawczy	2
2. Główny cel rozprawy	5
3. Cykl publikacji wchodzących w skład rozprawy	6
4. Heterogeniczne oddziaływanie na procesy rozprzestrzeniania informacji w sieciach społecznych	9
4.1. Podstawowe definicje	9
4.2. Wielokryterialne planowanie i ewaluacja procesów rozprzestrzeniania informacji w sieciach społecznych [A1, A2]	11
4.3. Oddziaływanie na inicjalizację procesów rozprzestrzeniania z udziałem próbkowania i model doboru wielkości próbek w ujęciu wielokryterialnym [A3, A4]	15
4.4. Oddziaływanie na proces propagacji informacji poprzez wielokryterialny dobór rankingów dla węzłów zasiewowych w podejściu sekwencyjnym [A5, A10]	19
4.5. Oddziaływanie poprzez nierównomierny rozrzut prawdopodobieństwa propagacji informacji [A6]	23
4.6. Oddziaływanie poprzez wielokryterialne targetowanie [A7, A8, A9]	26
5. Otwarte zorientowane obiektowo środowisko symulacyjne do badania procesu dyfuzji informacji w sieciach złożonych [A10]	29
6. Podsumowanie	31
7. Dorobek akademicki	33
7.1. Dorobek naukowy	33
7.1.1. Profile internetowe	33
7.1.2. Wykaz prac naukowych	33
7.1.3. Charakterystyka pozostałego dorobku naukowego	36
7.1.4. Pozostałe	38
7.2. Dorobek dydaktyczny	39
7.2.1. Kursy i sylabusy	39
7.2.2. Prace dyplomowe	39
7.3. Dorobek organizacyjny	40
7.4. Dorobek zawodowy	41
7.4.1. Historia zatrudnienia	41
7.4.2. Najciekawsze projekty	41
7.4.3. Najciekawsze szkolenia i certyfikaty zawodowe	41
Spis rysunków	43
Spis tablic	44
Bibliografia	45

1. Problem badawczy

Reprezentacja sieci społecznych w systemach elektronicznych ewoluowała od wczesnych systemów technicznych do zaawansowanych mediów społecznościowych integrujących mechanizmami komunikacji i interakcji podobne do tych znanych z realnego świata. Rozwój platform społecznościowych wpłynął na potrzebę zrozumienia zachowań, wzorców i predyspozycji milionów użytkowników online i ich powiązania z zachowaniami w świecie rzeczywistym [1].

Procesy rozprzestrzeniania informacji, obok relacji społecznych i aktywności online, należą do zjawisk absorbujących uwagę zarówno badaczy, jak i praktyków. W wielu przypadkach komunikacja elektroniczna, oparta na bazie procesów rozprzestrzeniania informacji w sieciach społecznych, daje wyniki lepsze niż tradycyjne kampanie reklamowe [2]. W związku z tym, rośnie liczba firm wykorzystujących te mechanizmy by dotrzeć do potencjalnych odbiorców. Fakt silnych więzi pomiędzy znajomymi, komunikacja w sieciach społecznych charakteryzuje się zwiększoną wiarygodnością komunikacji. To sprawia, że rekomendacje zorientowane społecznościowo mają większy wpływ na docelowych odbiorców niż tradycyjny przekaz [3].

Badania związane z dyfuzją treści cyfrowych zorientowane są wokół czynników wpływających na sukces kampanii [4, 5], czynników wpływających na uczestnictwo użytkowników w procesie rozprzestrzeniania informacji [6], czy też wyboru użytkowników w sieci do inicjalizacji kampanii [7, 8]. Ponadto, badany jest wpływ roli różnych miar centralności podczas selekcji początkowych influencerów [9], role treści i struktur w sieciach [10], motywacja użytkowników do przekazywania treści [5], jak również rola emocji [11, 12] i innych czynników [13] w procesie rozprzestrzeniania informacji.

Wiele wcześniejszych badań koncentrowało się na podejściach teoretycznych i empirycznych do maksymalizacji zasięgu, czyli zwiększaniu liczby węzłów, do których udało się dotrzeć w sieci [14]. Badania te mogą opierać się na modelach analitycznych stosowanych w epidemiologii [15] lub bardziej skupiać się na strukturach i cechach sieci [16]. Inną możliwością jest wykorzystanie teorii i modeli związanych z dyfuzją innowacji [17].

Chociaż zasięg, czyli liczba zainfekowanych węzłów w sieci, jest ważną miarą sukcesu kampanii, z praktycznego punktu widzenia kampanie rozprzestrzeniania informacji w sieciach społecznych mogą mieć różne cele i specyfikę [18]. Inna strategia może być wykorzystana w celu pozyskania dużej liczby potencjalnych odbiorców w bardzo krótkim czasie niż w przypadku potrzeby osiągnięcia organicznego wzrostu bazy odbiorców o zadanych cechach. Budżet kampanii wpływa na liczbę i cechy demograficzne początkowo infekowanych węzłów zasiewowych (ang. seeds). Jakość początkowych węzłów i ich liczba mogą być kluczowym czynnikiem wpływającym na zasięg kampanii i jej ogólne wyniki. Jednakże dodatkowa alokacja budżetu może posłużyć do zwiększenia dynamiki lub czasu trwania kampanii. Z drugiej strony, decydentowi może zależeć na maksymalizacji zasięgu przy ograniczonych kosztach inicjalizacji procesu, jednak bez nacisku na szybkość dotarcia do

użytkowników w kolejnych iteracjach. Aby uwzględnić różne cele, można wykorzystać wielokryterialną ocenę procesu i dobrać parametry oraz cele zgodnie z preferencjami i priorytetami.

Do niedawna, większość badań nad procesami rozprzestrzeniania informacji w sieciach społecznych zakładało jednorodność wszystkich węzłów (użytkowników) w sieci. Oznacza to, że dotarcie do każdego użytkownika sieci było dla decydenta punktowane tak samo, jak dotarcie do dowolnego innego. Nieliczne spośród najnowszych badań zaczynają koncentrować się na kampaniach celowanych [19, 20]. W procesach tych, spośród wszystkich węzłów sieci wybiera się zbiór węzłów, do których inicjator procesu chce dotrzeć. W kampaniach takich celem jest maksymalizacja nie tyle globalnego zasięgu w sieci, co zasięgu w grupie docelowej. Pozwala to skoncentrować wysiłki na dotarciu do faktycznej grupy docelowej, a z drugiej strony na ograniczeniu niechcianej korespondencji – co pozwala uniknąć negatywnych efektów agresywnych kampanii.

Jak zostało to przedstawione powyżej, problem rozprzestrzeniania informacji w sieciach społecznych jest zagadnieniem złożonym, łączącym w sobie różnorodne założenia i wyzwania. Ta różnorodność stanowiła motywację do podjęcia badań nad heterogenicznym oddziaływaniem na procesy rozprzestrzeniania informacji w sieciach społecznych, których wyniki przedstawione są w tej rozprawie. Heterogeniczność w rozprawie oznacza oddziaływanie na proces rozprzestrzeniania informacji różnymi podejściami, a nie ograniczając się do pojedynczych miar i sposobów, co charakteryzowałoby podejście homogeniczne. Oddziaływanie na proces rozprzestrzeniania informacji natomiast oznacza sposoby wpływania na zasięg, dynamikę i inne charakterystyki procesu rozprzestrzeniania informacji, między innymi poprzez dobór węzłów zasiewowych (ang. seeds), nierównomierny rozrzut prawdopodobieństwa propagacji, czy zróżnicowane sekwencje inicjalizacji procesu. W przypadku kampanii rozprzestrzeniania informacji w sieciach społecznych, inicjator procesu może być zainteresowany nie tylko maksymalizacją zasięgu kampanii, lecz również oddziaływaniem na jej dynamikę czy też ograniczeniem wymaganego budżetu. W sekcji 4.2 zaprezentowano wielokryterialne podejście do planowania i ewaluacji procesów rozprzestrzeniania informacji w sieciach społecznych.

Badania nad procesami rozprzestrzeniania informacji w sieciach złożonych opierają się często na modelach sieci rzeczywistych, które mogą zostać pobrane z licznych repozytoriów sieciowych. Badania często opierają się na modelach teoretycznych do budowy sieci syntetycznych. Sieci takie są sparametryzowane, co pozwala skoncentrować wysiłki badawcze na konkretnych cechach sieci (zob. sekcja 4.2). Kolejnym problemem wykorzystania sieci rzeczywistych jest ich wielkość. Przeprowadzanie symulacji na modelach opartych na dużej liczbie węzłów jest czasochłonne i zasobochłonne. Ten problem zaadresowany został w sekcji 4.3.

Wiele dotychczasowych badań w zakresie rozprzestrzeniania informacji w sieciach społecznych opiera się na jak najlepszej inicjalizacji kampanii. Wysiłki w takich kampaniach koncentrują się wobec tego na znalezieniu grupy użytkowników w sieci, którym przekazanie informacji skutkować będzie osiągnięciem jak największego zasięgu informacji w sieci. Działania takie ograniczają się na ogół do wybrania użytkowników i uruchomienia kampanii poprzez pojedynczy wysiew informacji (ang. seeding). W sekcji 4.4 przedstawiono oddziaływanie na procesy rozprzestrzeniania informacji w sieciach społecznych poprzez zróżnicowane sekwencje inicjalizacji kampanii.

Poza wyborem inicjalnych węzłów do uruchomienia kampanii, w celu zwiększenia zasięgu kampanii mogą być prowadzone akcje pomocnicze, polegające na bezpośrednim oddziaływaniu na prawdopodobieństwo przekazywania informacji w sieci przy stałej liczbie węzłów inicjalnych. W sekcji 4.5 przedstawiono badania nad oddziaływaniem na procesy rozprzestrzeniania informacji w sieciach społecznych poprzez nierównomierny rozrzut prawdopodobieństwa propagacji informacji. Zbadano, jak zmieniać się będzie zasięg informacji w sieci wraz ze wzrostem prawdopodobieństwa propagacji węzłów o średnich lub niskich wartościach miar centralności (potencjalnie łatwiejszych do przekonania), zamiast celować w węzły o wysokich rankingach (potencjalnie trudniejszych i droższych do przekonania).

W sekcji 4.6 przedstawiono badania nad oddziaływaniem na procesy rozprzestrzeniania informacji w sieciach społecznych poprzez kierowanie kampanii w określone grupy użytkowników w sieci. W szczególności, skoncentrowano się na sieciach o węzłach opisanych wieloma atrybutami.

2. Główny cel rozprawy

Głównym celem prezentowanej rozprawy doktorskiej jest opracowanie i weryfikacja algorytmów heterogenicznego oddziaływania na procesy propagacji informacji w sieciach złożonych z udziałem złożonych rankingów dynamicznych, mechanizmów topologicznych, z uwzględnieniem wielokryterialnej oceny efektywności.

Teza rozprawy:

Oddziaływanie na procesy rozprzestrzeniania informacji w sieciach złożonych zróżnicowanymi metodami umożliwi zwiększenie zasięgu procesu i jego dynamiki oraz innych charakterystyk procesu zgodnych z preferencjami decydenta.

3. Cykl publikacji wchodzących w skład rozprawy

Jako osiągnięcie naukowe w dyscyplinie Informatyka techniczna i telekomunikacja wskazuję cykl dziesięciu powiązanych tematycznie publikacji pt. **Heterogeniczne oddziaływanie na procesy rozprzestrzeniania informacji w sieciach społecznych**. Cykl ten obejmuje cztery artykuły opublikowane w czasopismach z otwartym dostępem, pięć artykułów wydanych w recenzowanych materiałach konferencyjnych oraz jeden rozdział monografii. Powiązania pomiędzy publikacjami zostały przedstawione na rysunku 3.1.

W skład cyklu publikacji wchodzi następujące prace:¹

- A1. Karczmarczyk, A.**, Jankowski, J., Wątróbski, J. (2018). Multi-criteria decision support for planning and evaluation of performance of viral marketing campaigns in social networks. PloS one, 13(12), e0209372.

Liczba cytowań: 34

Impact Factor: 2.776

Liczba punktów ministerialnych: 100

Udział w artykule: 60%

Wkład: Opracowanie koncepcji i założeń, przeprowadzenie badań i opracowanie wyników, wizualizacja, opracowanie tekstu.

- A2. Wątróbski, J.**, Jankowski, J., Ziemia, P., **Karczmarczyk, A.**, Ziolo, M. (2019). Generalised framework for multi-criteria method selection. Omega, 86, 107-124.

Liczba cytowań: 170

Impact Factor: 5.341

Liczba punktów ministerialnych: 140

Udział w artykule: 5%

Wkład: Opracowanie tekstu powiązane z analizą i porównaniem metod MCDA, udział w implementacji bazy reguł i systemu ekspertowego.

- A3. Karczmarczyk, A.**, Jankowski, J., Wątróbski, J. (2019, September). Multi-criteria approach to viral marketing campaign planning in social networks, based on real networks, network samples and synthetic networks. In 2019 Federated Conference on Computer Science and Information Systems (FedCSIS) (pp. 663-673). IEEE.

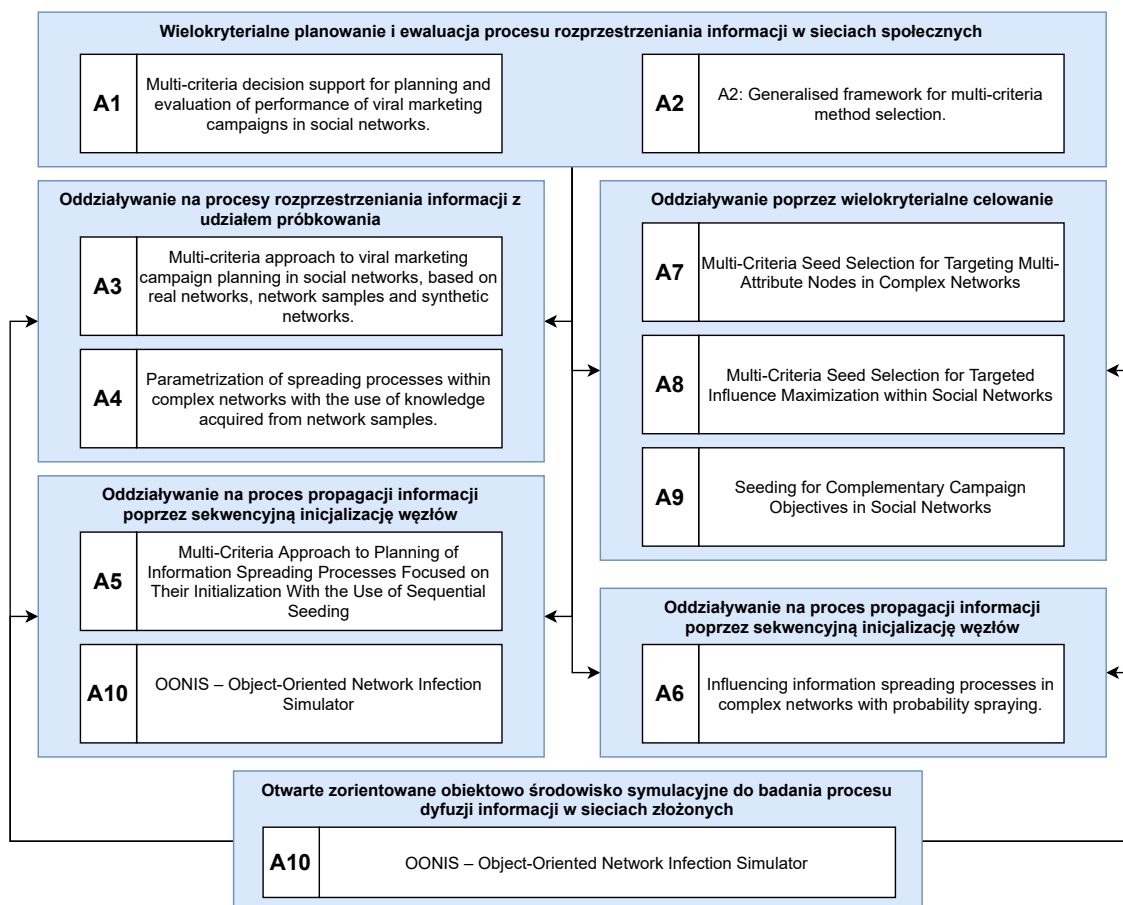
Liczba cytowań: 1

Indeksacja w WoS, Scopus

Udział w artykule: 70%

Wkład: Opracowanie koncepcji i założeń, przeprowadzenie badań i opracowanie wyników, wizualizacja, opracowanie tekstu.

¹Informacje o cytowaniach pochodzą z Google Scholar, stan na dzień 4.05.2021 r.



Rysunek 3.1. Wizualizacja powiązań pomiędzy poszczególnymi publikacjami A1–A10

A4. Karczmarczyk, A., Jankowski, J., Wątróbski, J. (2019). Parametrization of spreading processes within complex networks with the use of knowledge acquired from network samples. *Procedia Computer Science*, 159, 2279-2293.

Liczba cytowań: 1

Liczba punktów ministerialnych: 70

Udział w artykule: 70%

Wkład: Opracowanie koncepcji i założeń, przeprowadzenie badań i opracowanie wyników, wizualizacja, opracowanie tekstu.

A5. Karczmarczyk, A., Wątróbski, J., Jankowski, J. (2019). Multi-Criteria Approach to Planning of Information Spreading Processes Focused on Their Initialization With the Use of Sequential Seeding. In *Information Technology for Management: Current Research and Future Directions* (pp. 116-134). Springer, Cham.

Liczba cytowań: 1

Indeksacja w WoS, Scopus; rozdział w monografii

Udział w artykule: 70%

Wkład: Opracowanie koncepcji i założeń, przeprowadzenie badań i opracowanie wyników, wizualizacja, opracowanie tekstu.

A6. Karczmarczyk, A., Bortko, K., Bartków, P., Pazura, P., Jankowski, J. (2018,

August). Influencing information spreading processes in complex networks with probability spraying. In 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (pp. 1038-1046). IEEE.

Liczba cytowań: 1

Liczba punktów ministerialnych: 15

Udział w artykule: 50%

Wkład: Opracowanie koncepcji i założeń, opracowanie algorytmów, przeprowadzenie badań, opracowanie tekstu.

- A7. Karczmarczyk, A., Jankowski, J., Wątróbski, J. (2021).** Multi-Criteria Seed Selection for Targeting Multi-Attribute Nodes in Complex Networks. *Symmetry*, 13(4), 731.

Impact Factor za rok 2020: 2.645

Liczba punktów ministerialnych: 70

Udział w artykule: 65%

Wkład: Opracowanie koncepcji i założeń, przeprowadzenie badań i opracowanie wyników, wizualizacja, opracowanie tekstu.

