Załącznik 3 – Autoreferat

1 Imię i nazwisko

Radosław Witold Michalski

2 Posiadane dyplomy, stopnie naukowe lub artystyczne – z podaniem podmiotu nadającego stopień, roku ich uzyskania oraz tytułu rozprawy doktorskiej

2.1 **Stopień naukowy doktora w dziedzinie nauk technicznych w dyscyplinie naukowej informatyka**, nadany uchwałą **Rady Wydziału Informatyki i Zarządzania Politechniki Wrocławskiej** z dnia 25 listopada 2014 roku. Tytuł rozprawy doktorskiej: *„Maximizing the Spread of Influence in Temporal Social Networks"* (praca napisana w języku angielskim). Promotorem rozprawy był prof. dr hab. inż. Przemysław Kazienko, jej recenzentami prof. Bolesław Szymański (Rensselaer Polytechnic Institute) oraz dr hab. inż. Mikołaj Morzy (Politechnika Poznańska). Studia doktoranckie na wyżej wymienionym wydziale realizowałem w latach 2010–2014.

2.2 **Dyplom magistra inżyniera informatyka** (2006 rok) – ukończone jednolite studia magisterskie na kierunku *Informatyka* na **Wydziale Informatyki i Zarządzania Politechniki Wrocławskiej**, na specjalności *Systemy i sieci informatyczne*. Tytuł pracy magisterskiej: *„Weryfikacja tożsamości elektronicznej online"*.

3 Informacja o dotychczasowym zatrudnieniu w jednostkach naukowych lub artystycznych

3.1 W latach 2005–2011 praca zawodowa – **Hutmen S.A. we Wrocławiu**, początkowo jako **Administrator bazy Oracle i systemu ERP**, następnie jako **Kierownik Działu Informatyki**.

3.2 Równolegle do powyższego, w latach 2006–2009 **prowadziłem zajęcia dydaktyczne na kierunku informatyka w Wyższej Szkole Informatyki i Zarządzania „Copernicus"** we Wrocławiu.

3.3 W okresie od 6 lutego 2012 roku do 30 września 2015 roku byłem zatrudniony **na Politechnice Wrocławskiej, na Wydziale Informatyki i Zarządzania**. W latach 2012–2014 pracowałem **w Instytucie Informatyki** na stanowisku **asystenta naukowego do spraw realizacji projektu badawczego**, początkowo w projekcie *„Eksploracja danych w złożonych społecznych systemach sieciowych"*, następnie w projekcie *„Translation Research and Patient Safety in Europe (TRANSFoRm)"* – kierownikiem obu projektów był prof. dr hab. inż. Przemysław Kazienko. W latach 2014–2015 pracowałem zaś na stanowisku **asystenta naukowo-dydaktycznego w Katedrze Inteligencji Obliczeniowej**.

3.4 Od 1 października 2015 roku jestem zatrudniony na stanowisku **adiunkta na Politechnice Wrocławskiej, na Wydziale Informatyki i Zarządzania, w Katedrze Inteligencji Obliczeniowej**, będącej do końca 2020 roku pod kierownictwem prof. dr. hab. inż. Haliny Kwaśnickiej, zaś od stycznia 2021 roku kierowaną przez dr. hab. inż. Tomasza Kajdanowicza, prof. PWr.

3.5 Witryna internetowa
https://www.ii.pwr.edu.pl/~michalski
https://rmhere.github.io (adres alternatywny)

3.6 Identyfikatory w bazach

- ORCID: **0000-0002-0106-655X**
- ResearcherID: **I-6978-2013**
- Scopus: **42862111400**
- ResearchGate: **Radoslaw-Michalski-2**
- Loop/Frontiers: **143086**
- Google Scholar: **xMVnv3sAAAAJ**

4 Omówienie osiągnięć, o których mowa w art. 219 ust. 1 pkt. 2 ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2020 r. poz. 85 z późn. zm.)

4.1 Charakter osiągnięcia naukowego

Jako osiągnięcie naukowe w dyscyplinie *informatyka techniczna i telekomunikacja* przedkładam cykl publikacji naukowych powiązanych tematycznie zatytułowany **„Analiza sieci społecznych: próbkowanie, klasyfikacja i dyfuzja"**. Publikacje składające się na cykl wymienione są w punkcie 4.2, natomiast omówienie celu naukowego cyklu i osiągniętych wyników znajduje się w punkcie 4.4.

Cykl obejmuje czternaście publikacji, z czego jedenaście to publikacje czasopismowe, natomiast trzy – konferencyjne. Podstawowym zagadnieniem badawczym podejmowanym w cyklu są metody z obszaru dyfuzji w sieciach społecznych, w szczególności wybór węzłów początkowych oraz algorytmy aktywacji węzłów. Metody te pozwalają na uzyskanie zamierzonych efektów w procesach dyfuzji (redukcja budżetu, większy zasięg). Wsparciem do zaproponowanych metod są prace pozwalające na lepsze zrozumienie charakteru węzłów i interakcji w sieciach, głównie z obszaru próbkowania oraz klasyfikacji – ponieważ działania te podejmowane są z reguły przed realizacją procesów dyfuzji i wiedza uzyskana na tym etapie może być także wykorzystana w dyfuzji (w szczególności do wyboru węzłów początkowych), zostały one w opisie cyklu przedstawione wcześniej pod zbiorczym określeniem *wnioskowanie*.

Badania opisane w cyklu obejmują następujące zagadnienia:

- wnioskowanie w sieciach społecznych – opracowanie metod próbkowania, identyfikacji charakteru węzłów oraz sposobów kwantyfikacji stabilności interakcji w sieciach społecznych,
- metody wyboru węzłów początkowych w procesach dyfuzji – m.in. z wykorzystaniem algorytmów ewolucyjnych oraz rozwinięć opracowanych miar entropii do selekcji węzłów,
- metametody aktywacji węzłów w procesach dyfuzji – opracowanie algorytmów pozwalających na skuteczniejszą aktywację węzłów w procesach dyfuzji w trzech rodzinach sieci społecznych: sieciach statycznych, temporalnych i wielowarstwowych – metody te noszą zbiorczą nazwę *sekwencyjnej aktywacji węzłów*.

Dodatkowo w ramach prac zaproponowano i przebadano nowy model sieci temporalnej (CogSNet), który m.in. rozszerza zakres zastosowań modeli dyfuzji i metod wyboru

Profil ORCID: https://orcid.org/0000-0002-0106-655X
Profil ResearcherID/Publons: https://www.researcherid.com/rid/I-6978-2013
Profil Scopus: https://www.scopus.com/authid/detail.uri?authorId=42862111400
Profil ResearchGate: https://www.researchgate.net/profile/Radoslaw-Michalski-2
Profil Loop/Frontiers: https://loop.frontiersin.org/people/143086/overview
Profil Google Scholar: https://scholar.google.pl/citations?user=xMVnv3sAAAAJ

węzłów początkowych pierwotnie zaproponowanych tylko dla sieci statycznych o sieci temporalne przyczyniając się do zwiększenia precyzyjności modelowania. Ponadto, wnosząc wkład do lepszego zrozumienia procesów dyfuzji, przebadano także rzeczywiste procesy tego typu w sieciach wielowarstwowych.

Wyżej wymienione techniki były opracowane i poddane ewaluacji w różnych rodzajach sieci społecznych, zarówno ze względu na model: sieci statyczne, temporalne, wielowarstwowe, jak i ze względu na ich charakter: typowe sieci społeczne, sieci w organizacjach, jak i sieci zbudowane na bazie tzw. łańcucha bloków (ang. *blockchain*). Warto także podkreślić, że większość z opracowanych metod nie wyklucza ich zastosowania poza sieciami społecznymi – możliwe jest ich wykorzystanie w innych rodzajach sieci złożonych, takich jak np. sieci finansowe, sieci technologiczne, sieci biologiczne, czy też sieci transportowe.