- A8. Karczmarczyk, A., Jankowski, J., Wątróbski, J. (2021).** Multi-Criteria Seed Selection for Targeted Influence Maximization within Social Networks – in proceedings of International Conference on Computational Science: ICCS 2021

Publikacja zaakceptowana, w druku

Liczba punktów ministerialnych: 140

Udział w artykule: 65%

Wkład: Opracowanie koncepcji i założeń, przeprowadzenie badań i opracowanie wyników, wizualizacja, opracowanie tekstu.

- A9. Karczmarczyk, A., Wątróbski, J., Jankowski, J. (2021).** Seeding for Complementary Campaign Objectives in Social Networks - in proceedings of The Americas Conference on Information Systems: AMCIS 2021

Publikacja zaakceptowana, w druku

Liczba punktów ministerialnych: 140

Udział w artykule: 70%

Wkład: Opracowanie koncepcji i założeń, przeprowadzenie badań i opracowanie wyników, wizualizacja, opracowanie tekstu.

- A10. Karczmarczyk, A., Jankowski, J., Wątróbski, J. (2021).** OONIS—Object-Oriented Network Infection Simulator. *SoftwareX*, 14, 100675.

Liczba punktów ministerialnych: 200

Udział w artykule: 80%

Wkład: Opracowanie koncepcji i założeń, projektowanie i implementacja, przeprowadzenie badań i opracowanie wyników, wizualizacja, opracowanie tekstu.

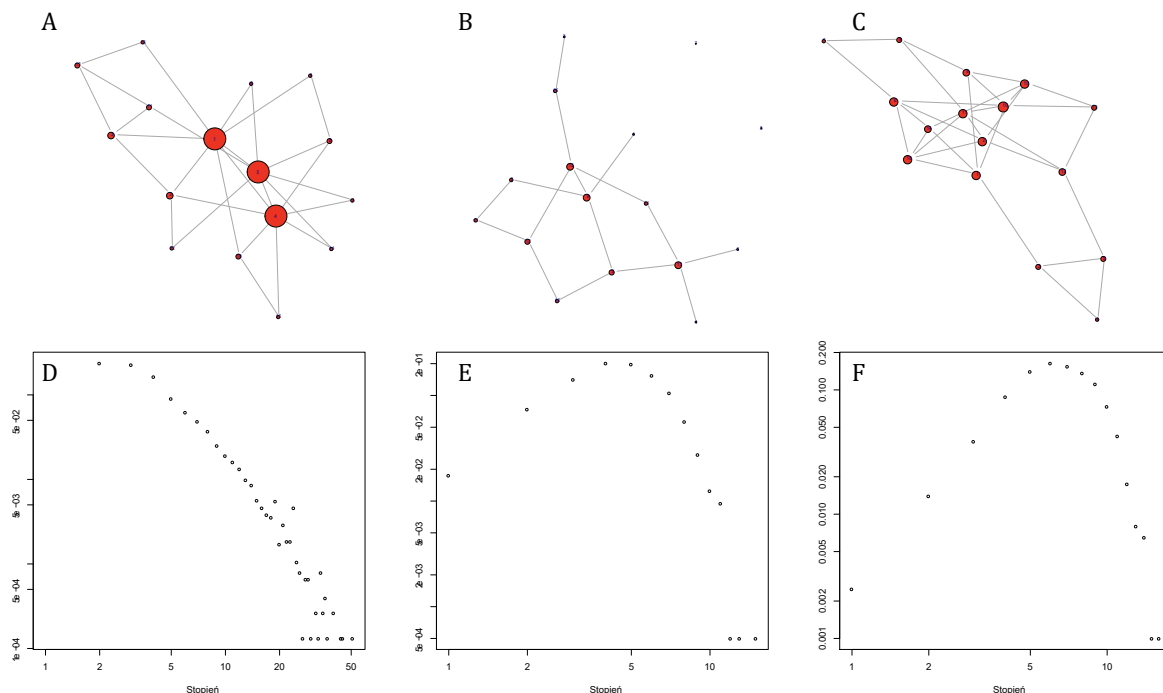
4. Heterogeniczne oddziaływanie na procesy rozprzestrzeniania informacji w sieciach społecznych

4.1. Podstawowe definicje

Działania prowadzone w sieciach społecznych mają często na celu motywowanie użytkowników do przekazywania informacji oraz treści cyfrowych znajomym oraz innym kontaktom w strukturach sieciowych. W związku z interdyscyplinarnością tego podejścia, badania prowadzone w tej dziedzinie angażują socjologów, fizyków, informatyków czy marketerów i obejmują szeroki wachlarz podejść i celów badawczych [3, 7]. Metodologiczne podstawy struktur sieciowych ewoluowały jednocześnie, ale oddzielnie, w różnych dyscyplinach [21].

Sieć społeczną G można zdefiniować jako zbiór węzłów (ang. nodes, vertices) $V(G)$ połączonych ze sobą za pośrednictwem zbioru krawędzi (ang. edges) $E(G)$. Sieć taką można opisać za pomocą notacji matematycznej: $G(V, E)$. Ścieżka (ang. path) w grafie G jest zbiorem krawędzi $\{\{v_1, v_2\}, \{v_2, v_3\}, \dots, \{v_{n-1}, v_n\}\}$, gdzie koniec krawędzi $\{v_i, v_{i+1}\}$ stanowi początek krawędzi $\{v_{i+1}, v_{i+2}\}$ dla każdego $i = 0, \dots, n - 2$, gdzie każdy węzeł i krawędź są unikalne. Długość ścieżki to liczba krawędzi wchodzących w skład tej ścieżki. Odległość $d(i, j)$ od węzła i do węzła j to długość najkrótszej ścieżki od i do j .

Sposoby w jaki modelowane jest rozprzestrzenianie informacji w sieciach społecznych obejmują kilka podstawowych kierunków. Wyróżnić tu trzeba model kaskadowy (ang. independent cascade model [16]), model progowy (ang. linear threshold model [17]), jak również modele epidemiczne [15]. W rozprawie podczas badań oparto się na modelu kaskadowym, w którym informacja rozprzestrzenia się w sieci poprzez kaskady. Każdy węzeł w sieci znajduje się w jednym z dwóch stanów: aktywnym lub nieaktywnym. Początkowo wszystkie węzły są w stanie nieaktywnym. Na początku symulacji inicjalizuje się proces poprzez przekazanie informacji do małej części (tzw. ang. seeding fraction) węzłów (ang. seeds). W momencie otrzymania informacji, stają się one aktywne. W każdym dyskretnym kroku symulacji, węzły aktywne starają się przekazać informację do węzłów nieaktywnych. O sukcesie lub porażce przekazania informacji stanowi prawdopodobieństwo propagacji (ang. propagation probability) cechujące krawędź sieci łączącą węzeł aktywny z węzłem nieaktywnym. Każda krawędź ma swoją własną wartość prawdopodobieństwa propagacji. W całości sieci określa się średnią wartość prawdopodobieństwa propagacji. Każdy aktywny węzeł ma tylko jedną szansę przekazać informację do węzła nieaktywnego. Węzły do których uda się przekazać informację, zostają aktywowane. Powyższy proces trwa tak długo, jak aktywowane są kolejne węzły, a kończy się gdy żaden nowy węzeł nie zostanie aktywowany. Liczba węzłów aktywowanych w sieci po zakończeniu procesu kaskadowego to tzw. zasięg kampanii (ang. coverage).



Rysunek 4.1. Grafy przedstawiające przykład 16-węzłowych sieci syntetycznych: A) BA, B) ER, C) WS; oraz rozkład stopni węzłów przykładowych 2000-węzłowych sieci syntetycznych: D) BA, E) ER, F) WS.

Większość strategii inicjalizacji kampanii w sieciach opiera się na doborze początkowych węzłów (ang. seeds) na bazie rankingów uzyskanych na podstawie miar centralności węzłów. Zakłada się tutaj, że im węzeł bardziej centralny, tym większy zasięg w sieci pozwoli osiągnąć. Jako najbardziej podstawowe miary centralności można wskazać centralność stopnia (ang. degree centrality), centralność bliskości (ang. closeness centrality), centralność pośrednictwa (ang. betweenness centrality) i centralność wektora własnego (ang. eigenvector centrality). Ponadto, stosowane są bardziej zaawansowane rozwiązania, jak podejście zachłanne [16].

Niestety, wiedza o sieciach społecznych, w których prowadzone mają być kampanie, często ogranicza się do kilku podstawowych cech. Podczas gdy zbieranie informacji o rzeczywistych sieciach jest trudne, można wykorzystać syntetyczne sieci oparte na modelach teoretycznych. Dodatkowym atutem sieci syntetycznych jest możliwość dostosowywania ich struktury w trakcie procesu ich generowania, co pozwala na głębszą analizę procesów zachodzących w złożonych sieciach. W badaniach symulacyjnych często wykorzystuje się sieci oparte na modelu bezskalowym (ang. free-scale) zaproponowanym przez Barabasi-Alberta (BA) [22], modelu małego świata zaproponowanego przez Watts-Strogatza (WS) [23] oraz modelu grafu losowego wprowadzonego przez zespół Erdos-Renyi (ER) [24].

Charakterystyki modeli teoretycznych BA i WS są zbliżone do rzeczywistych systemów. Model Barabasi-Alberta powstał w 1999 roku, w wyniku badania ówczesnej struktury sieci WWW. Budowa sieci BA opiera się na dwóch komplementarnych mechanizmach: rozwoju sieci oraz mechanizmie preferencyjnego dołączania. Model BA jest podobny do wielu systemów naturalnych i stworzonych przez człowieka, takich jak Internet, WWW,

sieci cytowań czy sieci społeczne. W systemach takich kilka wybranych węzłów (ang. hubs) ma wysoki stopień w porównaniu z pozostałymi węzłami sieci. Na rys. 4.1 (A) przedstawiono przykład sieci BA, a wykres na rys. 4.1 (D) przedstawia rozkład stopni węzłów przykładowego modelu.

Model sieci ER został po raz pierwszy opisany w 1959 r. Podczas jego konstrukcji, najpierw zdefiniowana jest liczba N węzłów, a następnie ze wszystkich $\binom{N}{2}$ par węzłów, wybiera się losowe E par, pomiędzy którymi tworzone są krawędzie. Przykładowy model ER i rozkład stopni przykładowego modelu ER przedstawiono odpowiednio na rys. 4.1 (B) i rys. 4.1 (E).

Model ER oferuje uniwersalny model o wielu zastosowaniach. Może on jednak być nieodpowiedni do modelowania niektórych zjawisk rzeczywistych, ponieważ nie generuje lokalnych klastrów węzłów. Aby rozwiązać ten problem, w 1998 roku powstał model Watts-Strogatza. Model WS uwzględnia klastrowanie, ale zachowuje krótkie średnie długości ścieżek z modelu ER. Rys. 4.1 (C) przedstawia przykład sieci WS, a wykres na rys. 4.1 (F) przedstawia rozkład stopni węzłów przykładowego modelu WS.

4.2. Wielokryterialne planowanie i ewaluacja procesów rozprzestrzeniania informacji w sieciach społecznych [A1, A2]¹

W przypadku procesów rozprzestrzeniania informacji w serwisach społecznościowych inicjator procesu może być zainteresowany nie tylko maksymalizacją zasięgu kampanii, ale także wpływaniem na jej dynamikę oraz utrzymaniem kosztów na zadanym poziomie. W związku z powyższym, planowanie takich kampanii jest problemem wielokryterialnym, który można przedstawić jako: (4.1) [25]:

$$\max \{c_1(a), c_2(a), \dots, c_k(a) | a \in A\}, \quad (4.1)$$

gdzie A oznacza zbiór możliwych strategii $\{a_1, a_2, \dots, a_n\}$, a $\{c_1(\cdot), c_2(\cdot), \dots, c_k(\cdot)\}$ oznacza zbiór kryteriów wykorzystywanych do ewaluacji tych strategii. Część kryteriów może być maksymalizowana, a część minimalizowana. Kryterialne wyniki każdej strategii w odniesieniu do każdego kryterium można wyrazić w postaci tabeli wyników. Intuicyjnie oczekuje się od decydenta określenia strategii optymalizującej wszystkie kryteria. Jednak zwykle nie ma alternatywy, która optymalizowałaby wszystkie kryteria jednocześnie.

Rozważmy przykładowy proces, dla którego przygotowano wiele alternatywnych strategii. Strategie te charakteryzują się trzema kryteriami: liczbą węzłów zasiewowych (ang. seeding fraction), prawdopodobieństwem propagacji informacji (ang. propagation probability) i potencjalnym zasięgiem, jaki można uzyskać (ang. coverage). W klasycznym homogenicznym podejściu maksymalizowałoby się zasięg. Zasięg jest bardzo ważnym kryterium, jednak generalnie strategia zapewniająca 100% zasięgu nie zawsze jest wybierana, ponieważ wymagałaby zainfekowania ogromnej liczby początkowych węzłów w

¹Sekcja powstała na podstawie opublikowanego artykułu **A1** o współczynniku IF 2.776 i 33 cytowaniach oraz **A2** opublikowanego w renomowanym czasopiśmie Omega wydawnictwa Elsevier o współczynniku IF 5.341 i 167 cytowaniach.

sieci lub zapewnienia wielu bodźców w celu zwiększenia prawdopodobieństwa propagacji pomiędzy węzłami pośrednimi w sieci. Z drugiej strony, jeśli zostanie wybrana strategia z minimalną liczbą inicjalnych węzłów i minimalnym prawdopodobieństwem propagacji, nie można oczekiwać, że obejmie ona całą sieć. Dlatego należy wybrać spośród strategii rozwiązanie kompromisowe. W proponowanym heterogenicznym podejściu, oprócz maksymalizacji zasięgu, rozważa się także inne kryteria.

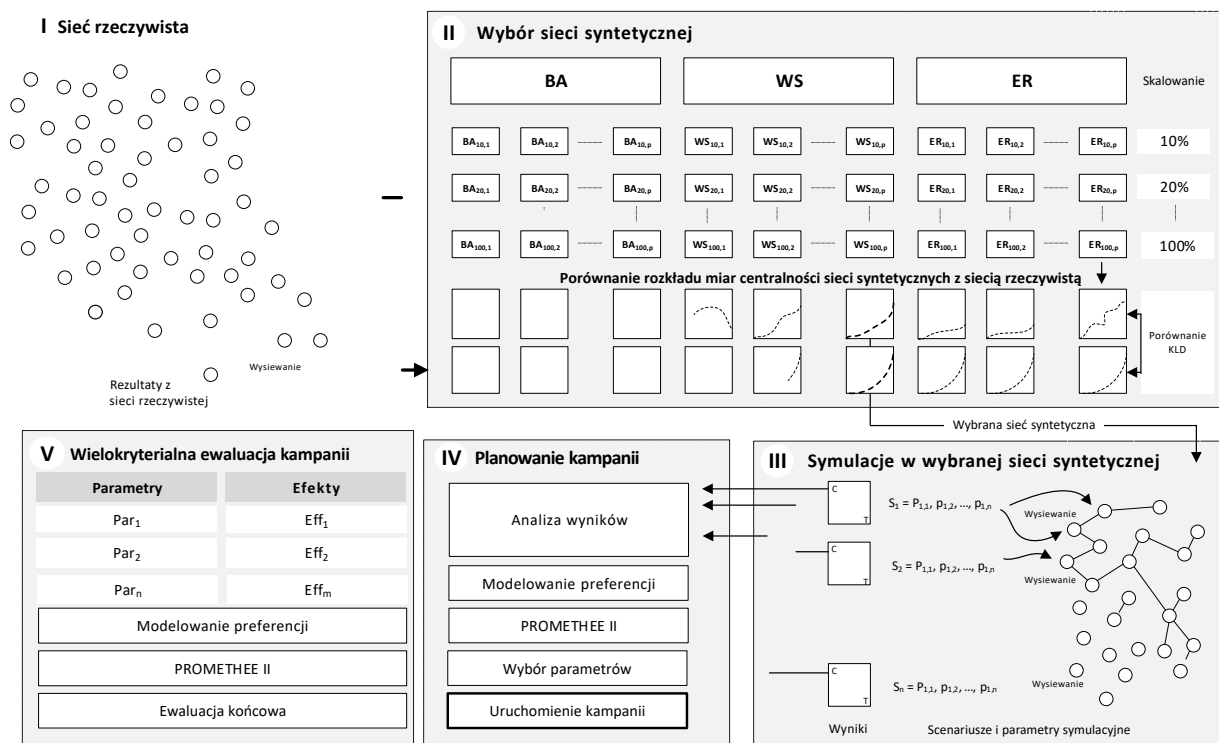
Należy zwrócić uwagę, że rozwiązanie problemu wielokryterialnego zależy nie tylko od wartości kryteriów każdej alternatywy, ale także od inicjatora procesu. Nie istnieje strategia absolutnie najlepsza dla wszystkich kampanii, natomiast najlepsza strategia kompromisowa zależy od preferencji decydenta.

Z przedstawionym wzorem (4.1) problemem decyzyjnym o charakterze wielokryterialnym można powiązać trzy relacje naturalnej dominacji: obojętność, preferencję i nieporównywalność. Rozważmy dwie alternatywy a i b . Jeśli dla każdego kryterium c_i , a jest tak samo dobre jak b , to obie strategie są obojętne (aIb). Jeśli dla każdego kryterium c_j , a jest tak samo dobre lub równe b i istnieje przynajmniej jedno kryterium c_k , dla którego a jest lepsze niż b , to a jest preferowane od b (aPb). Wreszcie, jeśli istnieje kryterium c_m , dla którego a jest lepsze niż b , ale istnieje również kryterium c_n , dla którego b jest lepsze niż a , to obie strategie są nieporównywalne (aRb).

Strategie, które są najlepsze według każdego kryterium mogą nie występować. W związku z tym, zazwyczaj większość strategii jest nieporównywalna bez dodatkowych informacji od decydenta. Informacje te mogą obejmować między innymi wagi wyrażające względne znaczenie każdego kryterium lub preferencje pomiędzy porównywanymi parami strategii, gdy każde kryterium jest rozpatrywane osobno [25]. Metody wielokryterialnego wspomagania decyzji (ang. Multi-Criteria Decision Analysis, MCDA) pomagają zredukować liczbę nieporównywalności (R) w grafie decyzyjnym między rozważanymi strategiami kampanii.

Metody MCDA generalnie dzieli się na dwie rodziny. Pierwsza, tzw. amerykańska, obejmuje metody agregujące wszystkie kryteria do pojedynczego kryterium – funkcji użyteczności. Druga, tzw. europejska, bazuje na relacjach przewyższania (ang. outranking) poprzez wzmacnianie relacji dominacji między alternatywami wszędzie tam, gdzie jest to możliwe. Przy takim podejściu nie wszystkie nieporównywalności są eliminowane, jednakże możliwy jest rzetelny wybór najlepszej alternatywy. Istnieją także metody łączące cechy obydwu rodzin. Szczegółowe zestawienie i ramy decyzyjne do wyboru metody MCDA dopasowanej do problemu decyzyjnego opisane zostały w publikacji **A2**. Na rysunku 4.2 przedstawiono schemat poglądowy proponowanego podejścia.

Na podstawie ram decyzyjnych opracowanych w **A2** i bazy reguł opublikowanej w [26] zdecydowano o wykorzystaniu metody PROMETHEE II w dalszej części tej sekcji. PROMETHEE to rodzina metod MCDA, które wykorzystują porównania alternatyw parami i przepływy przewyższania w celu stworzenia rankingu najlepszych wariantów decyzyjnych. Wagi wyrażające względne znaczenie każdego kryterium zostają określone przez decydenta. To skomplikowany proces oparty na priorytetach i spostrzeżeniach decydenta. Rzeczywiste wartości wag kryteriów mogą być dowolnie wybierane przez zamawiającego kampanię. Na szczęście metody wielokryterialnego wspomagania decyzji dostarczają narzędzi, takich jak analizy wrażliwości i odporności (ang. sensitivity and robustness analyses), które pozwalają zweryfikować wpływ wybranych wartości na otrzymane rankingi



Rysunek 4.2. Schemat poglądowy podejścia planowania i ewaluacji procesów rozprzestrzeniania informacji w sieciach społecznych.

i sekwencyjnie dostosowywać wybrane wagi. Szczegółowy opis metody PROMETHEE II można znaleźć w publikacji [25].

Wybór najlepszej strategii kampanii wirusowej w sieciach społecznych to złożony problem decyzyjny oparty na wielu kryteriach. Przeprowadzenie symulacji w rzeczywistej sieci jest najczęściej czasochłonne, a czasami niemożliwe. W związku z powyższym, w prezentowanym podejściu proponuje się prowadzenie procesu planowania na modelu syntetycznym, który ma podobne właściwości do docelowej sieci rzeczywistej, ale pozwala na przeprowadzenie wielu symulacji. Wynikiem tych symulacji jest zestaw danych zasilających tabelę wydajności (ang. performance table) wszystkich analizowanych strategii pod kątem kryteriów oceny. Tabela ta stanowi dane wejściowe do procesu oceny strategii.

W celu uzyskania modelu syntetycznego jak najbardziej przypominającego docelową sieć rzeczywistą, proponuje się wygenerować szereg sieci BA, ER i WS o zróżnicowanych parametrach oraz liczbie węzłów równą 10%, 20%, ..., 100% sieci rzeczywistej. Następnie, wskaźnik KLD (ang. Kullback-Leibler Divergence, [27]) może zostać użyty do weryfikacji która z wygenerowanych sieci jest najbliższa sieci rzeczywistej. Dodatkowo rozważone mogą być kryteria wywierające wpływ na złożoność obliczeniową symulacji, takie jak liczba węzłów i krawędzi. W zależności od potrzeb decydenta do procesu decyzyjnego można również dodać dodatkowe kryteria. Gdy tabela wydajności dla wszystkich sieci syntetycznych i wszystkich kryteriów zostanie już stworzona, należy wybrać sieć najbardziej preferowaną z wykorzystaniem metod MCDA.

Kolejnym elementem proponowanego podejścia jest proces strukturyzacji modelu decyzyjnego. W trakcie tego procesu dobierane są kryteria decyzyjne oceny możliwych strategii kampanii. W proponowanym podejściu kryteria można podzielić na dwie grupy. Pierwsza

grupa zawiera kryteria wejściowe do budowy strategii – Par1, Par2, ..., Parm. Druga grupa zawiera kryteria oceny efektywności strategii Eff1, Eff2, ..., Effn, których wartości opierają się na osiągniętych efektach, a ich wartości można uzyskać z symulacji każdej strategii na wybranej sieci syntetycznej. Niemniej jednak proponowane podejście zakłada swobodę decydenta w doborze kryteriów decyzyjnych i grupowaniu ich w klastry w zależności od wymagań inicjatora procesu.

Gdy wszystkie kryteria są już dobrane, a model decyzyjny został ustrukturyzowany, na wybranej sieci syntetycznej przeprowadza się szereg symulacji dla każdej potencjalnej strategii. W podejściu prezentowanym w publikacji A1 zastosowano model IC (ang. independent cascade, [16]). Wybór został podyktowany względnie niewielką liczbą węzłów inicjalnych (ang. seeds) wymaganych do wzbudzenia propagacji informacji, co może być istotne w niewielkich sieciach. W przypadku modelu LT (ang. linear threshold) mała liczba węzłów początkowych nie przyniosłaby efektu.

W wyniku przeprowadzenia symulacji otrzymuje się macierz wydajności kryterialnej wszystkich ewaluowanych strategii. Macierz tę wykorzystuje się następnie w metodzie PROMETHEE II do przeprowadzenia wielokryterialnej oceny strategii. Analiza ta obejmuje w szczególności:

- wygenerowanie kompletnego rankingu strategii, w oparciu o różne funkcje preferencji;
- wykorzystanie płaszczyzny GAIA (ang. Geometrical Analysis for Interactive Assistance, [25]) w celu weryfikacji jak poszczególne kryteria oddziałują na dobór strategii;
- przeprowadzenie analizy wrażliwości w celu weryfikacji stabilności uzyskanych rankingów dla wiodących strategii.

Należy zauważyć, że podczas analizy krok modelowania preferencji jest powtarzany wielokrotnie. Początkowe wagi preferencji kryteriów można następnie modyfikować w celu zweryfikowania odporności uzyskanego rozwiązania problemu wyboru strategii. Ostatecznie analityk rekomenduje, jaką strategię, czyli zbiór parametrów, zastosować do realizacji procesu w docelowej sieci rzeczywistej.

W celu weryfikacji proponowanego podejścia, zaproponowany został zestaw pięciu parametrów, z czego trzy pierwsze to kryteria dotyczące parametrów uruchomienia procesu, a pozostałe dwa to kryteria wydajnościowe:

- Par1 Liczba węzłów (użytkowników) inicjalnych w kampanii (ang. seeds) wyrażona jako ułamek całkowitej liczby węzłów w sieci (ang. seeding fraction).
- Par2 Motywacja do przekazywania treści - wyrażona jako średnie prawdopodobieństwo propagacji (ang. propagation probability).
- Par3 Miara centralności wykorzystywana do wyboru początkowych użytkowników sieci w kampanii, takie jak centralność stopnia (ang. degree centrality), centralność wektora własnego (ang. eigenvector centrality) itp. Z doбором węzłów początkowych może wiązać się koszt. Przykładowo węzły o wysokiej wartości centralności wektora własnego uznawane są za użytkowników wpływowych w sieci społecznej, więc koszt ich pozyskania może być wyższy.
- Eff4 Czas wymagany na dotarcie do założonej liczby użytkowników sieci. W przypadku modeli syntetycznych wartość ta reprezentowana jest przez liczbę kroków symulacji.

Eff5 Uzyskany zasięg w sieci, czyli liczba użytkowników sieci, do której dotarto przy założonej strategii kampanii względem całkowitej liczby użytkowników sieci.

Badania empiryczne proponowanego podejścia do planowania procesów rozprzestrzeniania zostały przedstawione w sekcji 3 publikacji **A1**. Badania te oparte zostały na sieci rzeczywistej [28] o 7610 węzłach i 15751 krawędziach, o średnim stopniu węzłów wynoszącym 4.14. W ramach badań empirycznych wygenerowano 150 sieci syntetycznych (po 50 każdego typu BA, ER i WS) zbudowanych z 10%, 20%, ..., 100% węzłów w stosunku do sieci rzeczywistej. Z wykorzystaniem miary KLD oraz analizy wielokryterialnej wybrano do dalszej analizy sieć BA o 761 węzłach i 3034 krawędziach. Następnie przeprowadzono łącznie 4000 symulacji dla 400 zestawów kryteriów Par1-Par3. W efekcie uzyskano wartości Eff4 i Eff5 dla wszystkich 400 strategii. Uzyskane wartości posłużyły do przeprowadzenia analizy wielokryterialnej poszczególnych kampanii z wykorzystaniem metody PROMETHEE II z różnymi funkcjami preferencji. Analiza GAIA pozwoliła określić wpływ poszczególnych kryteriów na wybór najlepszej dla decydenta strategii, a analiza wrażliwości pozwoliła określić stabilność rankingu.

Należy zwrócić uwagę, że zaproponowane podejście może być również wykorzystane do monitorowania wyników przeprowadzonej kampanii, a także do przeprowadzenia wielokryterialnej oceny strategii kampanii w sieci rzeczywistej. Empiryczne badanie proponowanego podejścia do ewaluacji kampanii rozprzestrzeniania informacji w rzeczywistej sieci [28] zostało przedstawione w sekcji 4 publikacji **A1**.

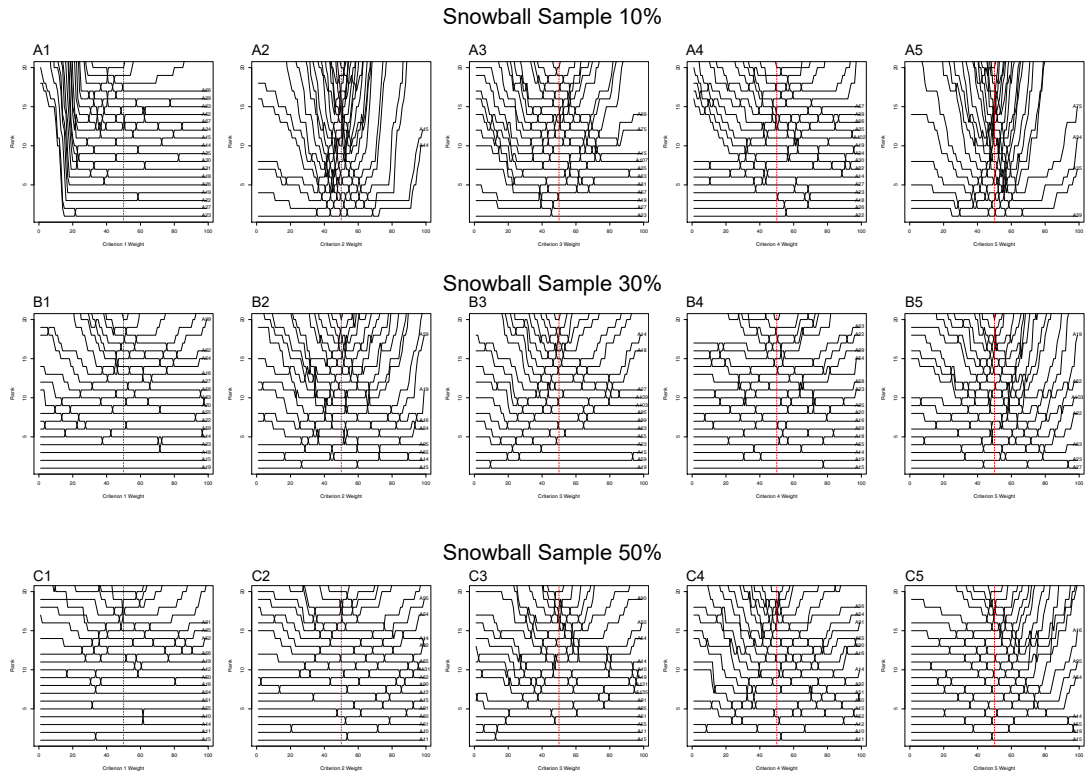
4.3. Oddziaływanie na inicjalizację procesów rozprzestrzeniania z udziałem próbkowania i model doboru wielkości próbek w ujęciu wielokryterialnym [A3, A4]²

W sekcji 4.2 przedstawiono podejście do planowania procesów propagacji z wykorzystaniem sieci syntetycznych opartych na modelach teoretycznych. Zastosowanie sieci syntetycznych o odpowiednio zmniejszonej liczbie węzłów pozwoliło przeprowadzić zaawansowane planowanie kampanii przy zredukowanej złożoności obliczeniowej symulacji. W celu jak najlepszego dopasowania sieci syntetycznej do docelowej sieci rzeczywistej zastosowano miarę KLD oraz analizę wielokryterialną. Powstała jednakże obawa, że modele teoretyczne w niektórych przypadkach mogą nie być wystarczająco dobrze dopasowane do sieci rzeczywistych. W związku z powyższym, w tej sekcji przedstawiono podejście do planowania i uruchamiania kampanii rozprzestrzeniania informacji w sieciach społecznych z wykorzystaniem próbek sieci rzeczywistej.

W publikacji **A3** rozszerzono podejście proponowane w sekcji 4.2 o wykorzystanie próbek sieci rzeczywistej. Badanie empiryczne oparto na sieci rzeczywistej [29], przedstawiającej części topologii sieci Gnutella. Odwzorowana sieć składa się z 8846 węzłów i 31839 krawędzi. Węzły reprezentują hosty w topologii sieci Gnutella, a krawędzie reprezentują połączenia pomiędzy hostami Gnutella w jednej z migawek sieci zebranych w sierpniu 2002 r. Średni stopień węzłów w sieci rzeczywistej wynosi 7.1985.

W badaniu wykorzystano taki sam zestaw kryteriów ewaluacyjnych strategii Par1-Eff5 jak w sekcji 4.2. Dodatkowo, w tym przypadku do ewaluacji poszczególnych 400 stra-

²Sekcja powstała w oparciu o publikacje **A3** i **A4**.



Rysunek 4.3. Analiza wrażliwości rankingu dla 20 najlepszych strategii z ewaluacji metodą TOPSIS na próbkach sieci rzeczywistej [29]. A1-A5 – sieć 10%, B1-B5 – sieć 30%, C1-C5 – sieć 50%.

tegi zamiast PROMETHEE II wykorzystano metodę TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution, [30]). Metoda TOPSIS jest reprezentantem amerykańskiej szkoły MCDA, która przekształca wszystkie kryteria problemu decyzyjnego w pojedynczą wartość punktową. W przypadku metody TOPSIS, w oparciu o wartości ocenianych kryteriów, tworzone są idealne i antyidealne strategie, czyli takie, które są najlepsze pod względem każdego z kryteriów i takie, które są najgorsze pod względem każdego z kryteriów. Następnie każdej ocenianej strategii przypisuje się punkty obliczane jako relatywna odległość między ocenianą strategią a zarówno rozwiązaniem idealnym, jak i antyidealnym. Gdy wszystkim strategiom przypisane zostaną punkty, wybierana jest taka strategia, która jest najbliższa strategii idealnej, ale jednocześnie jest jej tak daleko jak to możliwe od strategii antyidealnej pod względem wartości poszczególnych kryteriów.

Jedną z zalet metody TOPSIS jest to, że pozwala ona na zbudowanie idealnego modelu referencyjnego dla zadanego problemu ewaluacyjnego. W przypadku wykorzystanej w badaniu sieci rzeczywistej, strategia idealna opierałaby się na miarze centralności stopnia (ang. degree centrality) do wyboru inicjalnych użytkowników sieci (ang. seeds). Dodatkowo, kampania uruchomiona zostałaby przez przekazanie informacji do 1% użytkowników. Ponadto, zachęty do propagowania treści zorganizowane byłyby w taki sposób, żeby osiągać średnie prawdopodobieństwo propagacji na poziomie 1%. Z drugiej strony, przy takich parametrach sieci, strategia idealna skutkowałaby wynikami pokrycia sieci na poziomie 97.22% przy średniej długości 19.6 iteracji. Należy zauważyć, że choć taka strategia byłaby idealna, to jednak jest tylko modelem referencyjnym i nie istnieje w rzeczywistości.

Tablica 4.1. Macierz korelacji pomiędzy rankingami strategii obliczonymi na sieci rzeczywistej i jej próbkach.

	Rzeczywista	Próbka 10%	Próbka 30%	Próbka 50%
Rzeczywista	x	0.4629	0.6837	0.9222
Próbka 10%	0.4629	x	0.8227	0.6159
Próbka 30%	0.6837	0.8227	x	0.8718
Próbka 50%	0.9222	0.6159	0.8718	x

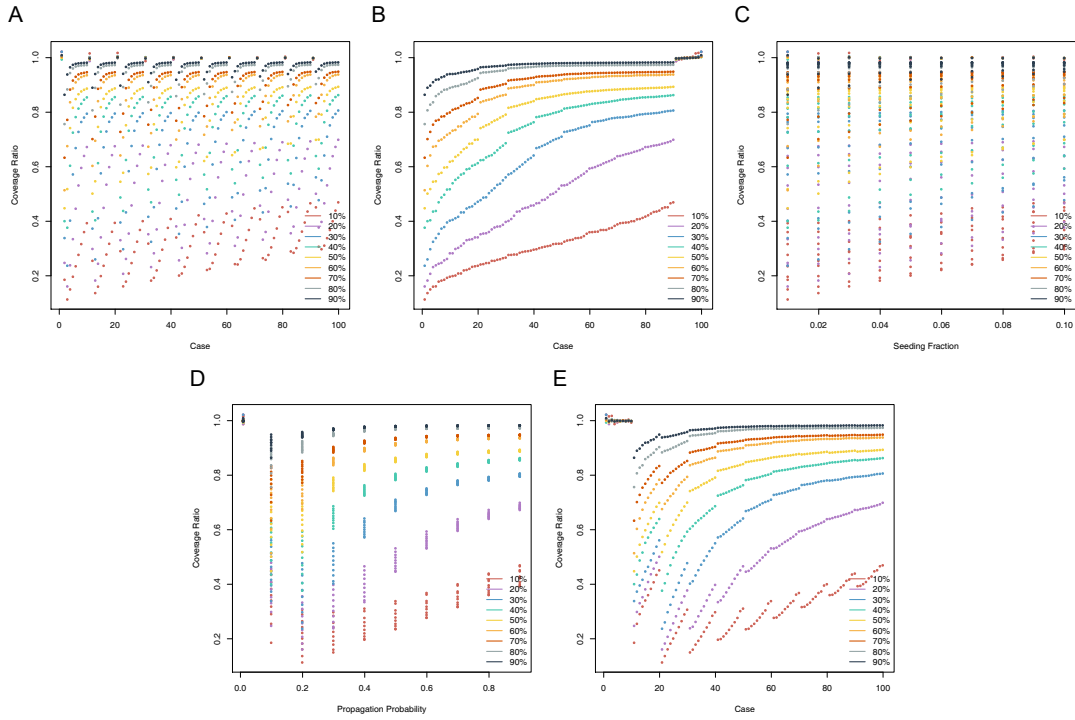
W badaniu empirycznym planowanie procesu na sieci rzeczywistej w oparciu o jej próbki przeprowadzono z wykorzystaniem trzech próbek sieci różnych wielkości: 10%, 30% i 50% węzłów oryginalnej sieci. Próbkę uzyskano z wykorzystaniem metody snowball sampling (zastosowano implementację *snowball.sampling* języka R z biblioteki *netdep* [31]). Następnie z wykorzystaniem metody TOPSIS dokonano ewaluacji wszystkich strategii na próbkach sieci rzeczywistej (zob. sekcja IV.D publikacji **A3**). Ponadto, wykonano analizę wrażliwości rankingów strategii. Wynik tej analizy (dla 20 najlepszych strategii) przedstawiony jest na rys. 4.3. Zwrócić należy uwagę, że o ile dla próbki 50% i 30% rankingi są dosyć stabilne, to przy próbce 10% niewielkie wahania wagi kryteriów Par2-Eff5 spowodowałyby znaczne zmiany kolejności strategii w rankingu.