Efektem wykorzystania ww. technik wnioskowania i dyfuzji w sieciach społecznych jest między innymi:

- pozyskiwanie dodatkowej wiedzy o węzłach w sieciach społecznych, niezawartej wprost w atrybutach węzłów i krawędziach sieci,
- skuteczniejszy wybór węzłów do aktywacji w procesach dyfuzji, np. w procesie wpływu społecznego,
- efektywniejszy sposób aktywacji węzłów w procesach dyfuzji – metoda sekwencyjnej aktywacji węzłów pozwalająca na uzyskanie większego zasięgu lub redukcję budżetu,
- adekwatniejsze modelowanie sieci temporalnych, umożliwiające zastosowanie w nich wybranych modeli dyfuzji opracowanych pierwotnie dla sieci statycznych,
- lepsze zrozumienie dynamiki kontaktów i procesów dyfuzji oraz dokładniej odwzorowujące rozkłady miar sieciowych próbkowanie węzłów w sieciach społecznych.

Poniżej zamieszczam informacje uzupełniające dotyczące cyklu publikacji przedstawionego w punkcie 4.2 oraz jego bibliometrycznego podsumowania znajdującego się w punkcie 4.3:

- współczynnik wpływu *impact factor* określa wartość IF obowiązującą dla czasopisma w roku opublikowania pracy,
- liczba punktów z wykazu czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych Ministra Nauki i Szkolnictwa Wyższego (obecnie Ministra Nauki i Edukacji), nazywanego dalej *wykazem MEiN*, podawana jest podwójnie: z wykazu obowiązującego w momencie opublikowania pracy oraz z wykazu obowiązującego w roku 2021 – zawsze ze wskazaniem, o który wykaz chodzi,
- ranking Core publikacji konferencyjnych podany jest jako obowiązujący w momencie wydania publikacji,
- wszystkie publikacje wchodzące w skład niniejszego cyklu zostały opublikowane w czasopismach i materiałach konferencyjnych przypisanych w wykazie MEiN do dyscypliny *informatyka techniczna i telekomunikacja*,
- liczba cytowań podana jest na dzień 23 kwietnia 2021 roku,
- liczba cytowań jest przedstawiana w postaci Google Scholar / Scopus / Web of Science; przykładowo, wartość 20/10/8 oznacza, że praca ma dwadzieścia cytowań w bazie Google Scholar, dziesięć w bazie Scopus oraz osiem w bazie Web of Science; jeśli w którejś pozycji wskaźnika pojawia się „–", oznacza to, że pracy nie zaindeksowano w danej bazie, np. 20/10/– oznacza, że publikacja nie była zaindeksowana w Web of Science,
- wkład współautorski został posortowany według szacowanego procentowego wkładu malejąco,

- oświadczenia wszystkich współautorów prac przedłożonych jako główne osiągnięcie naukowe określające ich wkład oraz oszacowanie procentowego udziału w poszczególne publikacje stanowią załącznik 5 do niniejszego wniosku o przeprowadzenie postępowania habilitacyjnego.

4.2 Publikacje wchodzące w skład głównego osiągnięcia naukowego

Jako cykl powiązanych tematycznie publikacji wchodzących w skład głównego osiągnięcia naukowego przedkładam następujące czternaście prac:

[H1] Jarosław Jankowski, **Radosław Michalski**, Piotr Bródka, Przemysław Kazienko, and Sonja Utz. Knowledge acquisition from social platforms based on network distributions fitting. *Computers in Human Behavior*, 51:685–693, Elsevier, 2015.

- mój wkład (szacowany na 40%): opracowanie założeń metody, opracowanie algorytmów próbkowania, przeprowadzenie badań symulacyjnych, analiza wyników i opracowanie manuskryptu
- wkład procentowy współautorów: **RM 40%**, JJ 40%, SU 10%, PB 5%, PK 5%
- współczynnik wpływu IF: 2,880
- liczba punktów z wykazu MEiN (z roku publikacji): 40
- liczba punktów z wykazu MEiN (z roku 2021): 140
- liczba cytowań: 8/7/7

[H2] Jarosław Jankowski, Piotr Bródka, Przemysław Kazienko, Bolesław K. Szymański, **Radosław Michalski**, and Tomasz Kajdanowicz. Balancing speed and coverage by sequential seeding in complex networks. *Scientific reports*, 7(1):1–11, Nature Publishing Group, 2017.

- mój wkład (szacowany na 10%): opracowanie rozwinięć metody sekwencyjnej aktywacji węzłów, przygotowanie tekstu manuskryptu
- wkład procentowy współautorów: JJ 50%, PB 30%, **RM 10%**, PK 5%, BS 3%, TK 2%
- współczynnik wpływu IF: 4,122
- liczba punktów z wykazu MEiN (z roku publikacji): 40
- liczba punktów z wykazu MEiN (z roku 2021): 140
- liczba cytowań: 37/31/24

[H3] Jarosław Jankowski, **Radosław Michalski**, and Piotr Bródka. A multilayer network dataset of interaction and influence spreading in a virtual world. *Scientific data*, 4(1):1–9, Nature Publishing Group, 2017.

- mój wkład (szacowany na 33%): opracowanie założeń badania, analiza stanu wiedzy, przygotowanie danych, zaprojektowanie i przeprowadzenie badań, interpretacja wyników badań, przygotowanie tekstu manuskryptu
- wkład procentowy współautorów: JJ 34%, **RM 33%**, PB 33%
- współczynnik wpływu IF: 5,311
- liczba punktów z wykazu MEiN (z roku publikacji): 15
- liczba punktów z wykazu MEiN (z roku 2021): 140
- liczba cytowań: 11/9/8

[H4] Jarosław Jankowski, Bolesław K. Szymański, Przemysław Kazienko, **Radosław Michalski**, and Piotr Bródka. Probing limits of information spread with sequential seeding. *Scientific reports*, 8(1):1–9, Nature Publishing Group, 2018.

- mój wkład (szacowany na 35%): udział w opracowaniu koncepcji, analiza stanu wiedzy, przygotowanie części środowiska eksperymentalnego, udział w przeprowadzeniu badania, interpretacja wyników badań, przygotowanie tekstu manuskryptu
- wkład procentowy współautorów: **RM 35%**, JJ 35%, PB 20%, BS 5%, PK 5%
- współczynnik wpływu IF: 4,011
- liczba punktów z wykazu MEiN (z roku publikacji): 40
- liczba punktów z wykazu MEiN (z roku 2021): 140
- liczba cytowań: 17/10/6

[H5] Marcin Kulisiewicz, Przemysław Kazienko, Bolesław K. Szymański, and **Radosław Michalski**. Entropy measures of human communication dynamics. *Scientific reports*, 8(1):1–8, Nature Publishing Group, 2018.

- mój wkład (szacowany na 25%): współudział w opracowaniu koncepcji, analiza stanu wiedzy, analiza wyników, przygotowanie tekstu manuskryptu
- wkład procentowy współautorów: MK 40%, **RM 25%**, BS 25%, PK 10%
- współczynnik wpływu IF: 4,011
- liczba punktów z wykazu MEiN (z roku publikacji): 40
- liczba punktów z wykazu MEiN (z roku 2021): 140
- liczba cytowań: 11/5/4

[H6] Michał Weskida and **Radosław Michalski**. Finding influentials in social networks using evolutionary algorithm. *Journal of Computational Science*, 31:77–85, Elsevier, 2019.

- mój wkład (szacowany na 50%): opracowanie koncepcji podejścia, analiza stanu wiedzy, opracowanie planu eksperymentów, analiza wyników, przygotowanie tekstu manuskryptu
- wkład procentowy współautorów: **RM 50%**, MW 50%
- współczynnik wpływu IF: 2,644
- liczba punktów z wykazu MEiN (z roku publikacji): 100
- liczba punktów z wykazu MEiN (z roku 2021): 100
- liczba cytowań: 10/6/3

[H7] Mateusz Nurek and **Radosław Michalski**. Combining machine learning and social network analysis to reveal the organizational structures. *Applied Sciences*, 10(5):1699, MDPI, 2020.