W tabeli 4.1 przedstawiono wartości współczynników korelacji pozycji poszczególnych strategii w rankingach uzyskanych dla sieci rzeczywistej oraz poszczególnych jej próbek. Należy zwrócić uwagę, że o ile dla próbki 50% korelacja wynosi 0.9222, co wskazuje na wysoką korelację pomiędzy rankingami dla sieci rzeczywistej i jej 50-procentowej próbki, to dla mniejszych próbek wartość współczynnika korelacji nie jest zadowalająca.

Wspomniana powyżej rozbieżność pomiędzy rankingami strategii dla sieci rzeczywistej i jej próbek różnych wielkości stanowiła podstawę opracowania publikacji **A4**. W publikacji tej przeprowadzono szereg symulacji dla sieci rzeczywistej Gnutella [29] oraz jej 10%, 20%, ..., 90% próbek. Dla każdej z tych 10 sieci przeprowadzono 1000 symulacji dla:

- 10 wartości liczby inicjalnych węzłów kampanii (ang. seeding fraction, SF): 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.10,
- 10 wartości średniego prawdopodobieństwa propagacji (ang. propagation probability, PP): 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.10,
- 10 uprzednio wylosowanych scenariuszy ustawień wag w sieci. Dla każdego węzła została wylosowana wartość. Jeśli wartość była mniejsza lub równa wartości PP podczas symulacji, informacja była przekazywana przez określony węzeł. Jeśli wartość była wyższa od wartości PP, informacja nie była propagowana.

Inicjalne węzły (ang. seeds) do uruchomienia procesu wybierane były na podstawie rankingu opartego na centralności stopnia (ang. degree centrality) poszczególnych węzłów. Szczegółowe analizy znajdują się w sekcji 4 publikacji **A4**. Porównanie wartości zasięgu dla każdej próbki względem sieci rzeczywistej przedstawione zostało na rysunku 4.4. Jak można zaobserwować, wyniki najbardziej zbliżone do sieci rzeczywistej uzyskano dla próbek o jak najwyższej liczbie węzłów.



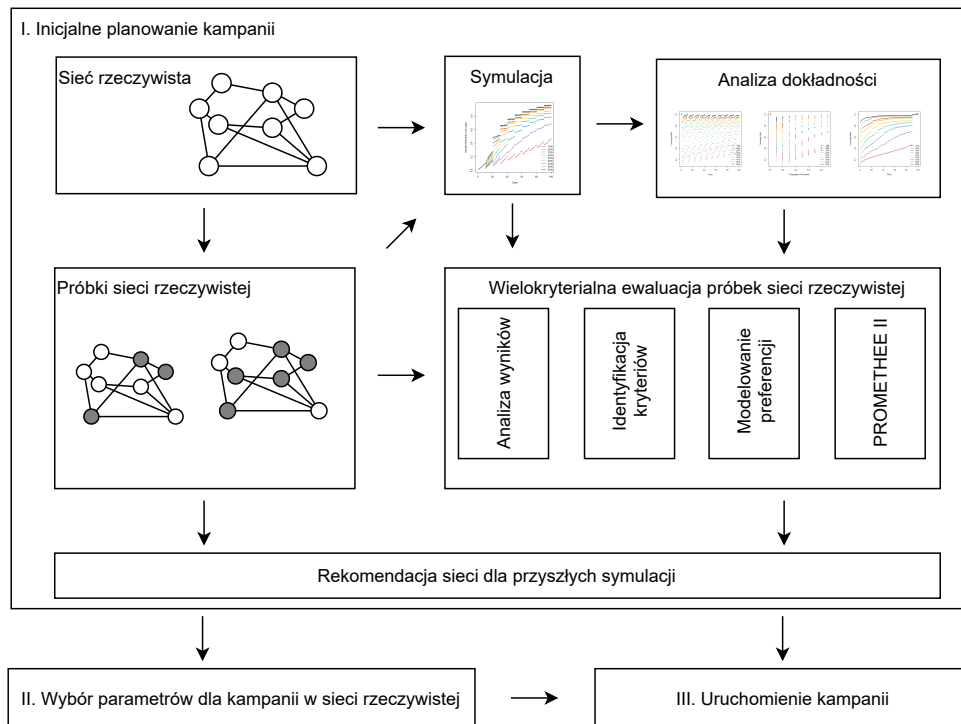
Rysunek 4.4. Stosunek zasięgu próbek do sieci rzeczywistej [29]. A: uporządkowane według przypadku symulacji, B: uporządkowane według stosunku zasięgu rosnąco, C: zgrupowane i uporządkowane według SF, D: zgrupowane i uporządkowane według PP, E: uporządkowane według rosnącej wartości zasięgu w sieci rzeczywistej.

W przypadku badanej sieci rzeczywistej, wybór 90% próbki sieci do planowania strategii rozprzestrzeniania informacji wydaje się najwłaściwszą opcją. Jednak w rzeczywistych zastosowaniach inicjator może zdecydować o rezygnacji z idealnej dokładności planowania kampanii, jeśli przyniosłoby to inne korzyści, które zrekompensowałyby utratę dokładności (kompensacja kryteriów). W publikacji **A4** zaproponowano wykorzystanie komponentu MCDA, aby ułatwić dobór wielkości próbki przy zmiennych preferencjach inicjatora procesu (Rys. 4.5).

Zaproponowano cztery kryteria do wyboru wielkości próbki sieci rzeczywistej, podzielone na dwie grupy – koszt i dokładność:

- C1 – grupa kosztów – wielkości próbki sieci, wyrażona jako stosunek liczby węzłów próbki do liczby wszystkich węzłów w sieci rzeczywistej;
- C2 – grupa kosztów – czas potrzebny na wygenerowanie próbki o określonej wielkości;
- C3 – grupa dokładności – różnica stosunku zasięgu pomiędzy próbką a siecią rzeczywistą od idealnego stosunku 1/1;
- C4 – grupa dokładności – różnica stosunku czasu trwania symulacji pomiędzy próbką a siecią rzeczywistą od idealnego stosunku 1/1.

Badanie empiryczne dla zaproponowanego modelu decyzyjnego przedstawiono w sekcji 4.3 publikacji **A4**. Na rysunku 4.6 przedstawiono analizę wizualną GAIA dla wyboru wielkości próbki sieci rzeczywistej przy równych wagach wszystkich czterech kryteriów. Można zwrócić uwagę, że dla zwykłej funkcji preferencji (ang. usual preference function) można



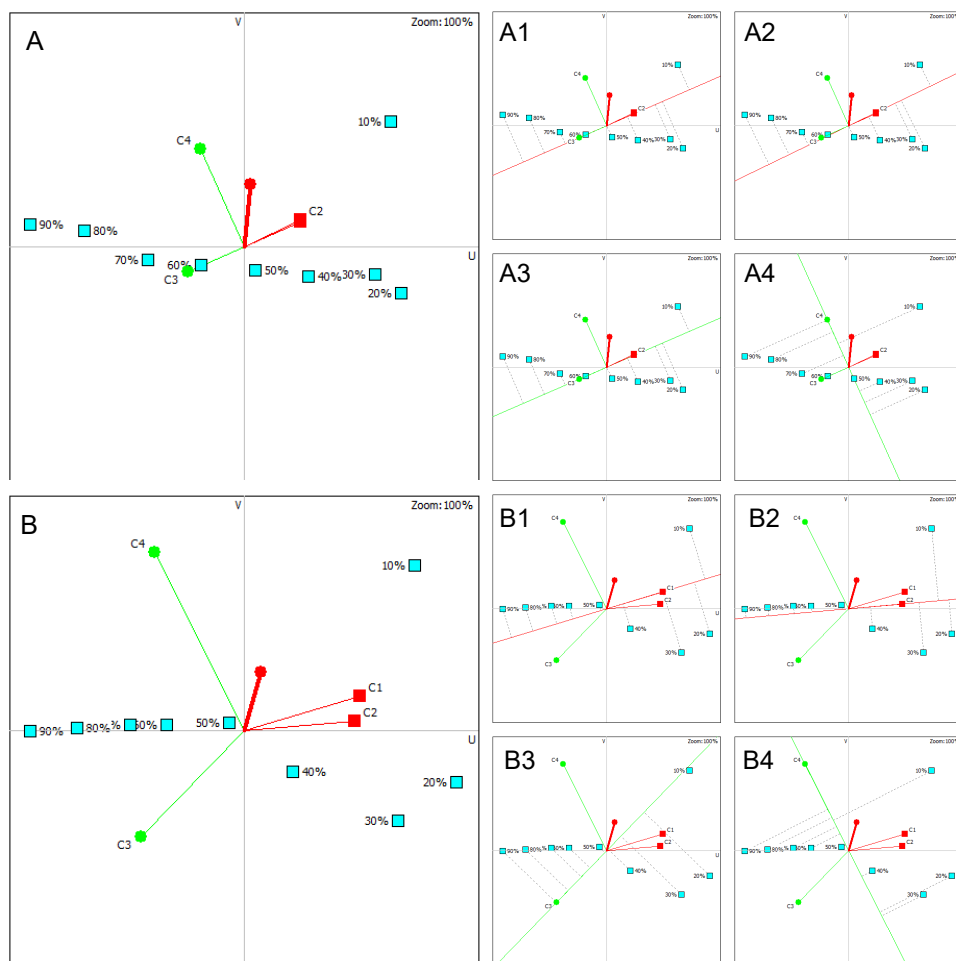
Rysunek 4.5. Schemat poglądowy proponowanego podejścia do wielokryterialnego doboru wielkości próbki przy zmiennych preferencjach inicjatora procesu.

wskazać zwycięską próbkę 10%, to wskazanie kolejności pozostałych wielkości próbek jest trudne – wiele z nich otrzymało taką samą ocenę w rankingu. Problem ten rozwiązany może zostać przez zastosowanie liniowej funkcji preferencji (ang. linear preference function) z uwzględnieniem wartości preferencji (ang. preference, aPb) i obojętności (ang. indifference, aIb) przy porównywaniu parami dowolnych dwóch wariantów. Na rysunku 4.6 (B) po zastosowaniu liniowej funkcji preferencji wyraźnie można wskazać, że drugim najlepszym wyborem byłoby użycie próbki 50%, a następnie kolejno 40%, 60% i 70%.

4.4. Oddziaływanie na proces propagacji informacji poprzez wielokryterialny dobór rankingów dla węzłów zasiewowych w podejściu sekwencyjnym [A5, A10]³

Wiele dotychczasowych badań opiera się na jednoetapowym doborze węzłów początkowych w celu maksymalizacji zasięgu w sieci. Sprowadza się to na ogół do wyboru użytkowników początkowych sieci (ang. seeds), uruchomienia kampanii i ewaluacji wyników. W sekcji 4.2 zademonstrowano, że kampanie rozprzestrzeniania informacji w sieciach społecznych mogą być planowane z wykorzystaniem mniejszych modeli teoretycznych, co znacznie obniża wymagane moce obliczeniowe. Przeprowadzone badania pokazały, że mimo iż wykorzystane sieci syntetyczne były znacznie mniejsze i mniej skomplikowane obliczeniowo, współczynnik korelacji wyników na sieci syntetycznej i sieci rzeczywistej przekroczył wartość 0.9.

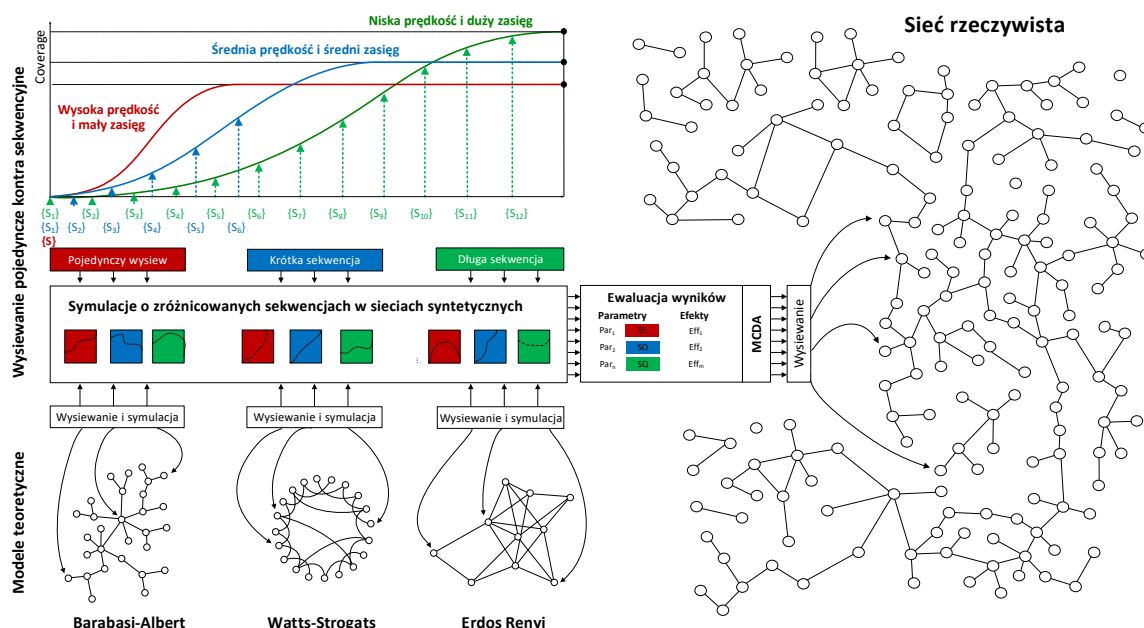
³Sekcja powstała w oparciu o publikację **A5** i wysoko-punktowaną publikację **A10** (200 punktów ministerialnych)



Rysunek 4.6. Analiza wizualna GAIA dla wyboru wielkości próbki sieci rzeczywistej. A - zwykła funkcja preferencji (ang. usual preference function). B - liniowa funkcja preferencji (ang. linear preference function).

Przegląd literatury pokazuje, że kryteria oceny strategii wykorzystane w sekcji 4.2 można dodatkowo rozszerzyć o sekwencyjną inicjalizację kampanii (ang. sequential seeding, [32]). Oznacza to, że strategia może przewidywać wystąpienie więcej niż jednej iteracji wprowadzania informacji do sieci (ang. seeding). Co więcej, takie iteracje mogą następować po sobie bez przerw lub być rozproszone po kampanii w określonych odstępach czasowych. To z kolei rodzi ciekawe pytanie badawcze, czy rozszerzenie modelu decyzyjnego doboru strategii kampanii rozprzestrzeniania informacji w sieciach społecznych o te nowe parametry pozwoliłoby na lepsze dopasowanie kampanii do potrzeb inicjatorów kampanii. W ramach prac badawczych początkowe podejście sekwencyjne z [32] rozszerzono poprzez zestawienie sieci syntetycznych ze zróżnicowanymi sekwencjami inicjalizacji kampanii w celu wielokryterialnego wyboru strategii rozprzestrzeniania informacji w sieci rzeczywistej.

W części metodycznej publikacji **A5** zaproponowano rozszerzenie wielokryterialnego podejścia z sekcji 4.2 o sekwencyjną inicjalizację kampanii (rys. 4.7). Założoną część węzłów początkowych (ang. seeding fraction) można wykorzystać do inicjalizacji kampanii na różne sposoby. Całość wyselekcjonowanych węzłów można zasilić informacją od razu na początku kampanii. Z drugiej strony, jednakże, można podzielić zbiór wyselekcjonowanych



Rysunek 4.7. Schemat poglądowy proponowanego podejścia do sekwencyjnej inicjalizacji węzłów z wykorzystaniem symulacji w sieciach syntetycznych.

węzłów na mniejsze podzbiory i aktywować je w kilku rzutach. Iteracje zasilania sieci w węzły początkowe mogą następować po sobie jedna po drugiej lub z przerwami.

Uwzględniając propozycję sekwencyjnej inicjalizacji, zaproponowano nowy zestaw parametrów do planowania strategii kampanii rozprzestrzeniania informacji w sieciach społecznych. Uwzględniono pięć kryteriów dotyczących parametrów kampanii (Par1 - Par5) oraz dwa kryteria wydajnościowe (Eff6 - Eff7):

Par1 – odsetek węzłów inicjalnych (ang. seeding fraction)

Ułamek wszystkich węzłów, które zostały wybrane do pierwotnego dostarczenia informacji celem dalszego przekazywania w sieci.

Par2 – prawdopodobieństwo propagacji (ang. propagation probability)

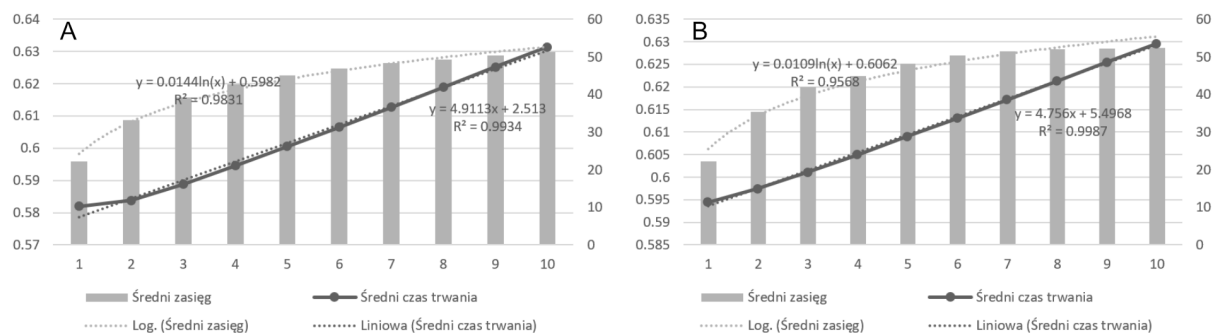
Zakładane prawdopodobieństwo przekazania informacji z jednego zainfekowanego węzła do innych niezainfekowanych węzłów. Poziom prawdopodobieństwa propagacji można dostosować, stosując zachęty użytkowników do przekazywania informacji.

Par3 – liczba iteracji inicjalizacji procesu (ang. seeding iterations' count)

W pierwotnym modelu decyzyjnym z sekcji 4.2, do uruchomienia procesu propagacji informacji w sieci wykorzystywany był pojedynczy rzut informacji do wyselekcjonowanych węzłów inicjalnych. Bazując na początkowych sukcesach uzyskanych w badaniach z sekcji 4.2, postanowiono rozszerzyć oryginalny model o procedury sekwencyjnej inicjalizacji kampanii (ang. sequential seeding). Ten parametr określa, ilokrotnie informacje będą wprowadzone do sieci (uwzględniając zarówno początkowe, jak i kolejne rzuty informacji do sieci).

Par4 – interwał pomiędzy sekwencjami wprowadzania informacji do sieci

Wstępna inicjalizacja kampanii jest zawsze wykonywana w pierwszej iteracji. Jeśli istnieje więcej niż jedna iteracja zasilania sieci w informacje (patrz Par3), Par4 określa



Rysunek 4.8. Wpływ A) liczby iteracji, B) interwału pomiędzy iteracjami inicjalnego zasilania sieci informacjami na zasięg i czas trwania kampanii na przykładzie sieci rzeczywistej Gnutella [29]

jaki jest odstęp czasu pomiędzy każdą procedurą wprowadzania informacji do sieci wewnątrz pojedynczej kampanii.

Par5 – miara centralności wykorzystywana do budowy rankingu węzłów

Węzły wybrane do początkowego umieszczenia informacji nie są wybierane losowo. Najpierw są sortowane według wybranej metryki, a następnie wybierane są te najlepsze. Każda możliwa metryka charakteryzuje się określonym kosztem obliczeniowym.

Eff6 – liczba iteracji

Parametr ten reprezentuje moment w symulacji, w którym doszło do ostatniej infekcji, tj. kiedy wygasł proces propagacji informacji. Wartość tego parametru wynosi co najmniej $1 + (Par3 - 1) \times Par4$.

Eff7 – uzyskany zasięg w sieci

Jest to całkowity zasięg osiągnięty przez symulację strategii opartej na parametrach Par1-Par5, tj. stosunek zainfekowanych węzłów do całkowitej liczby węzłów w sieci.

W badaniach empirycznych ponownie oparto się na sieci Gnutella [29]. Do planowania kampanii wykorzystano wstępnie 15 sieci syntetycznych o połowie wielkości sieci rzeczywistej. Za pomocą miary KLD [27] wybrano jedną z sieci BA do dalszych analiz. Następnie zademonstrowano działanie proponowanego podejścia dla dwóch przeciwnych celów kampanii. W pierwszym przypadku oczekiwano maksymalizacji zasięgu przy dużej dynamice kampanii skutkującej krótkim trwaniem kampanii. W drugim – maksymalizacji zasięgu w sieci, ale przy jak najdłuższym podtrzymywaniu trwania kampanii. Do ewaluacji wykorzystano metodę TOPSIS, co umożliwiło nie tylko łatwą ewaluację 9100 możliwych strategii, lecz również późniejsze przeprowadzenie analizy wrażliwości uzyskanych rozwiązań.

Prace empiryczne zakończono zbadaniem wpływu dwóch nowych kryteriów Par3 i Par4 na ostateczny zasięg i czas trwania procesu rozpowszechniania informacji. Dane z symulacji przeprowadzonych w sieci rzeczywistej Gnutella [29] zostały zagregowane i przedstawione na rysunku 4.8.

Wykres na A pokazuje, że wraz ze wzrostem liczby iteracji inicjalizacji kampanii (ang. seeding iterations) wzrastało zarówno średnie pokrycie, jak i czas trwania procesu. Średni wzrost czasu trwania symulacji można przybliżyć funkcją liniową $y = 4.9113x + 2.513$ z $R^2 = 0.9934$, natomiast średni wzrost zasięgu można przybliżyć funkcją logarytmiczną $y = 0.0144\ln(x) + 0.5982$ z $R^2 = 0.9831$. Podobny wzrost średniego zasięgu i średniego

czasu trwania procesu rozprzestrzeniania informacji można zaobserwować, gdy interwał pomiędzy iteracjami inicjalizacji kampanii jest zwiększony (patrz wykres B). Wzrost czasu trwania można przybliżyć funkcją liniową $y = 4.756x + 5.4968$ z $R^2 = 0.9987$, podczas gdy średni wzrost zasięgu można przybliżyć funkcją logarytmiczną $y = 0.0109\ln(x) + 0.6062$ with $R^2 = 0.9568$.

4.5. Oddziaływanie poprzez nierównomierny rozrzut prawdopodobieństwa propagacji informacji [A6]⁴

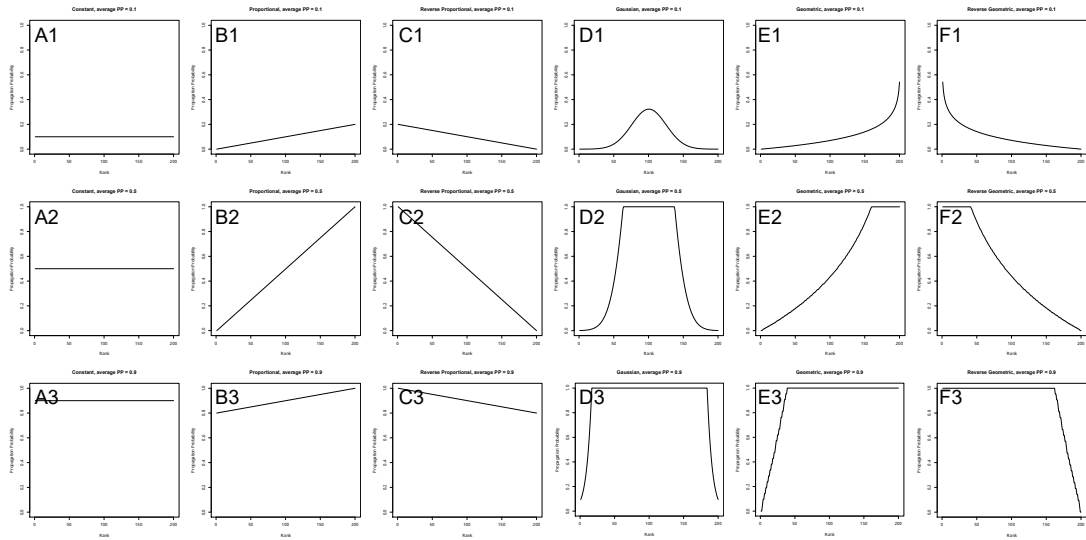
W tej sekcji przedstawione zostało wywieranie wpływu na dynamikę i zasięg procesu dyfuzji informacji poprzez nierównomierny rozrzut prawdopodobieństwa propagacji (ang. propagation probability spraying). W proponowanym podejściu, podobnie jak miało to miejsce we wcześniejszych sekcjach rozprawy, założono, że raz zainicjalizowane kampanie w sieciach społecznych mogą być następnie wspierane poprzez zwiększanie motywacji węzłów w sieci do przekazywania informacji do kolejnych węzłów. O ile we wcześniejszych badaniach zakładano, że prawdopodobieństwo propagacji w sieci rozrzucone jest losowo (lub jest jednolite), tak aby łącznie osiągnąć określony poziom średni (ang. average propagation probability), w publikacji **A6** zaproponowane zostało nowatorskie podejście.

W pracy zauważono, że na poszczególne węzły można oddziaływać np. zachętami w taki sposób, żeby wpływać na ich prawdopodobieństwo propagacji. Można rozważyć kilka strategii skupionych na zwiększeniu aktywności węzłów o wysokiej centralności. Jednak tacy użytkownicy (węzły) mogą być wymagający i trudno do nich dotrzeć. W związku z tym inną możliwością może być zwiększenie aktywności użytkowników o średnich lub nawet niewielkich wartościach miar centralności. Zwiększanie motywacji użytkowników o niskich wartościach miar centralności może być wydajniejsze pod względem kosztów, w związku z potencjalnie mniejszymi zachętami niż dla popularnych węzłów centralnych (ang. hubs).

Założmy, że prawdopodobieństwo rozpowszechnienia treści jest bezpośrednio związane z motywacją użytkownika. Z założonego zbioru i rozkładów D wybieramy rozkład D_i . Tworzony jest wektor $P_i[p_1, p_2, p_n]$ z n elementami, gdzie n odpowiada liczbie węzłów w sieci. Wektor P_i zawiera rozkład prawdopodobieństw, a każdy element reprezentuje prawdopodobieństwo, które należy przypisać do odpowiedniej pozycji w rankingu. Dla celów inicjalizacji kampanii w sieci (ang. seeding), w sieci z n węzłów tworzymy ranking węzłów reprezentowanych przez wektor $R_j[r_1, r_2, r_n]$ z węzłami uporządkowanymi według ich miary centralności typu j . Funkcja $f(p_i, r_i)$ przypisuje prawdopodobieństwo p_i elementowi o rankingu r_i . Funkcja $f(r_i, v_i)$ odwzorowuje prawdopodobieństwa przypisane węzłom w rankingu na poszczególne wierzchołki w sieci.

W trakcie badań nad proponowanym podejściem opracowano szereg algorytmów do uzyskiwania wektorów rozkładów prawdopodobieństw P_i dla zadanej liczby węzłów n i średniego prawdopodobieństwa propagacji P_{avg} . Przykładowe wyniki działania tych algorytmów dla różnorodnych rozkładów prawdopodobieństw przedstawiono na rysunku 4.9.

⁴Sekcja powstała na podstawie recenzowanej publikacji konferencyjnej **A6** indeksowanej w bazach WoS i Scopus

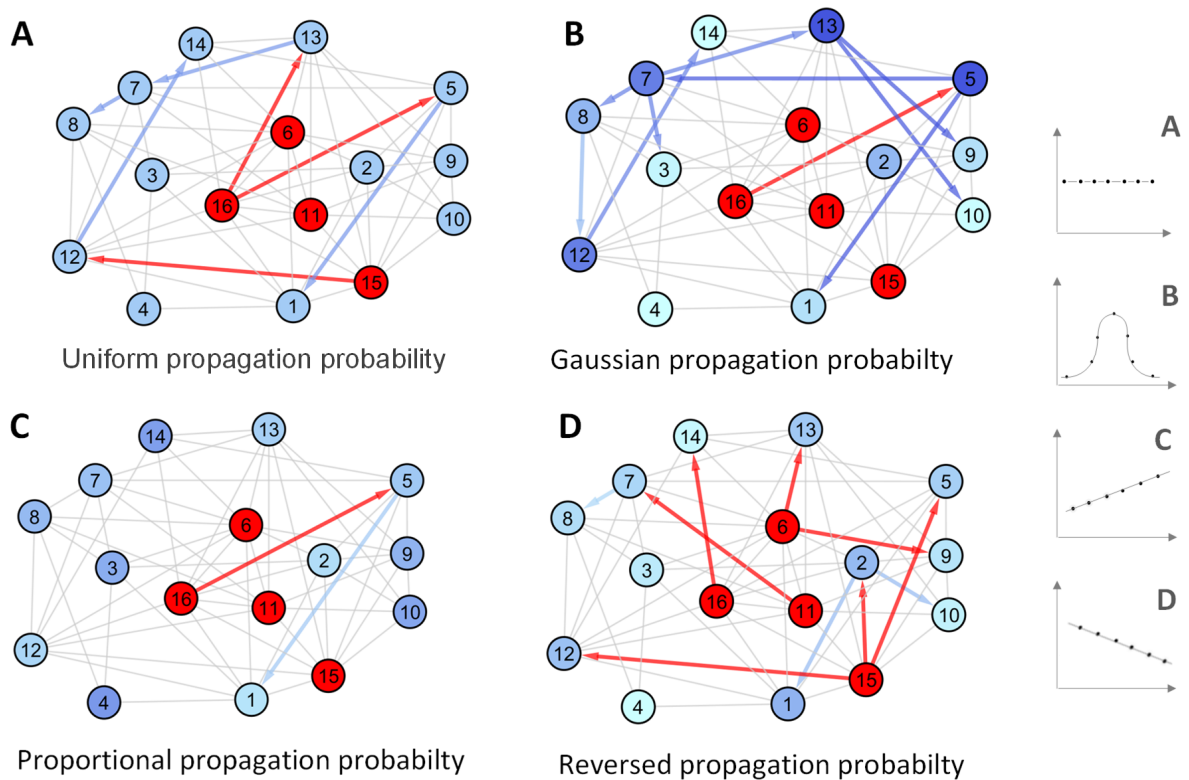


Rysunek 4.9. Przykładowe wyniki działania algorytmów generacji wektora P_i dla rozkładów **A** jednostajnego (ang. uniform), **B** proporcjonalnego (ang. proportional), **C** odwrotnie proporcjonalnego (ang. reversed proportional), **D** normalnego (ang. Gaussian), **E** geometrycznego (ang. geometric) i **F** odwróconego geometrycznego (ang. reversed geometric) przy założonym średnim prawdopodobieństwie propagacji równym **1** – 0.1, **2** – 0.5 i **3** – 0.9.

Tablica 4.2. Przedstawienie wartości prawdopodobieństwa dla wszystkich węzłów sieci [33] dla średniego prawdopodobieństwa propagacji wynoszącego 0.2

Ranking	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Węzeł	6	11	15	16	1	2	12	13	5	7	8	9	3	10	14	4
Stopień	10	9	9	9	8	8	8	8	7	7	7	7	6	5	5	3
r. jednorodny	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000
r. normalny	0.0002	0.0016	0.0092	0.0383	0.1195	0.2806	0.4957	0.6589	0.6589	0.4957	0.2806	0.1195	0.0383	0.0092	0.0016	0.0002
r. proporcjonalny	0.0000	0.0266	0.0533	0.0800	0.1066	0.1333	0.1600	0.1866	0.2133	0.2400	0.2666	0.2933	0.3200	0.3466	0.3733	0.4000
r. odwr. prop.	0.4000	0.3733	0.3466	0.3200	0.2933	0.2666	0.2400	0.2133	0.1866	0.1600	0.1333	0.1066	0.0800	0.0533	0.0266	0.0000

Efekty różnych strategii nierównomiernego rozrzutu prawdopodobieństwa zostały zilustrowane na prostym przykładzie na rysunku 4.10. Przedstawiony przykład jest oparty na symulacji w ramach sieci rzeczywistej składającej się z 16 węzłów [33]. W symulacjach założono, że inicjalnie zainfekowane zostanie 25% sieci, co przekładało się na 4 węzły (ang. seeds) o najwyższej centralności stopnia (ang. degree centrality). Na przykładzie przedstawiono cztery strategie rozrzutu prawdopodobieństwa propagacji (ang. propagation probability spraying), w oparciu o rozkłady jednorodny, normalny, proporcjonalny i odwrotnie proporcjonalny. Tabela 4.2 zawiera wartości prawdopodobieństwa propagacji (PP) dla każdego węzła w sieci, zgodnie z wybranymi dystrybucjami. Na rysunku 4.10 węzły wybrane do inicjalizacji kampanii oznaczono na czerwono. Odcień niebieskiego wskazuje na wielkość prawdopodobieństwa propagacji w odniesieniu do określonego rozkładu. Graf A przedstawia proces oparty na jednorodnym prawdopodobieństwie propagacji, w którym do każdego węzła przypisana jest ta sama wartość PP, równa 0.2. Proces kończy się z zasięgiem 68.75% z 11 zainfekowanymi węzłami w 4 krokach. Graf B ilustruje proces oparty na normalnym rozkładzie prawdopodobieństwa. Ostateczny zasięg 87.5% został osiągnięty w 6 krokach, przy 14 zainfekowanych węzłach. Dla rozkładu proporcjonalnego zilustrowanego na grafie C zasięg wynosi 37.5%, czyli 6 zainfekowanych węzłów w 3 krokach. Odwrócony rozkład proporcjonalny przedstawiony na grafie D skutkował zasięgiem 87.5%, czyli czternastoma zainfekowanymi węzłami w trzech krokach.



Rysunek 4.10. Przykładowy proces rozprzestrzeniania informacji z rozkładem **A** jednorodnym (ang. uniform); **B** normalnym (ang. Gaussian); **C** proporcjonalnym (ang. proportional); i **D** odwrotnie proporcjonalnym (ang. reversed proportional) prawdopodobieństwa propagacji informacji.

Powyższy przykład poglądowy pokazuje, jak oddziaływanie na procesy rozprzestrzeniania informacji w sieciach społecznych poprzez nierównomierny rozrzut prawdopodobieństwa propagacji informacji może wpływać na zasięg i czas trwania procesu. W kolejnym etapie badań przeprowadzone zostały symulacje na 25 sieciach syntetycznych i 5 sieciach rzeczywistych w celu oceny wpływu rozrzuconego rozkładu prawdopodobieństwa propagacji na końcowy zasięg kampanii. Opis badanych sieci oraz szczegółowe analizy przedstawione zostały w sekcji IV publikacji **A6**. Dwa z analizowanych podejść oparto na odwróconym rozkładzie geometrycznym i odwróconym rozkładzie proporcjonalnym, z większym wzrostem prawdopodobieństwa dla węzłów o wysokim rankingu. Skutkowały one największym wzrostem zasięgu w sieci w porównaniu z rozkładem jednorodnym prawdopodobieństwa. Zastosowanie geometrycznego i proporcjonalnego rozkładu prawdopodobieństw spowodowało natomiast zmniejszenie zasięgu w porównaniu z jednorodnym prawdopodobieństwem propagacji, ale ogólne koszty kampanii mogą być niższe. Podejście z rozkładem Gaussa okazało się skutkować najśłabszym zasięgiem, prawdopodobnie w związku z faktem, że znaczna część zachęt przeznaczona została w nim dla węzłów o przeciętnych rankingach.

4.6. Oddziaływanie poprzez wielokryterialne targetowanie [A7, A8, A9]⁵

Większość dotychczasowych badań nad procesami rozprzestrzeniania informacji w sieciach społecznych zakłada jednorodność wszystkich węzłów (użytkowników) w sieci. Oznacza to koncentrację na dotarciu do jak największej grupy obojętnie których węzłów w sieci. Nieliczne spośród najnowszych badań zaczęły koncentrować się na kampaniach targetowanych [19, 20]. W odróżnieniu od tradycyjnych kampanii, spośród wszystkich użytkowników sieci wskazuje się tutaj podzbiór tych użytkowników, do których zamawiający kampanię chce dotrzeć. Miarą efektywności takiej kampanii targetowanej nie jest globalny zasięg w sieci, lecz zasięg osiągnięty bezpośrednio w grupie docelowej. Część badań koncentruje się również na unikaniu powtarzania wiadomości, w celu uniknięcia efektu habituacji [34], czy też przeładowania użytkowników informacjami.

Należy jednak zwrócić uwagę, że dotychczasowe badania w dziedzinie targetowanych kampanii wirusowych w sieciach społecznych bazowały homogenicznie na pojedynczym atrybucie lub mierze centralności w celu doboru węzłów początkowych do uruchomienia kampanii (ang. seeds). W praktyce jednak, rzeczywiste zastosowania sieci społecznych w wirusowych kampaniach marketingowych często opierają się heterogenicznie na wyborze wielu atrybutów użytkowników, takich jak wiek, płeć czy lokalizacja grupy docelowej.