- mój wkład (szacowany na 75%): opracowanie metody klasyfikacji kolektywnej, współudział w przygotowaniu koncepcji i metodyki badania, walidacja planu eksperymentów, analiza wyników, udział w opracowaniu tekstu manuskryptu
- wkład procentowy współautorów: **RM 75%**, MN 25%
- współczynnik wpływu IF: 2,474
- liczba punktów z wykazu MEiN (z roku publikacji): 70
- liczba punktów z wykazu MEiN (z roku 2021): 70
- liczba cytowań: 7/4/3

[H8] **Radosław Michalski**, Daria Dziubałtowska, and Piotr Macek. Revealing the character of nodes in a blockchain with supervised learning. *IEEE Access*, 8:109639–109647, IEEE, 2020.

- mój wkład (szacowany na 40%): opracowanie pomysłu badania, zaproponowanie metod esktrakcji wiedzy o węzłach, przygotowanie platformy do pozyskiwania sieci do badań, analiza wyników, udział w opracowaniu manuskryptu

- wkład procentowy współautorów: **RM 40%**, DD 30%, PM 30%
- współczynnik wpływu IF: 3,745
- liczba punktów z wykazu MEiN (z roku publikacji): 100
- liczba punktów z wykazu MEiN (z roku 2021): 100
- liczba cytowań: 2/2/1

[H9] **Radosław Michalski**, Jarosław Jankowski, and Piotr Bródka. Effective influence spreading in temporal networks with sequential seeding. *IEEE Access*, 8:151208–151218, IEEE, 2020.

- mój wkład (szacowany na 65%): zaproponowanie metody sekwencyjnej aktywacji w sieciach temporalnych, opracowanie środowiska eksperymentalnego, przeprowadzenie badań, analiza wyników, udział w opracowaniu tekstu manuskryptu
- wkład procentowy współautorów: **RM 65%**, PB 20%, JJ 15%
- współczynnik wpływu IF: 3,745
- liczba punktów z wykazu MEiN (z roku publikacji): 100
- liczba punktów z wykazu MEiN (z roku 2021): 100
- liczba cytowań: 4/3/2

[H10] Piotr Bródka, Jarosław Jankowski, and **Radosław Michalski**. Sequential seeding in multilayer networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 31(3):033130, AIP Publishing, 2021.

- mój wkład (szacowany na 25%): udział w opracowaniu koncepcji dla sieci wielowarstwowych i planu eksperymentów, weryfikacja wyników eksperymentów, udział w opracowaniu manuskryptu
- wkład procentowy współautorów: PB 55%, **RM 25%**, JJ 20%
- współczynnik wpływu IF: 2,832
- liczba punktów z wykazu MEiN (z roku publikacji): 140
- liczba punktów z wykazu MEiN (z roku 2021): 140
- liczba cytowań: 0/0/0

[H11] **Radosław Michalski**, Bolesław K. Szymański, Przemysław Kazienko, Christian Lebiere, Omar Lizardo, and Marcin Kulisiewicz. Social networks through the prism of cognition. *Complexity*, 2021:4963903, Wiley/Hindawi, 2021.

- mój wkład (szacowany na 60%): zaproponowanie modelu CogSNet, analiza stanu wiedzy, przygotowanie środowiska badawczego, przeprowadzenie eksperymentów i analiza wyników, udział w opracowaniu manuskryptu
- wkład procentowy współautorów: **RM 60%**, BS 15%, PK 10%, CL 5%, OL 5%, MK 5%
- współczynnik wpływu IF: 2,462
- liczba punktów z wykazu MEiN (z roku publikacji): 70
- liczba punktów z wykazu MEiN (z roku 2021): 70
- liczba cytowań: 4/0/0

[H12] Patryk Pazura, Kamil Bortko, Jarosław Jankowski, and **Radosław Michalski**. A dynamic vote-rank based approach for effective sequential initialization of information spreading processes within complex networks. In *ICCS 2020: International Conference on Computational Science*, vol. 12137 of *Lecture Notes in Computer Science*, pp. 638–651. Springer, Cham, 2020.

- mój wkład (szacowany na 30%): analiza stanu wiedzy, współudział w opracowaniu planu eksperymentów, analiza wyników, współudział w przygotowaniu tekstu manuskryptu
- wkład procentowy współautorów: PP 55%, **RM 30%**, KB 10%, JJ 5%
- ranking Core: A
- liczba punktów z wykazu MEiN (z roku publikacji): 140
- liczba punktów z wykazu MEiN (z roku 2021): 140
- liczba cytowań: 0/0/0

[H13] **Radosław Michalski**. Analysing and predicting the adoption of anonymous transactions in cryptocurrencies. In *BIS 2020: International Conference on Business Information Systems*, vol. 394 of *Lecture Notes in Business Information Processing*, pp. 132–144. Springer, Cham, 2020.

- mój wkład: **100%** - publikacja jednoautorska
- ranking Core: B
- liczba punktów z wykazu MEiN (z roku publikacji): 70
- liczba punktów z wykazu MEiN (z roku 2021): 70
- liczba cytowań: 1/1/0

[H14] **Radosław Michalski**, Jarosław Jankowski, and Patryk Pazura. Entropy-based measure for influence maximization in temporal networks. In *ICCS 2020: International Conference on Computational Science*, vol. 12140 of *Lecture Notes in Computer Science*, pp. 277–290. Springer, Cham, 2020.

- mój wkład (szacowany na 65%): opracowanie koncepcji metody, analiza stanu wiedzy, opracowanie planu eksperymentów, przeprowadzenie eksperymentów, analiza wyników, przygotowanie tekstu manuskryptu
- wkład procentowy współautorów: **RM 65%**, PP 30%, JJ 5%
- ranking Core: A
- liczba punktów z wykazu MEiN (z roku publikacji): 140
- liczba punktów z wykazu MEiN (z roku 2021): 140
- liczba cytowań: 0/0/0

4.3 Podsumowanie bibliometryczne cyklu

W tabeli 1 przedstawiam wskaźniki bibliometryczne publikacji wchodzących w skład osiągnięcia naukowego. Informacje uzupełniające dotyczące sposobu ustalenia wartości wskaźników opisałem w punkcie 4.1. W punkcie 4.5.6 zawarłem szersze podsumowanie swojego dorobku wykraczające poza wskaźniki bibliometryczne samego cyklu.

Tabela 1: Wskaźniki bibliometryczne publikacji wchodzących w skład osiągnięcia naukowego

Łączna liczba publikacji w cyklu	14
Liczba publikacji czasopismowych	11
Liczba publikacji konferencyjnych	3
Suma współczynnika Impact Factor	38,237
Suma punktów z wykazu MEiN obowiązującego w roku publikacji	1105
Suma punktów z wykazu MEiN obowiązującego w 2021 roku	1630
Liczba cytowań[1]	112/78/58
Liczba cytowań bez autocytowań wszystkich autorów[1]	56/38/33

[1] w ujęciu: Google Scholar/Scopus/Web of Science

Tabela 2: Przykłady wierzchołków i krawędzi w sieciach społecznych

Wierzchołek	Krawędź
osoba	znajomość
naukowiec	współautorstwo publikacji
uczeń/student	uczestnictwo w tych samych zajęciach
członek stowarzyszenia	wspólne działania z innym członkiem
użytkownik telefonu	połączenie telefoniczne i/lub wiadomość tekstowa
właściciel adresu w sieci blockchain	transfer środków lub innych zasobów
pracownik	wymiana informacji z innym pracownikiem

4.4 Omówienie celu naukowego ww. prac i osiągniętych wyników wraz z potencjalnymi zastosowaniami

4.4.1 Wprowadzenie

W niniejszej sekcji pokrótce podsumowano stan wiedzy w obszarze, którego dotyczy przedłożone osiągnięcie naukowe, aby umiejscowić je pośród zagadnień problemowych.