W tej sekcji przedstawiono nowatorskie podejście do badania procesów rozprzestrzeniania informacji w sieciach społecznych. Proponowana metodologia uzupełnia szeroko stosowany model Independent Cascade (IC) [16] poprzez uwzględnienie problemu docierania do docelowych węzłów wieloatrybutowych w sieciach społecznych. W proponowanym podejściu zakłada się, że węzły sieci charakteryzują się nie tylko relacjami centralności pomiędzy nimi a innymi węzłami, ale także zestawem niestandardowych atrybutów $C_1, C_2, \dots, C_n$.

Wartości tych atrybutów dla poszczególnych węzłów można wyrazić jako dokładne wartości liczbowe, takie jak wiek [lata] czy dochód [dolary]. Alternatywnie, jeśli atrybuty reprezentują jakościowe właściwości węzłów, ich wartości można przeliczyć na wartości liczbowe za pomocą 5-stopniowej skali Likerta [35, 36] (1 - zdecydowanie się nie zgadzam, 5 - zdecydowanie się zgadzam) lub wyliczenia (np. wiek: 1 - młody, 2 - w średnim wieku, 3 - stary; lub płeć: 1 - mężczyzna, 2 - kobieta).

Węzły można również scharakteryzować za pomocą miar centralności, takich jak centralność stopnia, bliskości, pośrednictwa czy wektora własnego. Dodatkowe atrybuty można również wyprowadzić jako złożenie dwóch wyżej wymienionych typów atrybutów. Na przykład, jeśli atrybut C_i reprezentuje stopień węzła, tj. całkowitą liczbę jego sąsiadów, C_{i_1} może reprezentować liczbę mężczyzn w jego sąsiedztwie, a C_{i_2} liczbę kobiet w jego sąsiedztwie.

W proponowanym podejściu podjęto próbę dotarcia do węzłów o określonych wartościach wybranych atrybutów. Przykładowo, w programie profilaktyki raka piersi [37] podejmuje się próbę dotarcia do kobiet w średnim wieku, tj. między 50 a 69 rokiem życia.

⁵Sekcja powstała w oparciu o artykuł **A7** opublikowany w czasopiśmie Symmetry o współczynniku IF 2.645 oraz o dwie wysoko punktowane publikacje **A8** i **A9** (140 punktów ministerialnych każda).

W modelu IC [16] proces propagacji informacji w sieci poprzedzony jest selekcją inicjalnych węzłów do uruchomienia kampanii (ang. seeds). Zwykle nasiona reprezentują określoną część wszystkich węzłów sieci. Na przykład założone może być zainicjalizowanie kampanii poprzez przekazanie informacji do 5% użytkowników w sieci. Istnieje wiele sposobów wyboru inicjalnych węzłów sieci, które generalnie sprowadzają się do opracowania rankingu wszystkich węzłów w sieci i wyboru tych najwyżej punktowanych. Podczas gdy inne podejścia koncentrują się na generowaniu rankingu na podstawie pojedynczej miary centralności, takiej jak stopień, w podejściu proponowanym w tej sekcji rozważa się wiele atrybutów w celu wybrania węzłów inicjalnych.

Należy zauważyć, że w proponowanym podejściu ostateczny globalny zasięg sieci, czyli odsetek węzłów, do których dotarły informacje, może być niższy niż w przypadku tradycyjnych podejść opartych na miarach centralności. Jednak, co ukazały badania w publikacjach **A7**, **A8**, proponowane podejście daje szansę na zwiększenie zasięgu w grupach węzłów docelowych wewnątrz sieci.

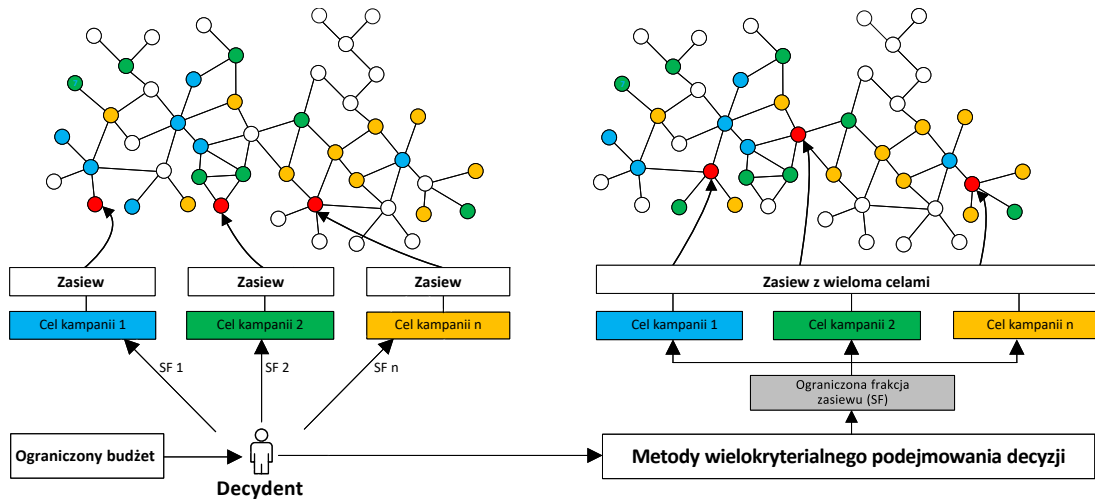
W badaniach empirycznych w publikacjach **A7** i **A8** oparto się na sieciach rzeczywistych i syntetycznych. Ponieważ znalezienie modelu sieci o węzłach scharakteryzowanych wieloma atrybutami okazało się niemożliwe, postanowiono na zwykłe grafy, reprezentowane przez macierz sąsiedztwa, sztucznie nałożyć atrybuty na podstawie danych demograficznych [38] - płci (kobieta, mężczyzna) i grupy wiekowej (młodzi, średni, starsi). Następnie próbowano dotrzeć do różnych grup docelowych, przykładowo do kobiet w średnim wieku, bądź młodszych mężczyzn.

Dwa wyżej wspomniane atrybuty węzłów sieci, tj. płeć i wiek naturalnie zostały wybrane jako kryteria decyzyjne do procesu wyboru węzłów inicjalnych (ang. seeds) w kampanii. Kolejnym wybranym kryterium był stopień każdego węzła, jako najbardziej tradycyjna miara centralności do wyboru węzłów startowych. Ponadto, utworzono dodatkowe kryteria ewaluacyjne jako złożenie miary stopnia i pozostałych atrybutów sieci. Finalnie zaproponowano model decyzyjny oparty na 8 kryteriach:

- C1 – stopień węzła, tj. liczba jego sąsiadów;
- C2 – płeć użytkownika;
- C3 – liczba sąsiadów płci męskiej;
- C4 – liczba sąsiadów płci żeńskiej;
- C5 – wiek użytkownika;
- C6 – liczba sąsiadów w młodszym wieku;
- C7 – liczba sąsiadów w średnim wieku;
- C8 – liczba sąsiadów w starszym wieku.

W badaniach empirycznych w publikacjach **A7** i **A8** do wyboru węzłów inicjalnych zastosowano metody wielokryterialne TOPSIS i PROMETHEE (do ewaluacji węzłów) oraz AHP (do doboru wag kryteriów). Zastosowanie proponowanego podejścia umożliwiło uzyskanie w grupie docelowej zasięgu większego o 7.14% niż w przypadku tradycyjnego podejścia opartego o miarę centralności stopnia.

W związku z dobrymi osiąganiami proponowanego podejścia, zdecydowano o dalszym jego badaniu. Poskutkowało to jego rozszerzeniem w kierunku zrównoważenia (ang. susta-



Rysunek 4.11. Schemat poglądowy proponowanego zrównoważonego podejścia do realizacji pojedynczej kampanii realizującej n celów zamiast n kampanii realizujących pojedyncze cele.

inability) procesu doboru węzłów inicjalnych dla komplementarnych celów kampanii w sieciach społecznych (publikacja **A9**). W tak rozszerzonym podejściu zakłada się, że inicjator procesu rozprzestrzeniania informacji w sieci społecznej stara się osiągnąć n celów, tj. chce zmaksymalizować wpływ w n zbiorach docelowych węzłów (ang. targeted nodes). W tradycyjnym podejściu można byłoby przeprowadzić n kampanii, po jednej dla każdej grupy docelowej. Jednakże w proponowanym podejściu zakłada się ograniczony budżet, a tym samym ograniczony zbiór użytkowników początkowych sieci (ang. seeds), na bazie których można oprzeć kampanię. W związku z tym, proponuje się podejście zrównoważone, w którym decydent (ang. decision-maker, DM) stosuje metodę MCDA i na podstawie celów kampanii $1, 2, \dots, n$, oceny eksperckiej oraz doświadczenia analityka wybiera ograniczoną część węzłów sieci do zainicjowania pojedynczej kampanii w celu dotarcia do węzłów odpowiadających wielu uzupełniającym się celom zamawiającego (zob. rys. 4.11).

Ze względu na zrównoważony charakter proponowanego podejścia zakłada się, że globalny zasięg sieci może zostać zmniejszony, lecz celem jest zwiększenie zasięgu wśród docelowych węzłów sieci. W badaniach empirycznych, przedstawionych w publikacji **A9**, dla przykładowych kampanii łączących dwa rozłączne cele udało się osiągnąć zasięg w grupach docelowych większy o 1.74% od tradycyjnego podejścia, przy zmniejszonym zasięgu globalnym o 2.1%.

5. Otwarte zorientowane obiektowo środowisko symulacyjne do badania procesu dyfuzji informacji w sieciach złożonych [A10]¹

Grafy i złożone sieci, a także zachodzące w nich procesy rozprzestrzeniania informacji to interdyscyplinarny temat badawczy, znajdujący zainteresowanie w takich dyscyplinach jak informatyka, fizyka, medycyna, epidemiologia [15, 7, 3]. Model kaskadowy (ang. independent cascade model, IC, [16]) oraz symulacje agentowe pozwalają na badanie procesów propagacji informacji w sieciach złożonych.

Istnieją biblioteki w języku R, takie jak `igraph`² oraz `netdep`³, które umożliwiają reprezentację grafów i złożonych sieci w środowisku skryptowym języka R. Jednak aby przeprowadzić eksperymenty w modelu IC, badacze zmuszeni dotychczas byli do pisania własnych skryptów, zwłaszcza jeśli ich badania koncentrowały się na adaptacyjnym lub sekwencyjnym inicjalizowaniu kampanii (ang. adaptive and sequential seeding). Przemieszczenie logiki modelu IC z właściwą logiką rozpowszechniania informacji wymaga od naukowców ponownego implementowania całości skryptu przy każdym kolejnym projekcie badawczym. Co więcej, podejście takie pomija osiągnięcia i zalety paradygmatu programowania obiektowego [39]. Stworzyło to interesującą lukę, którą w trakcie prac nad rozprawą udało się wypełnić zorientowaną obiektowo biblioteką wraz z wykorzystującym ją środowiskiem do symulacji, badania oraz oddziaływania na procesy rozprzestrzeniania informacji w sieciach złożonych – w tym sieciach społecznych.

Zastosowanie paradygmatów programowania obiektowego pozwoliło na enkapsulację warstwy złożonej i powtarzalnej logiki symulacji procesów rozprzestrzeniania informacji w sieciach do postaci łatwo reużywalnej biblioteki OONIS⁴. Ponadto zastosowane podejścia modułowego i podziału odpowiedzialności pozwoliło na zbudowanie środowiska, w którym tworzenie scenariuszy symulacyjnych jest łatwe, skalowalne i szybkie. Wdrożenie w projektach badawczych biblioteki OONIS pozwoliło podczas tworzenia tej rozprawy na tworzenie scenariuszy eksperymentalnych badających różne podejścia, bez potrzeby utrzymywania warstwy logiki symulacyjnej, a w szczególności – mechaniki symulacji wysiewu informacji (ang. information seeding), infekowania węzłów i rejestrowania wyników. Przygotowana i przetestowana wielokrotnie podczas prac nad tą rozprawą biblioteka OONIS została następnie udostępniona do użytku publicznego na licencji GNU GPLv3 za sprawą publikacji o otwartym dostępie **A10**.

Zaprojektowane oprogramowanie składa się z dwóch części. Pierwsza, to zorientowana obiektowo biblioteka w R do symulacji procesów rozprzestrzeniania informacji w sieciach

¹Sekcja powstała w oparciu o wysoko punktowaną publikację **A10** (200 punktów ministerialnych).

²<https://igraph.org>

³<https://cran.r-project.org/web/packages/netdep/index.html>

⁴OONIS – Object-Oriented Network Infection Simulator

```

1  # SEEDER
2  # choose and configure the simulation seeder component
3  seeder <- SomeDefaultOrCustomSeeder(...);
4
5  # CONTAMINATOR
6  # choose and configure the simulation contamination component
7  contaminator <- SomeDefaultOrCustomContaminator(...)
8
9  # RESULT PRINTER
10 resultPrinter <- SomeDefaultOrCustomResultPrinter(...);
11
12 # SIMULATION
13 maxIterations <- 100; minIterations <- 1;
14 ir <- InfectionRunner(infectionSeeder = seeder, contaminator = contaminator, resultPrinter = resultPrinter)
15 ir$readFromEdgesTxt('/path/to/network.txt', FALSE)
16 ir$run(maxIterations, minIterations)

```

Rysunek 5.1. Przykładowy kod źródłowy prostego scenariusza symulacji procesu rozprzestrzeniania informacji w sieci złożonej z wykorzystaniem proponowanego środowiska [40].

złożonych. Druga część to proponowane środowisko do przeprowadzania eksperymentów z wykorzystaniem owej biblioteki. Biblioteka składa się z głównej klasy *InfectionRunner*, która odpowiada za mechanikę symulacji w modelu IC [16]. Przed uruchomieniem symulacji, do obiektu klasy *InfectionRunner* muszą zostać podłączone instancje trzech modułów: *seeder* (do wprowadzania informacji do sieci), *contaminator* (do przekazywania informacji między węzłami) oraz *result printer* (do rejestrowania wyników).

W skład opublikowanej biblioteki wchodzi 7 przykładowych modułów zasiewania oraz trzy przykładowe moduły do obsługi przekazywania informacji. Domyślnie biblioteka umożliwia rejestrowanie wyników do postaci plików rozdzielanych przecinkami.

Opublikowana biblioteka, wraz z bazującym na niej środowiskiem symulacyjnym, za sprawą enkapsulacji oraz podziału odpowiedzialności poszczególnych wymiennych modułów, umożliwia w łatwy sposób przeprowadzać eksperymenty heterogenicznych scenariuszy oddziaływania na procesy rozprzestrzeniania informacji w sieciach społecznych, takich jak:

- oddziaływanie przez zmienną frakcję inicjalnych węzłów w kampanii;
- oddziaływanie przez wysiewanie pojedyncze lub sekwencyjne w wielu iteracjach (publikacja **A5**);
- oddziaływanie przez nierównomierny rozrzut prawdopodobieństwa propagacji w sieci (publikacja **A6**);
- ewaluacja i planowanie kampanii marketingu wirusowego w sieciach społecznych w oparciu o sparametryzowane wartości frakcji wysiewu, prawdopodobieństwa propagacji, miar centralności węzłów i rankingi (publikacje **A1**, **A2**, **A3**, **A4**)

Na rysunku 5.1 zaprezentowano przykład wykorzystania proponowanego środowiska, ukazujący jak łatwe jest tworzenie rozmaitych scenariuszy symulacyjnych dla sieci złożonych z wykorzystaniem biblioteki i środowiska powstałych w trakcie tworzenia tej rozprawy. Wystarczy zainicjalizować obiekt *InfectionRunner* jednym z domyślnych lub samodzielnie oprogramowanych modułów wysiewu, zarażania i zbierania wyników, wczytać sieć za pomocą listy krawędzi i uruchomić symulację.

6. Podsumowanie

Procesy rozprzestrzeniania informacji w sieciach społecznych generują szereg wyzwań dla informatyki w obszarach modelowania, optymalizacji i efektywności obliczeniowej. Aktualne kierunki badawcze odnoszą się między innymi do rozkładu zasiewu w czasie, procesów adaptacyjnych, optymalizacji zasobów czy modelowania procesów konkurujących. Badania zostały posadowione w dyscyplinie *Informatyka techniczna i telekomunikacja*, a w szczególności zgodnie z klasyfikacją ACM Digital Library Computing Classification System w obszarach *Human-centered computing · Collaborative and social computing · Collaborative and social computing theory, concepts and paradigms · Social networks*.

W ramach realizowanych badań cel rozprawy osiągnięto poprzez wprowadzenie szeregu rozszerzeń do aktualnych rozwiązań. Dotychczas uwaga środowiska badawczego była zwrócona na problem maksymalizacji zasięgu poprzez odpowiedni dobór zbioru węzłów zasiewowych wykorzystywanych do jednorazowej inicjalizacji procesu. W rzeczywistych procesach poza samym zasięgiem istotne są też parametry czasowe, charakterystyka grup docelowych czy dostępne zasoby. W niniejszej rozprawie podjęto próbę zbliżenia modeli teoretycznych do potrzeb systemów rzeczywistych i zaproponowano ramy do wielokryterialnego planowania i ewaluacji procesów rozprzestrzeniania informacji w sieciach społecznych. Następnie przedstawiono szereg sposobów heterogenicznego oddziaływania na procesy rozprzestrzeniania informacji w nich zachodzące. W rozprawie oddziaływano na procesy rozprzestrzeniania informacji poprzez sekwencyjność inicjalizacji, nierównomierny rozrzut prawdopodobieństwa propagacji informacji oraz targetowanie wielokryterialne.

W szczególności osiągnięcia uzyskane w rozprawie, które stanowią wkład do *Informatyki technicznej i telekomunikacji* obejmują:

- opracowanie autorskiego podejścia do wielokryterialnego planowania i ewaluacji procesów rozprzestrzeniania informacji w sieciach społecznych;
- zaprezentowanie podejścia wykorzystującego sieci syntetyczne i próbki sieci rzeczywistych do planowania procesów rozprzestrzeniania informacji w strukturach sieciowych;
- opracowanie wielokryterialnego podejścia do wyboru rozmiaru próbek sieci rzeczywistej w celu uproszczonego obliczeniowo pozyskiwania przyszłych rekomendacji odnośnie parametrów inicjalizacji procesów w sieci rzeczywistej;
- opracowanie podejścia wykorzystującego sieci syntetyczne i zróżnicowanie sekwencji inicjalizacji węzłów wysiewowych w celu doboru strategii do inicjalizacji i realizacji procesu w sieci rzeczywistej;
- zbadanie oddziaływania nierównomiernego rozrzutu prawdopodobieństwa propagacji informacji w sieciach na osiągany zasięg i dynamikę procesu rozprzestrzeniania informacji;

- opracowanie szeregu algorytmów do generowania wektorów nierównomiernego rozrzutu prawdopodobieństwa propagacji informacji (ang. probability spraying);
- zaproponowanie podejścia do wielokryterialnego targetowania w wieloatrybutowe (heterogeniczne) węzły w sieciach społecznych;
- zaproponowanie zrównoważonego podejścia do adresowania wielu celów działań w sieciach o wieloatrybutowych węzłach z wykorzystaniem pojedynczych kampanii;
- opracowanie otwartego zorientowanego obiektowo środowiska symulacyjnego do badania procesu dyfuzji informacji w sieciach złożonych.

Dodatkowym efektem rozprawy doktorskiej w dziedzinie Informatyka techniczna i telekomunikacja było opublikowanie otwartej zorientowanej obiektowo biblioteki oraz środowiska w języku R do przeprowadzania rozmaitych eksperymentów symulacyjnych z zakresu oddziaływania na procesy rozprzestrzeniania informacji w sieciach złożonych.

Przeprowadzone badania pozwoliły na potwierdzenie postawionej w rozprawie tezy. Oddziaływanie na procesy rozprzestrzeniania informacji w sieciach złożonych zróżnicowanymi metodami umożliwiło zwiększenie zasięgu procesu i jego dynamiki oraz innych charakterystyk zgodnie z preferencjami decydenta, które nie znajdowały odzwierciedlenia we wcześniejszych rozwiązaniach.

Prace prowadzone w ramach rozprawy były częściowo realizowane w ramach grantów NCN (OPUS) numer 2016/21/B/HS4/01562 (**A1**, **A3**, **A4**, **A6**, **A10**) oraz 2017/27/B/HS4/01216 (**A7**, **A8**, **A9**), w których doktorant był wykonawcą/stypendystą.

W trakcie prac nad rozprawą doktorską zidentyfikowano potencjalne obszary dalszych prac badawczych. Korzystne do dalszych badań byłoby mapowanie rzeczywistych sieci o węzłach charakteryzowanych wieloma atrybutami na środowisko realizacji procesów propagacji. Planowany jest także dalszy rozwój zrównoważonego podejścia do inicjalizacji pojedynczych procesów realizujących jednocześnie wiele celów, tradycyjnie uzyskiwanych przez odrębne kampanie.

7. Dorobek akademicki

Niniejszy rozdział prezentuje dorobek akademicki mgr inż. Artura Karczmarczyka, kandydata do stopnia naukowego doktora. W ramach prezentowanego dorobku wyodrębniono osiągnięcia naukowe, dydaktyczne i organizacyjne.

7.1. Dorobek naukowy

7.1.1. Profile internetowe¹

Wskaźniki z profili internetowych Web of Science, Scopus oraz Google Scholar zebrane zostały w tabeli 7.1.

Tablica 7.1. Profile internetowe (stan na dzień 6.05.2021).

Profil	Liczba artykułów	Liczba cytowań	h-index
Web of Science	16	201	9
Scopus	27	429	12
Google Scholar	30	547	13

7.1.2. Wykaz prac naukowych²

Poniżej przedstawiono wykaz prac naukowych opublikowanych oraz w druku. Łączny uzyskany współczynnik IF wynosi 19.663. Łączna liczba punktów ministerialnych za publikacje bez uwzględnienia oświadczeń: 295 (do 2017) oraz 760 (od 2018).

1. Wątróbski, J., Jankowski, J., Ziemia, P., **Karczmarczyk, A.**, Ziolo, M. (2019). Generalised framework for multi-criteria method selection. Omega, 86, 107-124.
[IF: 5.341, 140 pkt, 170 cytowań]
2. Wątróbski, J., Małecki, K., Kijewska, K., Iwan, S., **Karczmarczyk, A.**, Thompson, R. G. (2017). Multi-criteria analysis of electric vans for city logistics. Sustainability, 9(8), 1453
[IF: 2.075, 20 pkt, 55 cytowań]

¹Stan na dzień 4.05.2021 r.

²Stan na dzień 4.05.2021 r.

3. Ziemia, P., Wątróbski, J., Ziolo, M., **Karczmarczyk, A.** (2017). Using the PROSA method in offshore wind farm location problems. *Energies*, 10(11), 1755.
[IF: 2.676, 25 pkt, 51 cytowań, okładka]
4. Wątróbski, J., Ziemia, E., **Karczmarczyk, A.**, Jankowski, J. (2018). An index to measure the sustainable information society: the Polish households case. *Sustainability*, 10(9), 3223.
[IF: 2.075, 20 pkt, 35 cytowań]
5. **Karczmarczyk, A.**, Jankowski, J., Wątróbski, J. (2018). Multi-criteria decision support for planning and evaluation of performance of viral marketing campaigns in social networks. *PloS one*, 13(12), e0209372.
[IF: 2.776, 100 pkt, 34 cytowania]
6. Sałabun, W., **Karczmarczyk, A.** (2018). Using the comet method in the sustainable city transport problem: an empirical study of the electric powered cars. *Procedia computer science*, 126, 2248-2260.
[WOS 15 pkt, 30 cytowań]
7. Wątróbski, J., Jankowski, J., Ziemia, P., **Karczmarczyk, A.**, Ziolo, M. (2019). Generalised framework for multi-criteria method selection: Rule set database and exemplary decision support system implementation blueprints. *Data in brief*, 22, 639.
[40 pkt, 29 cytowań]
8. Wątróbski, J., Sałabun, W., **Karczmarczyk, A.**, Wolski, W. (2017, September). Sustainable decision-making using the COMET method: An empirical study of the ammonium nitrate transport management. In *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)* (pp. 949-958). IEEE.
[WOS 20 pkt, 28 cytowań]
9. Sałabun, W., **Karczmarczyk, A.**, Wątróbski, J., Jankowski, J. (2018, November). Handling data uncertainty in decision making with COMET. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 1478-1484). IEEE.
[WOS 20pkt, 27 cytowań]
10. Sałabun, W., **Karczmarczyk, A.**, Wątróbski, J. (2018, November). Decision-making using the hesitant fuzzy sets COMET method: An empirical study of the electric city buses selection. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 1485-1492). IEEE.
[20 cytowań]
11. Jankowski, J., Ziolo, M., **Karczmarczyk, A.**, Wątróbski, J. (2018). Towards sustainability in viral marketing with user engaging supporting campaigns. *Sustainability*, 10(1), 15.
[IF: 2.075, 20 pkt, 18 cytowań]
12. Ziemia, P., Wątróbski, J., **Karczmarczyk, A.**, Jankowski, J., Wolski, W. (2017, September). Integrated approach to e-commerce websites evaluation with the use of surveys and eye tracking based experiments. In *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)* (pp. 1019-1030). IEEE.
[WOS 15 pkt, 15 cytowań]

13. Wątróbski, J., Jankowski, J., **Karczmarczyk, A.**, Ziemba, P. (2017, September). Integration of eye-tracking based studies into e-commerce websites evaluation process with eQual and TOPSIS methods. In EuroSymposium on Systems Analysis and Design (pp. 56-80). Springer, Cham.
[WOS 15 pkt, 14 cytowań]
14. Wątróbski, J., **Karczmarczyk, A.** (2017). Application of the fair secret exchange protocols in the distribution of electronic invoices. *Procedia computer science*, 112, 1819-1828.
[WOS 15 pkt, 4 cytowania]
15. **Karczmarczyk, A.**, Jankowski, J., Sałabun, W. (2017). Linguistic query based quality evaluation of selected image search engines. *Procedia computer science*, 112, 1809-1818.
[WOS 15 pkt, 3 cytowania]
16. **Karczmarczyk, A.**, Wątróbski, J., Ladorucki, G., Jankowski, J. (2018, September). MCDA-based approach to sustainable supplier selection. In 2018 Federated Conference on Computer Science and Information Systems (FedCSIS) (pp. 769-778). IEEE.
[WOS 15 pkt, 2 cytowania]
17. **Karczmarczyk, A.**, Bortko, K., Bartków, P., Pazura, P., Jankowski, J. (2018, August). Influencing information spreading processes in complex networks with probability spraying. In 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (pp. 1038-1046). IEEE.
[WOS 15 pkt, 2 cytowania]
18. Wątróbski, J., **Karczmarczyk, A.**, Jankowski, J., Ziemba, P., Wolski, W. (2017). Hierarchical Representation of Website Evaluation Model Using Survey and Perceptual Based Criteria. In *Information Technology for Management. Ongoing Research and Development* (pp. 229-248). Springer, Cham.
[WOS 15 pkt, 2 cytowania]
19. **Karczmarczyk, A.**, Wątróbski, J., Jankowski, J. (2019). Multi-Criteria Approach to Planning of Information Spreading Processes Focused on Their Initialization With the Use of Sequential Seeding. In *Information Technology for Management: Current Research and Future Directions* (pp. 116-134). Springer, Cham.
[1 cytowanie]
20. **Karczmarczyk, A.**, Jankowski, J., Wątróbski, J. (2019, September). Multi-criteria approach to viral marketing campaign planning in social networks, based on real networks, network samples and synthetic networks. In 2019 Federated Conference on Computer Science and Information Systems (FedCSIS) (pp. 663-673). IEEE.
[1 cytowanie]
21. **Karczmarczyk, A.**, Jankowski, J., Wątróbski, J. (2019). Parametrization of spreading processes within complex networks with the use of knowledge acquired from network samples. *Procedia Computer Science*, 159, 2279-2293.
[70pkt, 1 cytowanie]
22. Jankowski, J., Michalski, R., Bródka, P., **Karczmarczyk, A.** (2017, July). Increasing Coverage of Information Diffusion Processes by Reducing the Number of Initial Seeds.

In Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017 (pp. 713-720).
[WOS 15 pkt, 1 cytowanie]

23. Sałabun, W., **Karczmarczyk, A.**, Mejsner, P. (2017). Experimental Study of Color Contrast Influence in Internet Advertisements with Eye Tracker Usage. In Neuroeconomic and Behavioral Aspects of Decision Making (pp. 365-375). Springer, Cham.
[WOS 15 pkt, 1 cytowanie]
24. **Karczmarczyk, A.**, Jankowski, J., Wątróbski, J. (2021). OONIS—Object-Oriented Network Infection Simulator. SoftwareX, 14, 100675.
[200 pkt]
25. **Karczmarczyk, A.**, Jankowski, J., Wątróbski, J. (2021). Multi-Criteria Seed Selection for Targeting Multi-Attribute Nodes in Complex Networks. Symmetry, 13(4), 731.
[IF: 2.645, 70pkt]
26. Rymaszewski, S., Wątróbski, J., **Karczmarczyk, A.** (2020). Identification of reference multi criteria domain model-Production line optimization case study. Procedia Computer Science, 176, 3794-3801.
[70 pkt]
27. Wątróbski, J., **Karczmarczyk, A.**, Rymaszewski, S. (2020). Multi-criteria decision making approach to production line optimization. Procedia Computer Science, 176, 3820-3830.
[70 pkt]
28. **Karczmarczyk, A.**, Wątróbski, J., Jankowski, J. (2018). Comparative Study of Different MCDA-Based Approaches in Sustainable Supplier Selection Problem. In Information Technology for Management: Emerging Research and Applications (pp. 176-193). Springer, Cham.
29. **Karczmarczyk, A.** (2011). Zastosowanie sprawiedliwej wymiany sekretów w dystrybucji faktur elektronicznych. Studia i Materiały Polskiego Stowarzyszenia Zarządzania Wiedza/Studies Proceedings Polish Association for Knowledge Management, (57).
30. **Karczmarczyk, A.**, Jankowski, J., Wątróbski, J. (2021). Multi-Criteria Seed Selection for Targeted Influence Maximization within Social Networks – in proceedings of International Conference on Computational Science: ICCS 2021
[w druku, planowany czas publikacji - druga połowa 2021]
31. **Karczmarczyk, A.**, Wątróbski, J., Jankowski, J. (2021). Seeding for Complementary Campaign Objectives in Social Networks - in proceedings of The Americas Conference on Information Systems: AMCIS 2021
[w druku, planowany czas publikacji - druga połowa 2021]

7.1.3. Charakterystyka pozostałego dorobku naukowego

Pozostały dorobek naukowy spoza cyklu **A1-A10** obejmuje 21 publikacji naukowych. Są to publikacje o zasięgu krajowym i światowym. Dużą ich część stanowią publikacje w

międzynarodowych czasopismach oraz materiały z konferencji międzynarodowych indeksowane w bazach Scopus oraz WebOfScience. Tematyka zrealizowanych prac koncentruje się wokół zagadnień metodycznych i praktycznych związanych z wykorzystaniem oraz rozwojem metod MCDA. Doświadczenie pozyskane w realizacji prac poza głównym cyklem wspomogło późniejsze opracowanie rozwiązań MCDA w tematyce propagacji informacji w sieciach społecznych.

Prace (4)³, (12), (13), (18) koncentrują się na aplikacji i doskonaleniu metod MCDA w obszarze cyfrowego zrównoważenia, a w szczególności: poszukiwania zrównoważonych interfejsów, zrównoważoną ocenę jakości systemów informatycznych oraz witryn internetowych. Wzrost roli informacji i technologii informacyjno-komunikacyjnych (ICT) doprowadził do wykształcenia nowych pojęć, a w tym pojęcia społeczeństwa informacyjnego (ang. information society). Pojęcie to stanowi adaptację pryncypiów zrównoważonego rozwoju na grunt społeczeństwa informacyjnego (ang. sustainable information society, SIS). Problem badawczy podjęty w pracy (4) stanowiło opracowanie metodyki pomiaru zrównoważonego społeczeństwa informacyjnego. Uwzględniono przy tym literaturę przedmiotu, gdzie do podstawowych zadań realizowanych podczas pomiaru SIS zalicza się: możliwość elastycznej definicji kryteriów i celów pomiaru SIS oraz ocenę stopnia realizacji zakładanych celów. W pracy zaproponowano autorski indeks pomiaru SIS oparty na wybranych technikach MCDA.

Cykl publikacji (12), (13), (18) stanowi próbę budowy nowej metody oceny jakości serwisów internetowych. Stosowane w literaturze metody oceny serwisów internetowych jak eQual, Ahn, czy SiteQual oparte są na trywialnym aparacie matematycznym. Najczęściej stosowana jest w nich średnia arytmetyczna, służąca do agregacji ocen. Warto zauważyć, że zastosowanie metod MCDA do oceny serwisów internetowych niesie większy potencjał niż tylko konstrukcja rankingu. Przedstawiona w pracach (12), (13), (18) autorska metoda Pequal oparta jest na funkcji dostosowania jakości (Quality Function Deployment) oraz algorytmach wybranych metod MCDA. Funkcja dostosowania jakości jest ustrukturalizowanym procesem zapewniającym środki identyfikacji i dostarczającym opinii użytkowników o jakości produktu na kolejnych etapach jego tworzenia. Metodycznie, w pracach wykorzystano metody Promethee II, TOPSIS, AHP oraz COMET, również w wersjach rozmytych. Podczas badań, oprócz starannie zgromadzonych danych ankietowych, dążąc do obiektywizacji eksperckiej oceny, rozszerzono zbiór danych wejściowych o dane pomiarowe z urządzeń typu eye-tracker.

Prace (2), (3), (6), (8), (16), (28) koncentrują się na poszukiwaniu rozwiązań metodycznych i algorytmicznych wybranych problemów zrównoważonego rozwoju. Prace posadowione są w ważnych i aktualnych problematykach odnawialnych źródeł energii oraz zrównoważonego transportu i logistyki. Obszary te są szeroko dyskutowane w literaturze przedmiotu, a konkluzje autorów wyraźnie wskazują na potrzebę doskonalenia warsztatu istniejących modeli wspomagania tego procesu decyzyjnego. W tym zakresie publikacje te stanowią odpowiedź na aktualne wyzwania badawcze.

W pracy (3) opracowano nową metodę wielokryterialną PROSA. Przy jednoczesnym zachowaniu w niej uniwersalnych właściwości metody Promethee II, ograniczono w niej efekt liniowej kompensacji czynników. W wyniku przeprowadzonych badań dowiedziono, że metoda PROSA preferuje bardziej zrównoważone alternatywy decyzyjne spełniając paradygmat strong sustainability.

³W sekcji 7.1.3 numeracja literatury odnosi się do pozycji wykazanych w sekcji 7.1.2.

Grupa prac (2), (6), (8), (16), (28) posadowiona jest w obszarze zrównoważonego transportu i logistyki. W pracy (2) wykorzystano metody wielokryterialne w ocenie możliwości wdrożenia frachtowych, elektrycznych samochodów w obszarze logistyki miejskiej. Kontynuacją i dopełnieniem tej pracy jest publikacja (6), gdzie używając wielokryterialnej metody COMET dokonano identyfikacji pełnego modelu domenowego (obszar samochodów elektrycznych). W obszarze zrównoważonego zarządzania transportem posadowiona jest praca (8). Tutaj również dokonano porównania efektywności szeregu modeli wielokryterialnych. W pracach (16) oraz (28) podjęto się wielokryterialnego modelowania ważnego i aktualnego problemu oceny i doboru zrównoważonego dostawcy. Dokonano tutaj oceny efektywności wybranych metod MCDA.

W kolejny nurt badań wpisują się prace (7), (9) oraz (10). Obejmują one obszar rozwoju nowych metod MCDA. W szczególności prace dotyczą adaptacji arytmetyki przedziałowej czy kolejnych rozwinięć arytmetyki rozmytej (hesistant fuzzy sets) w wybranych metodach wielokryterialnych.

Dodatkowo, prowadzone były prace powiązane z dystrybucją faktur elektronicznych (14), (29), oceną jakości wybranych wyszukiwarek obrazów (15), oceną intensywności interaktywnego przekazu reklamowego (23), jak też budową wielokryterialnych modeli optymalizacji linii produkcyjnej (26), (27).

7.1.4. Pozostałe

- Nagroda Rektora ZUT w Szczecinie za wybitne osiągnięcia naukowe stopnia drugiego za rok 2018.
- Nagroda Best Paper Award za publikację [41] na konferencji FedCSIS 2017.
- Publikacja [42] została wyróżniona spośród wszystkich 248 pozycji poprzez umieszczenie na okładce wydania czasopisma *Energies, Volume 10, Issue 11 (November 2017)*.
- Publikacja A2 od 2019 roku znajduje się na liście najczęściej pobieranych artykułów otwartych czasopisma Omega The International Journal of Management Science wydawnictwa Elsevier o wskaźniku IF 5.324. Ponadto została oznaczona jako "Highly Cited Paper" i "Hot Paper" w bazie WoS.
- Uczestnictwo i wygłoszenie referatów na międzynarodowych konferencjach ASONAM 2018 (The 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining) w Barcelonie i FedCSIS (14th Federated Conference on Computer Science and Information Systems) w Lipsku.
- Funkcja Guest Editor w Special Issue Exploring of Sustainable Supplier Selection w 2018 roku w czasopiśmie Sustainability wydawnictwa MDPI – IF 2.576.
- Zaproszenie do pełnienia funkcji Guest Editor w Special Issue Production Line Optimization and Sustainability w 2021 roku w czasopiśmie Sustainability wydawnictwa MDPI – IF 2.576.
- Trzykrotny finalista Ogólnopolskiej Olimpiady Języka Angielskiego Dla Studentów Wyższych Uczelni Technicznych w latach 2006, 2007 i 2009.
- W roku 2007 uczestnictwo w finale regionalnym konkursu ACM Central European Programming Contest w Pradze.