Jako sieci można traktować wiele różnych systemów, których elementy są ze sobą powiązane. Przykładem mogą być sieci komputerowe, gdzie poszczególne urządzenia (węzły sieci) wymieniają dane między sobą z wykorzystaniem medium transmisyjnego (krawędzie). Innym przykładem – rozważanym w pracach badawczych w ramach przedkładanego cyklu – są sieci społeczne, w których członkowie sieci (osoby) komunikują się, wchodzą w pewne interakcje (choćby poprzez spotkania czy nawiązywanie relacji), a także dołączają i odchodzą z tychże sieci. Typową sieć matematycznie można przedstawić jako graf $G(V, E)$, składający się z niepustego i skończonego zbioru wierzchołków V oraz skończonego zbioru krawędzi E, którego elementami są pary różnych wierzchołków ze zbioru V. Przyjęło się, że sieci, w których zależności zachodzące pomiędzy ich elementami nie są łatwe do opisania i zrozumienia, określane są mianem sieci złożonych [1, 16], a ich przykłady to choćby sieci społeczne, transportowe, biologiczne czy informacyjne. Wyróżnikiem sieci społecznych, których w znaczącej mierze dotyczy przedkładane osiągnięcie naukowe, jest fakt, że wierzchołkami tych sieci są osoby. W rzeczywistości może być jednak tak, jak to przedstawiono w tabeli 2, gdzie osoby te reprezentowane są przez np. numery telefonów, adresy e-mail czy adresy sieci blockchain. Jednak w takim przypadku wciąż sieć zbudowaną z danych pozyskanych z takich systemów interpretuje się i bada jak sieć społeczną, chyba że zachodzą okoliczności wymagające innej interpretacji charakteru węzłów.

Sieci złożone (w tym społeczne) długo wymykały się możliwościom opisu i zrozumienia ze względu na posiadanie nietrywialnych zależności topologicznych, szczególnie ze względu na charakter węzłów i relacji w tych sieciach. Jednak dzięki fundamentom położonym na przełomie lat pięćdziesiątych i sześćdziesiątych XX wieku związanym z modelowaniem sieci losowych [14, 18], a później także znaczącym dokonaniom w analizie krat (ang. *lattices*) [74], można było powiedzieć, że dokonano pewnego przełomu w modelowaniu i wnioskowaniu z grafów konstruowanych sztucznie. Jednak rzeczywiste systemy złożone, do których należą między innymi sieci biologiczne, technologiczne, informacyjne czy społeczne, ciągle cechują zależności, które trudno w prosty sposób opisać matematycznie i generalizować za pomocą modeli [3, 9]. Dostęp do ogromnej liczby danych

gromadzonych przez systemy informatyczne, przykładowo pochodzących z sieci telekomunikacyjnych czy Internetu, pokazał, że ciągle jest przestrzeń do proponowania nowych modeli. Modele te adekwatnie odzwierciedlają wiele rzeczywistych sieci, np. sieci społeczne, które mają charakter bezskalowy [1, 80] czy też *small-world* [76]. I choć w ostatnich latach aplikowalność niektórych opracowanych modeli została wystawiona na próbę [11, 20], najnowsze prace pokazują, że jednak całkiem dobrze reprezentują one strukturę rzeczywistych sieci [4, 24]. Można by więc stwierdzić, że teoretycznie powinniśmy dysponować wystarczającym aparatem narzędziowym z dziedziny modelowania oraz danymi, dzięki którym możemy parametryzować modele. Jednak mimo tego jesteśmy wciąż daleko od pełnego zrozumienia różnych procesów zachodzących w sieciach, przykładowo dyfuzji – w szczególności jeśli myślimy o takich wariantach sieci jak sieci temporalne lub wielowarstwowe. Warto podkreślić, że od niemal dwóch dekad jesteśmy w o tyle komfortowej sytuacji, że dane możliwe obecnie do pozyskania z wielu systemów informatycznych uzupełniają się z modelami, co pozwala na walidację modeli teoretycznych oraz tworzenie nowych na bazie obserwacji w danych. Stąd analiza sieci złożonych (a w szczególności społecznych) jest obecnie obszarem, który łączy m.in. informatyków, fizyków i socjologów, gdzie reprezentanci tych gałęzi nauki wnoszą swoje kompetencje do zrozumienia zjawisk występujących w tych sieciach.

Historycznie, strukturami, które były analizowane jako pierwsze w sieciach, były struktury statyczne. Są to sieci pozbawione warstw, w których też nie zachodziły zmiany w atrybutach węzłów oraz krawędzi, a także nie zmieniała się sama struktura sieci poprzez pojawianie się lub znikanie węzłów czy krawędzi. Dla sieci statycznych opracowano modele, metody i miary, które można traktować jako ugruntowane, choć, oczywiście, ich zbiór cały czas rośnie. Jednocześnie coraz większe znaczenie zyskują inne modele sieci, a mianowicie sieci temporalne [26] i wielowarstwowe [2]. W przypadku pierwszego z nich – sieci temporalnej – zakłada się zmienność struktury w czasie, natomiast drugi – sieć wielowarstwowa – uwzględnia wiele rodzajów krawędzi. Szczególnie intensywny rozwój analizy obu tych rodzin modeli to ostatnia dekada, gdyż coraz więcej prac naukowych wskazywało ograniczenia modelowania sieci jako sieci statycznych w przypadku zachodzących w tych sieciach dynamicznych procesów, takich jak badane w ramach cyklu procesy dyfuzji.

Oprócz samej struktury, także właśnie analiza procesów mających miejsce w sieciach stanowi przedmiot zainteresowania naukowców. Początkowo strukturą, w której rozpatrywano i modelowano te procesy, była typowa sieć statyczna. Tam też modelowane były takie procesy jak dyfuzja informacji [23], innowacji [58] czy rozprzestrzenianie się wpływu społecznego [53]. Jednym z istotnych problemów postawionych na początku XXI wieku był problem maksymalizacji wpływu w sieciach społecznych, który w klasycznym ujęciu dotyczył takiego wyboru węzłów początkowych do aktywacji, które w wybranym modelu wpływu uzyskają najlepszy efekt końcowy, tj. aktywują największą część sieci [40]. Formalnie ten problem optymalizacyjny można zapisać jak w równaniu 1:

$$S^* = \arg\max_{S \subseteq V} \sigma(S), \tag{1}$$

gdzie V oznacza zbiór wszystkich węzłów w sieci, S jest zbiorem węzłów początkowych (ang. *seed set*), $|S| = k$ to budżet na aktywację, $\sigma(\cdot)$ jest funkcją dyfuzji, a $\sigma(S)$ liczbą aktywowanych węzłów w procesie dyfuzji. Przy tak zdefiniowanym

problemie widać, że na zasięg aktywacji w sieci mają wpływ dwa zasadnicze elementy: (i) sposób wyboru zbioru S oraz (ii) model dyfuzji. Stąd na przestrzeni ostatnich kilkunastu lat badacze usiłowali zaproponować metody pozwalające skutecznie generować zbiór węzłów początkowych dla jednego lub wielu modeli dyfuzji. Próby rozwiązania tego problemu dla sieci rzeczywistych podsumowałem w pracy [69]. Jednak studiując to zagadnienie miało się wrażenie, że obszarem, który nie był wystarczająco zbadany, były inne warianty tego problemu przedstawione w [21], takie jak minimalizacja liczby iteracji lub budżetu.

Dodatkowo, niezależnie od wariantu, po wyborze węzłów początkowych aktywacja ich następowała od razu. To skutkowało tym, że część budżetu przeznaczonego do aktywacji była niepotrzebnie zużywana na węzły, które podlegały jej w sposób naturalny. W toku prac badawczych w ramach cyklu zaproponowano i przebadano nowe podejście, które jest więc odpowiedzią na ten problem – sekwencyjną aktywację węzłów. Niezależnie od metody wyboru węzłów początkowych S algorytm sekwencyjny określa sposób aktywacji prowadzący albo do zaoszczędzenia budżetu, albo zwiększenia ostatecznego zasięgu $\sigma(S)$. Warto dodać, że sekwencyjna aktywacja węzłów została przebadana wielowymiarowo w różnych rodzajach sieci społecznych (statyczne, temporalne, wielowarstwowe) i w ramach prac badawczych wskazano jakie są możliwości i ograniczenia metody także w odniesieniu do rozwiązań optymalnych. Ponadto problem ten najczęściej rozpatrywany jest w kategoriach sieci nieatrybutowanych, gdzie nie dysponujemy informacją o charakterze węzłów – stąd koszt pozyskania każdego węzła w sieci jest identyczny. Dopiero ostatnie dwa lata przyniosły interesujące pomysły jak optymalizować budżet w przypadku różnego kosztu aktywacji węzłów (ang. *budgeted influence maximisation*) [8, 55, 62]. Przez to wydaje się, że potrzeba pozyskania faktycznej informacji o charakterze węzłów w sieci, nie tylko ich pozycji w grafie wyrażonej jako wartość miary sieciowej, będzie przybierała na sile, gdyż pozwoli na lepszą alokację budżetu.

Powyżej opisane prace poruszały się w domenie sieci statycznych, jednak ostatnia dekada to także bardzo dynamiczny rozwój badań nad wspomnianymi już sieciami temporalnymi [25, 26] oraz wielowarstwowymi [2, 41]. Wynika to z faktu, że te sieci mimo wprowadzenia dodatkowej złożoności (czas lub warstwy) oferują często większe możliwości w zakresie modelowania procesów takich jak właśnie dyfuzja. Skutkuje to także stopniowym przenoszeniem wyzwań opisanych powyżej do domeny sieci temporalnych [49] i wielowarstwowych [7], gdzie struktura i dynamika sieci są mocno powiązane z dyfuzją [6, 13, 39, 60]. Badania nad tym zagadnieniem także stanowiły przedmiot zainteresowań grantu NCN SONATA, którego byłem kierownikiem (grant realizowany w latach 2016–2020).

Podsumowując stan wiedzy, można stwierdzić, że w sieciach społecznych ciągle występuje wiele problemów dalekich od rozwiązania. Wśród nich można wymienić brak kompletnej wiedzy o węzłach i interakcjach w sieciach, a także niepełną wiedzę o tym jak zachodzą procesy dyfuzji w tych sieciach oraz jak tę wiedzę skutecznie wykorzystać do osiągnięcia założonego celu. Jednocześnie, chcąc w lepszym stopniu modelować procesy dyfuzji, warto rozważyć inne modele sieci niż tylko sieci statyczne, np. sieci temporalne i wielowarstwowe.

4.4.2 Cel prac badawczych

Celem prac badawczych nawiązujących do problemów przedstawionych w poprzednim punkcie była chęć pogłębienia zrozumienia sieci społecznych, szczególnie w obszarze dyfuzji – prace te zaowocowały przedkładanym cyklem publikacji. Istotną potrzebą było znalezienie odpowiedzi na pytanie czy możliwe jest zaproponowanie metod, które pozwolą na skuteczniejsze realizowanie procesów dyfuzji. W efekcie opracowano szereg metod i algorytmów w obszarze dyfuzji w sieciach społecznych oraz w obszarach wspierających, tj. próbkowaniu, klasyfikacji i modelowaniu. Istotną nowością zaproponowanego cyklu jest rodzina metod określanych jako sekwencyjna aktywacja węzłów, która wykracza poza standardowe podejście w zagadnieniu maksymalizacji wpływu.

Jak opisano w punkcie 4.4.1, typowo problem maksymalizacji wpływu analizowany był w sieciach statycznych i zadanie to ograniczało się najczęściej do wariantu stałego budżetu i jego jednoczesnej alokacji do aktywowania węzłów początkowych (tzw. *seed set*). Próbując spojrzeć nieco szerzej na to zagadnienie, w ramach prac badawczych zadano pięć pytań, na które poszukiwano odpowiedzi w różnych modelach i przykładach sieci społecznych:

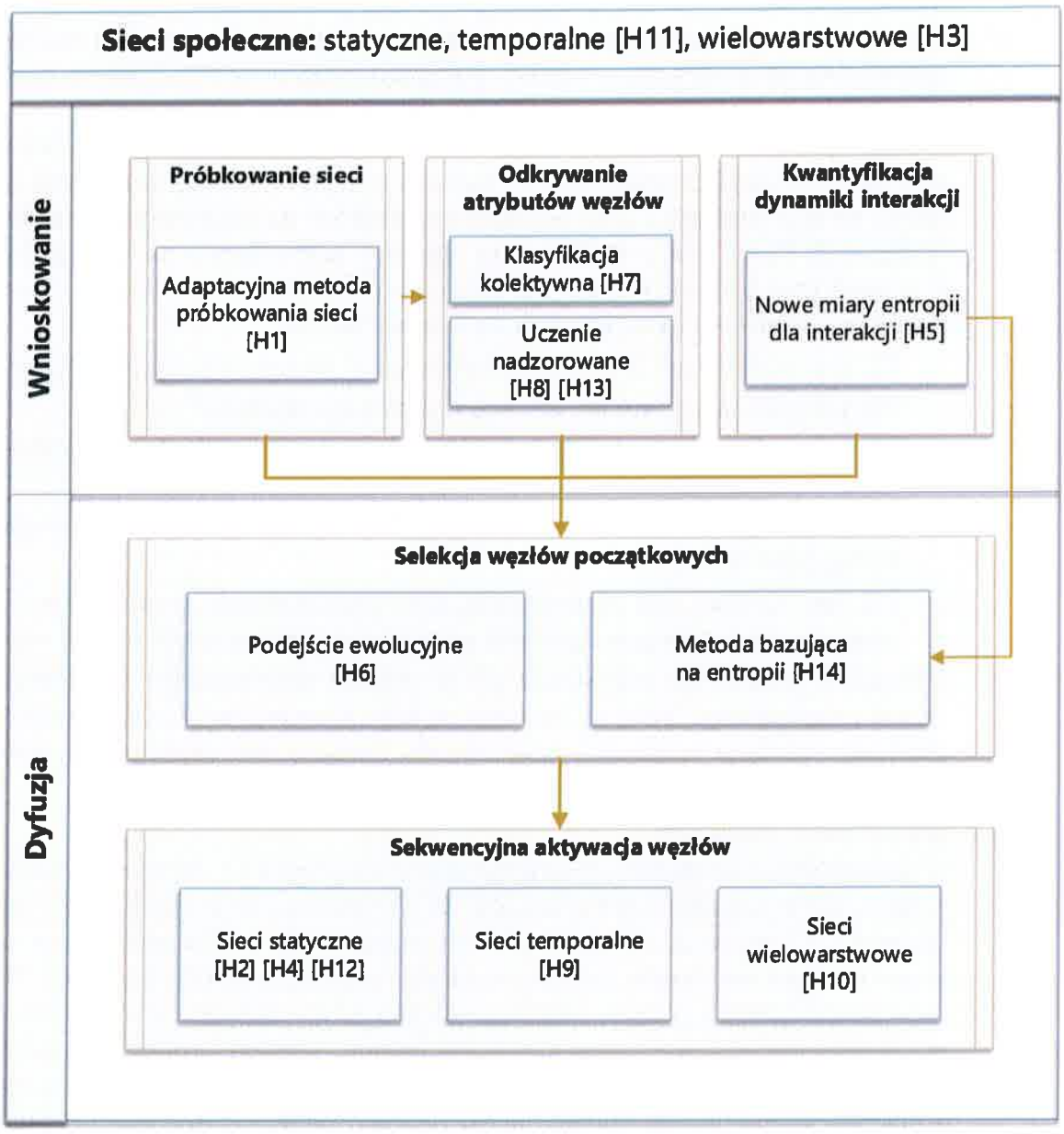
- P1: czy możliwe jest pozyskanie informacji o nieznanych atrybutach węzłów?
- P2: jak mierzyć dynamikę interakcji pomiędzy węzłami?
- P3: czy możliwe jest zaproponowanie nowych skutecznych algorytmów selekcji węzłów początkowych?
- P4: które algorytmy aktywacji węzłów w sieciach będą lepsze w wykorzystaniu dostępnego budżetu?
- P5: czy możliwe jest modelowanie sieci temporalnych, które będzie bliższe procesom kognitywnym i pozwoli na zastosowanie uznanych modeli wpływu?