- W roku 2011 I miejsce w konkursie e-point na najlepsze prace dyplomowe z zakresu rozwiązań internetowych realizowanych w technologii JAVA.

7.2. Dorobek dydaktyczny

7.2.1. Kursy i sylabusy

W ramach dotychczasowej pracy dydaktycznej prowadziłem zajęcia na poziomie S1 zgodnie z wykazem:

1. Programowanie systemów i aplikacji internetowych
– *cykl 15 wykładów i 15 laboratoriów*
2. Systemy i platformy biznesu elektronicznego
– *cykl 8 wykładów i 15 laboratoriów*
3. Systemy i platformy biznesu cyfrowego
– *cykl 15 wykładów i 15 laboratoriów*
4. Programowanie dokumentów dynamicznych
– *cykl 8 wykładów*
5. Aplikacje internetowe 1
– *cykl 15 wykładów i 15 laboratoriów*
6. Pracownia dyplomowa
– *cykl 8 seminariów*
7. Inżynierski projekt zespołowy
– *cykl 4 wykładów i 15 spotkań projektowych*
8. Wprowadzenie do informatyki
– *wyłoszenie 3 wykładów w ramach większego cyklu prowadzonego przez wielu wykładowców*

Ponadto współtworzyłem sylabusy do następujących kursów na poziomie S1:

1. Aplikacje internetowe 1
2. Aplikacje internetowe 2
3. Aplikacje internetowe

7.2.2. Prace dyplomowe

Podczas dotychczasowej pracy dydaktycznej byłem promotorem następujących prac inżynierskich:

1. Mobilny system informacji dla pasażerów komunikacji miejskiej
praca inżynierska, obrona w 2018 r.
2. System rejestracji i zarządzania wizytami w klinikach weterynaryjnych
praca inżynierska, obrona w 2019 r.

3. Aplikacja webowa wspierająca użycie fiszek w procesie nauki i zapamiętywania
praca inżynierska, obrona w 2019 r.
4. Wyszukiwarka gier wideo
praca inżynierska, obrona w 2020 r.
5. Internetowy system zarządzania i wynajmu mieszkań
praca inżynierska, obrona w 2021 r.
6. System do zarządzania i współdzielenia ulubionych witryn WWW
praca inżynierska, obrona w 2021 r.
7. Aplikacja mobilna do wyszukiwania najkorzystniejszej stacji benzynowej w okolicy
praca inżynierska, obrona planowana w II kwartale 2021 r.
8. Mobilna aplikacja informująca pasażerów komunikacji miejskiej o wydarzeniach na drodze
praca inżynierska, obrona planowana w II kwartale 2021 r.
9. Projekt i implementacja w wersji mobilnej dynamicznego modelu wielokryterialnego do wspomagania decyzji wyboru produktów konsumenckich na przykładzie samochodów
praca magisterska, obrona planowana w II kwartale 2021 r.
10. Projekt i implementacja systemu wspomagającego trwałe zapamiętywanie obcojęzycznego słownictwa z wykorzystaniem teorii krzywej zapominania Hermanna Ebbinghausa
praca inżynierska, obrona planowana w III kwartale 2021 r.
11. Projekt i implementacja dwuwymiarowej gry online Okręty z wykorzystaniem technologii webowych i API
praca inżynierska, obrona planowana w 2022 r.
12. System do webowej i mobilnej nauki języków obcych
praca inżynierska, obrona planowana w 2022 r.
13. Projekt i implementacja trójwymiarowej gry strategicznej
praca inżynierska, obrona planowana w 2022 r.
14. System do ewidencjonowania czasu pracy oraz zarządzania pracownikami w przedsiębiorstwie
praca inżynierska, obrona planowana w 2022 r.
15. Projekt i implementacja systemu gromadzenia i przetwarzania statystyk z gier komputerowych
praca inżynierska, obrona planowana w 2022 r.
16. Internetowy system do agregacji i analizy treningów sportowych
praca inżynierska, obrona planowana w 2022 r.

7.3. Dorobek organizacyjny

— Jestem aktywnym członkiem Association for Computing Machinery (ACM) od 2008 roku.

- W latach 2007-2009 byłem prezesem koła naukowego TWIPS działającego na Wydziale Informatyki ZUT w Szczecinie.
- W ramach prowadzonej działalności gospodarczej prowadzonej pod nazwą IdeaSpot Artur Karczmarczyk przeprowadziłem praktyki studenckie dla 5 osób oraz praktyki absolwenckie dla 8 osób w latach 2017-2021.
- Jestem członkiem Kolegium Wydziału Informatyki ZUT w Szczecinie.
- Jestem przewodniczącym sesji *WS01: 4th Symposium on Information Systems and Technologies for Management and Economy* na konferencji *25th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems 2021* z ponad 20 zgłoszonymi i recenzowanymi publikacjami.
- Uczestniczyłem we współpracy pomiędzy Technoparkiem Pomerania i Wydziałem Informatyki ZUT w Szczecinie w tworzeniu bloku przedmiotów obieralnych Komercjalizacja.

7.4. Dorobek zawodowy

7.4.1. Historia zatrudnienia

2006-2014 Unizeto Technologies S.A.

ścieżka kariery obejmująca stanowiska kwalifikowanego podwykonawcy, wdrożeniowca, programistę, skończywszy na inżynierze oprogramowania w dziale kierowania projektami

2014-obecnie IdeaSpot Artur Karczmarczyk

właściciel jednoosobowej działalności gospodarczej, w ramach której w szczytowym momencie zatrudnione było 12 osób

2017-obecnie Wydział Informatyki ZUT w Szczecinie

pracownik naukowo-dydaktyczny na stanowisku asystent

7.4.2. Najciekawsze projekty

- Współtworzyłem zaawansowany sklep internetowy wraz z czytelnią online norm dla Polskiego Komitetu Normalizacyjnego.
- Uczestniczyłem w tworzeniu prototypu Internetowego Konta Pacjenta – program pilotażowy dla grupy pacjentów diabetycznych z południa Polski.
- Współtworzę system anonimowej eskalacji problemów związanych z mobbingiem wykorzystywany m.in. w brytyjskim NHS czy francuskim Deezer.
- Kieruję wytwarzaniem oprogramowania dla platformy esport wyposażonej między innymi w systemy zaproszeń znajomych czy komunikator tekstowy.

7.4.3. Najciekawsze szkolenia i certyfikaty zawodowe

- Zend Certified PHP Engineer

- Szkolenie Programowanie Zespołowe w ramach Microsoft IT Academy Programme
- Certyfikat REQB - Certyfikowany Profesjonalista Inżynierii Wymagań
- Szkolenie w ramach Uniwersytetu Ernst&Young
- CS169.1x: Software as a Service (BerkeleyX, The University of California, Berkeley through edX)
- Introducing Agile Software Development
- Managing Agile Software Development
- Planning Agile Software Development
- Create Work Breakdown Structure (PMBOK Guide Fourth Edition)
- Estimating Activity Resources and Durations (PMBOK Guide Fourth Edition)
- Managing Projects within Organisations (PMBOK Guide Fourth Edition)
- Project Management Overview (PMBOK Guide Fifth Edition)
- Project Management Process Groups

Spis rysunków

3.1.	Wizualizacja powiązań pomiędzy poszczególnymi publikacjami A1–A10	7
4.1.	Grafy przedstawiające przykład 16-węzłowych sieci syntetycznych: A) BA, B) ER, C) WS; oraz rozkład stopni węzłów przykładowych 2000-węzłowych sieci syntetycznych: D) BA, E) ER, F) WS.	10
4.2.	Schemat poglądowy podejścia planowania i ewaluacji procesów rozprzestrzeniania informacji w sieciach społecznych.	13
4.3.	Analiza wrażliwości rankingu dla 20 najlepszych strategii z ewaluacji metodą TOPSIS na próbkach sieci rzeczywistej [29]. A1-A5 – sieć 10%, B1-B5 – sieć 30%, C1-C5 – sieć 50%.	16
4.4.	Stosunek zasięgu próbek do sieci rzeczywistej [29]. A: uporządkowane według przypadku symulacji, B: uporządkowane według stosunku zasięgu rosnąco, C: zgrupowane i uporządkowane według SF, D: zgrupowane i uporządkowane według PP, E: uporządkowane według rosnącej wartości zasięgu w sieci rzeczywistej.	18
4.5.	Schemat poglądowy proponowanego podejścia do wielokryterialnego doboru wielkości próbki przy zmiennych preferencjach inicjatora procesu.	19
4.6.	Analiza wizualna GAIA dla wyboru wielkości próbki sieci rzeczywistej. A - zwykła funkcja preferencji (ang. usual preference function). B - liniowa funkcja preferencji (ang. linear preference function).	20
4.7.	Schemat poglądowy proponowanego podejścia do sekwencyjnej inicjalizacji węzłów z wykorzystaniem symulacji w sieciach syntetycznych.	21
4.8.	Wpływ A) liczby iteracji, B) interwału pomiędzy iteracjami inicjalnego zasilania sieci informacjami na zasięg i czas trwania kampanii na przykładzie sieci rzeczywistej Gnutella [29]	22
4.9.	Przykładowe wyniki działania algorytmów generacji wektora P_i dla rozkładów A jednostajnego (ang. uniform), B proporcjonalnego (ang. proportional), C odwrotnie proporcjonalnego (ang. reversed proportional), D normalnego (ang. Gaussian), E geometrycznego (ang. geometric) i F odwróconego geometrycznego (ang. reversed geometric) przy założonym średnim prawdopodobieństwie propagacji równym 1 – 0.1, 2 – 0.5 i 3 – 0.9.	24
4.10.	Przykładowy proces rozprzestrzeniania informacji z rozkładem A jednorodnym (ang. uniform); B normalnym (ang. Gaussian); C proporcjonalnym (ang. proportional); i D odwrotnie proporcjonalnym (ang. reversed proportional) prawdopodobieństwa propagacji informacji.	25
4.11.	Schemat poglądowy proponowanego zrównoważonego podejścia do realizacji pojedynczej kampanii realizującej n celów zamiast n kampanii realizujących pojedyncze cele.	28
5.1.	Przykładowy kod źródłowy prostego scenariusza symulacji procesu rozprzestrzeniania informacji w sieci złożonej z wykorzystaniem proponowanego środowiska [40].	30

Spis tablic

4.1.	Macierz korelacji pomiędzy rankingami strategii obliczonymi na sieci rzeczywistej i jej próbkach.	17
4.2.	Przedstawienie wartości prawdopodobieństwa dla wszystkich węzłów sieci [33] dla średniego prawdopodobieństwa propagacji wynoszącego 0.2	24
7.1.	Profile internetowe (stan na dzień 6.05.2021).	33

Bibliografia

- [1] R. Hanna, A. Rohm, and V. L. Crittenden, “We’re all connected: The power of the social media ecosystem,” *Business horizons*, vol. 54, no. 3, pp. 265–273, 2011.
- [2] D. J. Watts, J. Peretti, and M. Frumin, *Viral marketing for the real world*. Harvard Business School Pub., 2007.
- [3] J. L. Iribarren and E. Moro, “Impact of human activity patterns on the dynamics of information diffusion,” *Physical review letters*, vol. 103, no. 3, p. 038702, 2009.
- [4] J. Berger and K. L. Milkman, “What makes online content viral?,” *Journal of marketing research*, vol. 49, no. 2, pp. 192–205, 2012.
- [5] J. Y. Ho and M. Dempsey, “Viral marketing: Motivations to forward online content,” *Journal of Business research*, vol. 63, no. 9-10, pp. 1000–1006, 2010.
- [6] X. Zhang, D.-D. Han, R. Yang, and Z. Zhang, “Users’ participation and social influence during information spreading on twitter,” *PloS one*, vol. 12, no. 9, p. e0183290, 2017.
- [7] O. Hinz, B. Skiera, C. Barrot, and J. U. Becker, “Seeding strategies for viral marketing: An empirical comparison,” *Journal of Marketing*, vol. 75, no. 6, pp. 55–71, 2011.
- [8] Y. Liu-Thompkins, “Seeding viral content: The role of message and network factors,” *Journal of Advertising Research*, vol. 52, no. 4, pp. 465–478, 2012.
- [9] C. Kiss and M. Bichler, “Identification of influencers: measuring influence in customer networks,” *Decision Support Systems*, vol. 46, no. 1, pp. 233–253, 2008.
- [10] M. Bampo, M. T. Ewing, D. R. Mather, D. Stewart, and M. Wallace, “The effects of the social structure of digital networks on viral marketing performance,” *Information systems research*, vol. 19, no. 3, pp. 273–290, 2008.
- [11] S. Stieglitz and L. Dang-Xuan, “Emotions and information diffusion in social media: sentiment of microblogs and sharing behavior,” *Journal of management information systems*, vol. 29, no. 4, pp. 217–248, 2013.
- [12] A. Dobeles, A. Lindgreen, M. Beverland, J. Vanhamme, and R. Van Wijk, “Why pass on viral messages? because they connect emotionally,” *Business Horizons*, vol. 50, no. 4, pp. 291–304, 2007.
- [13] C. Camarero and R. San José, “Social and attitudinal determinants of viral marketing dynamics,” *Computers in Human Behavior*, vol. 27, no. 6, pp. 2292–2300, 2011.
- [14] S. Helm, “Viral marketing-establishing customer relationships by ‘word-of-mouse’,” *Electronic markets*, vol. 10, no. 3, pp. 158–161, 2000.
- [15] K. Kandhway and J. Kuri, “How to run a campaign: Optimal control of sis and sir information epidemics,” *Applied Mathematics and Computation*, vol. 231, pp. 79–92, 2014.

- [16] D. Kempe, J. Kleinberg, and É. Tardos, “Maximizing the spread of influence through a social network,” in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 137–146, ACM, 2003.
- [17] E. M. Rogers, *Diffusion of innovations*. Simon and Schuster, 2010.
- [18] D. Cruz and C. Fill, “Evaluating viral marketing: isolating the key criteria,” *Marketing Intelligence & Planning*, vol. 26, no. 7, pp. 743–758, 2008.
- [19] A. Mochalova and A. Nanopoulos, “A targeted approach to viral marketing,” *Electronic Commerce Research and Applications*, vol. 13, no. 4, pp. 283–294, 2014.
- [20] H. T. Nguyen, T. N. Dinh, and M. T. Thai, “Cost-aware targeted viral marketing in billion-scale networks,” in *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*, pp. 1–9, IEEE, 2016.
- [21] S. Wasserman and K. Faust, *Social network analysis: Methods and applications*, vol. 8. Cambridge university press, 1994.
- [22] A.-L. Barabási and R. Albert, “Emergence of scaling in random networks,” *science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [23] D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks,” *nature*, vol. 393, no. 6684, p. 440, 1998.
- [24] P. ERdS and A. R&WI, “On random graphs i,” *Publ. Math. Debrecen*, vol. 6, pp. 290–297, 1959.
- [25] J.-P. Brans and B. Mareschal, “Promethee methods,” in *Multiple criteria decision analysis: state of the art surveys*, pp. 163–186, Springer, 2005.
- [26] J. Wątróbski, J. Jankowski, P. Ziemia, A. Karczmarczyk, and M. Ziolo, “Generalised framework for multi-criteria method selection: Rule set database and exemplary decision support system implementation blueprints,” *Data in brief*, vol. 22, p. 639, 2019.
- [27] S. Kullback and R. A. Leibler, “On information and sufficiency,” *Ann. Math. Statist.*, vol. 22, pp. 79–86, 03 1951.
- [28] M. E. Newman, “The structure of scientific collaboration networks,” *Proceedings of the national academy of sciences*, vol. 98, no. 2, pp. 404–409, 2001.
- [29] M. Ripeanu, I. Foster, and A. Iamnitchi, “Mapping the Gnutella Network: Properties of Large-Scale Peer-to-Peer Systems and Implications for System Design,” *arXiv:cs/0209028*, Sept. 2002. arXiv: cs/0209028.
- [30] C.-L. Hwang and K. Yoon, “Methods for multiple attribute decision making,” in *Multiple attribute decision making*, pp. 58–191, Springer, 1981.
- [31] “Snowball Sampling Function - R Documentation.”
- [32] J. Jankowski, P. Bródka, P. Kazienko, B. K. Szymanski, R. Michalski, and T. Kajdaniowicz, “Balancing speed and coverage by sequential seeding in complex networks,” *Scientific reports*, vol. 7, no. 1, p. 891, 2017.
- [33] K. E. Read, “Cultures of the central highlands, new guinea,” *Southwestern Journal of Anthropology*, pp. 1–43, 1954.
- [34] G. Voss, A. Godfrey, and K. Seiders, “Do satisfied customers always buy more? the roles of satiation and habituation in customer repurchase,” *Marketing Science Institute Working Paper Series 2010*, pp. 10–101, 2010.

- [35] R. Likert, "A technique for the measurement of attitudes.," *Archives of psychology*, 1932.
- [36] T. L. Saaty and L. G. Vargas, "The legitimacy of rank reversal," *Omega*, vol. 12, no. 5, pp. 513–516, 1984.
- [37] M. Zdrowia, "Program profilaktyki raka piersi (mammografia) - ministerstwo zdrowia - portal gov.pl," Apr 2018.
- [38] Gus, "Ludność. stan i struktura ludności oraz ruch naturalny w przekroju terytorialnym. stan w dniu 30 vi 2015 r.," Feb 2016.
- [39] B. J. Cox, "Object-oriented programming: an evolutionary approach," 1986.
- [40] A. Karczmarczyk, J. Jankowski, and J. Wątróbski, "Oonis—object-oriented network infection simulator," *SoftwareX*, vol. 14, p. 100675, 2021.
- [41] P. Ziemba, J. Wątróbski, A. Karczmarczyk, J. Jankowski, and W. Wolski, "Integrated approach to e-commerce websites evaluation with the use of surveys and eye tracking based experiments," in *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pp. 1019–1030, IEEE, 2017.
- [42] P. Ziemba, J. Wątróbski, M. Ziolo, and A. Karczmarczyk, "Using the prosa method in offshore wind farm location problems," *Energies*, vol. 10, no. 11, p. 1755, 2017.