Usiłowanie znalezienia odpowiedzi na te pytania doprowadziło do opracowania metod i algorytmów, które traktowane są jako poszczególne etapy procesu wnioskowania i dyfuzji w sieciach społecznych. Zostaną one omówione w kolejnym punkcie, odwołując się także do powyższej listy pytań badawczych.

4.4.3 Omówienie wyników

Schematycznie cykl publikacji przedstawiono na rysunku 1. W ramach badań poruszano się w ramach dwóch obszarów. W pierwszym z nich, określonym zbiorczo mianem *wnioskowania*, analizowano trzy zagadnienia: próbkowanie sieci, odkrywanie wartości atrybutów węzłów oraz kwantyfikację dynamiki interakcji. W drugim, istotniejszym z punktu widzenia całego cyklu – *dyfuzji* – dwa kolejne: selekcję węzłów początkowych oraz aktywację węzłów. Metody zaproponowane w tych obszarach zostaną opisane poniżej w odpowiednich sekcjach, przy czym metody z obszaru wnioskowania opisano przed dyfuzją, ponieważ działania te podejmowane są z reguły we wcześniejszych etapach, gdyż służą do lepszego poznania sieci społecznych.

Opracowane algorytmy dotyczą sieci społecznych reprezentujących trzy modele: sieci statyczne, sieci temporalne oraz sieci wielowarstwowe – stąd w górnej części rysunku 1 wymieniono również te modele wraz z dwiema publikacjami z cyklu, które wniosły wkład w obszar modelowania sieci. Należy mieć także na uwadze, że zaproponowane na rysunku 1 przypisanie prac cyklu do poszczególnych obszarów należy traktować jako określenie, do którego obszaru poszczególne prace najmocniej przynależą. Oprócz opisu prac znajdującego się w dalszej części tego punktu, w tabeli 3 przedstawiono zbiorcze ich podsumowanie.



Rysunek 1: Schemat powiązań koncepcji opracowanych w osiągnięciu naukowym.

Próbkowanie w sieciach społecznych

Wyniki wielu prac badawczych dowodzą, że dysponowanie atrybutami węzłów lub krawędzi w sieciach społecznych umożliwia opracowywanie znacznie skuteczniejszych metod, niezależnie czy chodzi o lepsze zrozumienie struktury sieciowej [12], modelowanie sieci [44], predykcję krawędzi [45], czy też identyfikację węzłów początkowych w procesach dyfuzji [82]. Metody opracowane dla sieci atrybutowanych (ang. *attributed networks*) posiadają mniejszą uniwersalność, ale, jak wskazują m.in. powyższe prace, osiągają lepsze efekty w poszczególnych zadaniach. Tym samym pierwszym wyzwaniem, które zostało podjęte w ramach prac badawczych prowadzących do przedłożonego cyklu, była odpowiedź na pytanie **P1**: czy możliwe jest pozyskanie informacji o nieznanych atrybutach węzłów?

W pierwszym obszarze badań mającym udzielić odpowiedzi na to pytanie skupiono się na zaproponowaniu takiego sposobu próbkowania sieci społecznych, który dostarczy wiedzy na temat atrybutów węzłów reprezentatywnych dla rozkładów miar sieciowych. Praca **[H1]** dotyczy adaptacyjnych metod próbkowania sieci społecznych i proponuje algorytm MSCS (Multistage Survey Control System), która ma stanowić wsparcie do pozyskiwania informacji w sieciach w przypadku brakującej wiedzy, np. o etykietach węzłów. Informacje te mogą być wykorzystane następnie właściwszego wyboru węzłów początkowych w dyfuzji, ale też np. do propagacji wiedzy w zadaniu klasyfikacji kolektywnej. Opracowana metoda MSCS ma za zadanie w taki sposób proponować węzły do próbkowania (np. ankietowania), aby w jak najkrótszym czasie doprowadzić do pozyskania informacji z reprezentatywnej próbki sieci, przy czym miara reprezentatywności może być dowolnie określana. Taką próbkę można także interpretować jako desygnowaną do odkrycia informacji (*utility score* w klasyfikacji kolektywnej). Przeprowadzone badania wykazały wysoką skuteczność metody, w szczególności gdy próbka musi być mała (np. gdy dysponujemy niskim budżetem na próbkowanie). Naturalnie, jak wskazano w pracy [75], budowa złożonych miar reprezentatywności bazujących na nieskorelowanych miarach sieciowych może okazać się trudnym zadaniem, jednak wyzwanie to wykracza poza samą ideę metody MSCS wzbogacającej rodzinę metod adaptacyjnego próbkowania sieci. Wnioski płynące z pracy **[H1]** stanowią zatem częściową odpowiedź na pytanie **P1**, gdyż w pracy zaproponowano podejście wymagające niższego kosztu pozyskania informacji o etykietach węzłów (lub innych atrybutów). Kontynuacją poszukiwania odpowiedzi na to pytanie zajęto się w pracach poświęconych klasyfikacji w sieciach społecznych opisanych poniżej.

Klasyfikacja

Jak już napisano w punkcie 4.4.1, sieci złożone (w tym społeczne) posiadają nietrywialne zależności topologiczne. W związku z tym, jeśli dysponujemy nawet cząstkową wiedzą o sieci, warto tę wiedzę wykorzystać. Ponadto w niektórych rodzinach sieci obserwowane są zjawiska wynikające z natury sieci. Przykładowo, w sieciach społecznych, występują homofilia oraz wpływ społeczny [47] – czasem ciężko rozróżnialne [43], jednak wiedza o ich istnieniu może także zostać wykorzystana do wnioskowania.

W ramach dalszych prac badawczych usiłujących odpowiedzieć na pytanie jak w sieciach społecznych wnioskować o nieznanych etykietach węzłów (**P1**), skupiono się na dwóch rodzajach sieci społecznych: sieciach organizacyjnych (ang. *organisational networks*) oraz sieciach bazujących na łańcuchu bloków (ang. *blockchain*). Prace cyklu obejmujące ten obszar to **[H7, H8, H13]** – zostaną one

omówione poniżej.

Klasycznym podejściem w uczeniu nadzorowanym jest przyjęcie założenia o zmiennych niezależnych o identycznym rozkładzie *(i.i.d.)*. Jednocześnie wiele danych pochodzących z systemów rzeczywistych nie zawsze spełnia to założenie, czego najlepszym przykładem mogą być sieci społeczne – choćby wspomniana wcześniej homofilia skutkuje migracją członków sieci społecznych do rejonów sieci, w której podziela się podobne poglądy. Tym samym wydaje się słusznym poszukiwanie innych sposobów na odkrywanie nieznanych atrybutów węzłów, a jednym z nich może być klasyfikacja kolektywna [50]. Istota klasyfikacji kolektywnej zakłada, że struktura sieciowa determinuje relacje pomiędzy zmiennymi losowymi, w przeciwieństwie właśnie do podejścia i.i.d., gdzie tych relacji się nie zakłada. W pracy [**H7**] zbadano możliwości klasyfikacji kolektywnej w odniesieniu do problemu odkrywania nieznanych atrybutów węzłów sieci w organizacji – w tym przypadku poziomu zajmowanego przez pracowników w hierarchii organizacyjnej. Wykorzystano do tego dwa zbiory danych, gdzie taka informacja była dostępna – jako źródła do budowy sieci społecznej wykorzystano komunikację e-mailową. W pracy zaproponowano algorytm klasyfikacji kolektywnej, który następnie poddano porównaniu z klasycznymi metodami uczenia nadzorowanego: drzewami decyzyjnymi, lasami losowymi, siecią neuronową i SVM. Badania wykazały podstawową przewagę zaproponowanego algorytmu klasyfikacji kolektywnej nad typowymi podejściami – lepsze radzenie sobie z klasami niezbalansowanymi. Jednocześnie zbadano także jak wygląda skuteczność wszystkich przebadanych algorytmów klasyfikacji w zależności od odsetka odkrytych etykiet oraz sposobu odkrywania tych etykiet (ang. *utility score*). Wnioski płynące z pracy mogą więc zostać wykorzystane przy adaptacyjnych metodach próbkowania sieci społecznych (np. właśnie w celu pozyskania etykiet), którymi zajmowano się w opisanej już pracy [**H1**].

Kolejne dwie prace z tej tematyki [**H8, H13**] ciągle poruszają się w domenie sieci społecznych, jednak sieci te były zbudowane na bazie innego źródła – łańcucha bloków (ang. *blockchain*). W obu pracach zmierzono się z zagadnieniem możliwości zastosowania uczenia nadzorowanego do wnioskowania o etykietach węzłów. W pracy [**H8**] dokonano etykietowania niemal 9 000 adresów w sieci Bitcoin do kategorii: górników (ang. *miners*), giełd, kopalni (ang. *mining pools*) i innych, aby następnie zbadać jakie metody uczenia nadzorowanego bazującego na zaproponowanych cechach utworzonych na bazie historii transakcji w tej sieci (bloków) będą w stanie radzić sobie najlepiej. Istotnym aspektem badania było także zbadanie długości historii transakcji na jakość klasyfikacji, co ma duży wpływ na czas przetwarzania rozbudowanego grafu transakcji. Dodatkowo w pracy przebadano możliwości metody osadzeń node2vec, która zyskuje w ostatnich latach sporą uwagę badaczy w zakresie możliwości wektoryzacji grafów [22]. Aby przebadać skuteczność podejścia node2vec, w pracy [**H8**] zaproponowano metodę budowy grafu z bloków, który to graf został wykorzystany do stworzenia wektorów osadzeń. Uzyskane wyniki mogą być wykorzystane do deanonimizacji typów węzłów w sieci typu blockchain, co szczególnie w trzech ostatnich latach zaczęło leżeć w obszarze zainteresowań badawczych [46]. Warto dodać, że etykietowany zbiór danych został opublikowany na platformie Harvard Dataverse[1].

Uzupełnieniem opisywanej pracy są badania przeprowadzone w artykule [**H13**], gdzie przeanalizowano jak wygląda adopcja transakcji anonimowych w kryptowalutach przez członków tej sieci i czy jest ona uwarunkowana zdarzeniami ze-

[1]Etykietowany zbiór adresów w sieci transakcji Bitcoin: https://doi.org/10.7910/DVN/KEWU0N

wnętrznymi wynikającymi z rozwoju ekosystemu danej kryptowaluty. Możliwość przeprowadzenia takiego badania wynika z faktu, że łańcuchy bloków wielu kryptowalut są transparentne i można prześledzić przepływ środków. Jednocześnie kilka kryptowalut, np. Zcash, oferuje możliwość wykonywania transakcji na bazie tzw. dowodów z wiedzą zerową [5], które umożliwiają wykonywanie i autoryzowanie transakcji, których nie da się prześledzić. W pierwszej części pracy poszukiwano odpowiedzi na pytanie na ile rozwój ekosystemu danej kryptowaluty i zdarzenia zewnętrzne wpływają na popularyzację transakcji tego typu wśród posiadaczy portfeli Zcash, aby – po uzyskaniu negatywnej odpowiedzi na to pytanie – spróbować wykorzystać uczenie nadzorowane do próby predykcji czy dany użytkownik będzie usiłował przenieść środki do puli środków pozwalających na wykonywanie transakcji anonimowych. W pracy zaproponowano i przebadano skuteczność zestawu cech określających węzeł (zarówno sieciowych jak i niesieciowych) wykazując wysoką skuteczność możliwości przewidzenia zaangażowania adresów w transakcje z pulą anonimową. Praca ta uzupełnia wnioski płynące z pracy [**H8**] pokazując, że pewne aspekty deanonimizacji są możliwe w sieciach społecznych zbudowanych na łańcuchu bloków.

W obu powyższych pracach analizujących sieci zbudowane z łańcucha bloków wykorzystano platformę Dru [72], która została stworzona w ramach projektu przemysłowego kierowanego przeze mnie – jej zadaniem jest umożliwienie prostszego dostępu do informacji o transakcjach w sieciach blockchain, ze szczególnym uwzględnieniem metod nauki o sieciach[2].

Kwantyfikacja dynamiki interakcji

Typowym dla sieci statycznych podejściem związanym z wyrażaniem intensywności interakcji pomiędzy węzłami w sieciach społecznych jest osadzanie tej informacji jako atrybuty na krawędziach, najczęściej w postaci wagi krawędzi. Takie modelowanie interakcji pozwala w prosty sposób skwantyfikować częstotliwość interakcji, np. jako wyrażenie udziału kontaktów z danym węzłem w stosunku do wszystkich kontaktów z innymi węzłami. Jednak takie podejście niesie także sporo zagrożeń. Jak wskazano w [25], modelowanie interakcji w sieciach za pomocą modeli statycznych całkowicie usuwa informację o tzw. ścieżce uwzględniającej czas (ang. *time respecting path*), co, w efekcie, utrudnia modelowanie m.in. procesów dyfuzji. Wnioski te doprowadziły do dynamicznego rozwoju podejść, które pozwalają zachowywać kolejność interakcji, ogólnie określanych mianem sieci temporalnych, oraz miar w tych sieciach [48, 56]. Miary te, niestety, ściśle powiązane są jednak z wybranym modelem sieci temporalnej, dla którego zostały opracowane, dlatego też ograniczone są możliwości ich zastosowania w innych modelach.

W związku z powyższym, tam, gdzie to możliwe, w niniejszym cyklu w pracach dotyczących sieci temporalnych starano się bazować na elementarnym i bezstratnym sposobie opisu interakcji: sekwencjach kontaktów. Każdy element tej sekwencji to krotka postaci $< v_i, v_j, t >$, gdzie v_i i v_j oznaczają węzły: odpowiednio inicjatora interakcji i węzła, z którym ta interakcja zachodzi, t oznacza czas interakcji. Podejście to gwarantuje zachowanie właściwej kolejności kontaktów, co w niektórych modelach, takich jak np. sieci temporalne bazujące na oknach czasowych, nie jest możliwe. Jednocześnie też warto zauważyć, że w przypadku modelowania dyfuzji, typowe modele dyfuzji jednak zostały opracowane dla sieci statycznych. Stąd pojawia się pewien dysonans pomiędzy modelami sieci a mode-

[2]Opis platformy Dru: https://dru.readthedocs.org

lami dyfuzji mogący w niedługim czasie być w obszarze zainteresowań szerszego grona badaczy. Choć poza pierwszymi próbami połączenia obu zagadnień [38], nie widać wyraźnej kontynuacji, prawdopodobnie ze względu na stosunkowo mało informacji o dyfuzji w rzeczywistych sieciach społecznych, co także poruszono w pracy [**H3**] opisanej nieco dalej.

W pytaniu **P2** podniesiono problem kwantyfikacji dynamiki interakcji w sieciach. Z jednej strony można go interpretować ściśle jako zagadnienie na poziomie całej sieci i starać się kwantyfikować tzw. *wybuchy*, czyli okresy intensywnej interakcji pomiędzy węzłami (ang. *bursts*). Z drugiej strony, co wydaje się nieść więcej informacji mających bezpośredni wpływ na procesy dyfuzji, w pracy [**H5**] zaproponowano podejście zorientowane na węzły, czyli skalę mikro zamiast skali makro. Wykorzystując powyżej opisaną sekwencję kontaktów jako bezstratny model sieci temporalnej i bazując na koncepcji entropii jako fundamencie [61], opracowano trzy miary pozwalające na kwantyfikację dynamiki kontaktów węzłów w sieciach temporalnych: pierwszego rzędu – entropię węzłową (w trzech wariantach), drugiego rzędu – entropię krawędziową – oraz trzeciego rzędu – entropię następstwa. Uzupełnieniem tych miar jest także miara znormalizowana, tj. wyrażenie wartości miary do maksymalnej możliwej wartości. Rodzina zaproponowanych miar pozwala określać dynamikę i wymienność kontaktów pomiędzy węzłami w sieciach – wychodząc od poszczególnych węzłów informacja ta jest agregowana na poziomie całej sieci. Ewaluacji miar dokonano na kilku rzeczywistych zbiorach reprezentujących interakcje w sieciach społecznych i organizacyjnych, wykorzystując także losowy model referencyjny jako punkt odniesienia [17]. Pozwoliło to wykazać, że interakcje te – zgodnie z socjologicznymi oczekiwaniami – z upływem czasu stają się mniej losowe, ale także wykazano, że dzięki zaproponowanym miarom możliwe jest określanie trendów i anomalii w różnorodności kontaktów z upływem czasu. Obserwacje dynamiki kontaktów przez pryzmat entropii wykorzystano później do zaproponowania miary służącej do selekcji węzłów początkowych w procesie dyfuzji, opisanej w pracy [**H14**], natomiast wnioski płynące z rozważań dotyczących modelowania sieci temporalnych bazujących na sekwencjach zdarzeń doprowadziły do opracowania modelu CogSNet opisanego w pracy [**H11**] – obie te prace zostaną przedstawione w dalszej części niniejszego opisu.

Od wnioskowania do dyfuzji

Powyższe trzy obszary, tj. próbkowanie sieci, odkrywanie atrybutów węzłów oraz kwantyfikacja dynamiki interakcji wpisują się w pierwszy zakres tematyki cyklu publikacji, tj. część dotyczącą **wnioskowania w sieciach społecznych**. W kolejnych punktach zostaną przedstawione opracowane **metody i algorytmy związane z dyfuzją**. Jak napisano wcześniej, wnioskowanie można potraktować jako działanie wspierające mające doprowadzić do zwiększenia efektu dyfuzji, gdyż pozwala ono na lepsze zrozumienie sieci – stąd metody z tego obszaru zostały opisane wcześniej. Ponadto pewne koncepcje zaproponowane w obszarze wnioskowania (np. podejście bazujące na entropii) wykorzystałem jako fundament metod w obszarze dyfuzji, stąd taka kolejność opisu wydawała się bardziej naturalna.

W przedkładanym cyklu artykułów w obszarze dyfuzji rozpatrywane są dwa zasadnicze etapy procesu związanego z maksymalizacją efektu końcowego, które opisane są rozłącznie: selekcja węzłów początkowych oraz sekwencyjna aktywacja węzłów. Aby lepiej wprowadzić czytelnika w dalszą część opisu, pokrótce poniżej opisano jak rozumiane są te etapy.

Problem selekcji węzłów początkowych, niezależnie od tego czy chodzi o marketing, epidemiologię czy też wprowadzanie innowacji, jest często podejmowany w analizie procesów dyfuzji [40]. Wynika to właśnie z faktu, że skuteczne metody wyboru tych węzłów mogą radykalnie wpłynąć na efekt końcowy w sieciach społecznych, np. zwiększenie sprzedaży, zahamowanie epidemii czy popularyzację pewnych pożądanych zachowań. Odpowiednie wykorzystanie wiedzy o strukturze sieci, jej dynamice, atrybutach węzłów lub krawędzi może znacząco skrócić czas lub też budżet do osiągnięcia zamierzonego efektu, zależnie od wariantu rozpatrywanego problemu [21]. W typowym problemie selekcji węzłów początkowych chodzi o taki wybór węzłów początkowych w sieci społecznej (tzw. *seed set*) w ramach określonego budżetu na aktywację węzłów, aby w danym modelu dyfuzji maksymalizować liczbę aktywowanych węzłów w sieci. W kolejnej sekcji omówione zostaną dwie prace cyklu z zakresu selekcji węzłów początkowych, które poruszają się w tak zdefiniowanej przestrzeni problemowej, nawiązując do pytania **P3**.

Natomiast jak opisano w punkcie 4.4.1, węzły wybrane w procesie selekcji aktywowane były od razu. Prowadziło to do sytuacji, w której pewna część budżetu przepadała na aktywację węzłów, które aktywowane byłyby w naturalnym procesie. Inne spojrzenie na ten problem doprowadziło do opracowania rodziny algorytmów sekwencyjnej aktywacji węzłów opisanych w pracach [**H2, H4, H9, H10, H12**]. Taka perspektywa doprowadziła do zdefiniowania kolejnego podproblemu w maksymalizacji dyfuzji, gdzie selekcja węzłów początkowych jest jednym zagadnieniem, a metody ich aktywacji – drugim. Powyższe opublikowane prace przyczyniły się w znacznym stopniu do popularyzacji takiego spojrzenia [27], jednocześnie przynosząc szereg rozwiązań i wniosków dla różnych rodzajów sieci: statycznych, temporalnych i wielowarstwowych. Stąd pytanie **P4** dotyczy algorytmów aktywacji węzłów a nie ich wyboru. A zatem podejścia zaproponowane w części cyklu poświęconej aktywacji mogą wykorzystywać dowolną metodę selekcji węzłów początkowych.

Selekcja węzłów początkowych

W pierwszej pracy cyklu dotyczącej problemu selekcji węzłów początkowych, [**H6**], zaproponowano wykorzystanie podejścia ewolucyjnego do generowania zestawu węzłów początkowych. Podejście to bazuje na obserwacji, że każdy *seed set* można traktować jako osobnika populacji, a zatem możliwe jest poszukiwanie skutecznego zestawu węzłów początkowych poprzez algorytm ewolucyjny. Zalążek tego podejścia zaproponowany został wcześniej w innej pracy własnej [77] – praca uzyskała nagrodę Best Paper Award na warsztacie SNAA 2016 odbywającego się przy konferencji ASONAM. Mimo że w pracy [77] wskazano potencjał tego kierunku, w szczególności poprzez porównanie z metodą zachłanną, to dopiero w artykule [**H6**] w sposób systematyczny zbadano zależności pomiędzy parametrami a jakością wyników oraz przezwyciężono problem wydajnościowy wynikający z ograniczonych możliwości pamięci jednostek GPU, na których wykonywano obliczenia. W toku badań uzyskano kilka istotnych wniosków dotyczących zaproponowanej metody. Po pierwsze, przebadanie jakości wyników w zależności od czasu danego algorytmowi ewolucyjnemu wykazało stosunkowo szybkie zbliżanie się do rozwiązania najlepszego możliwego do uzyskania tą metodą – pozwala to na zastosowanie metody w przypadku systemów, w których konieczne jest szybkie decydowanie o węzłach do aktywacji, np. w szybkozmiennych sieciach temporalnych. Po drugie, istotnym wnioskiem było wykazanie, że określenie dobrego zestawu parametrów algorytmu ewolucyjnego na jednej sieci często jest

wystarczające by zastosować je na sieci podobnego typu bez konieczności ponownego badania przestrzeni parametrów.

W drugiej pracy z tego obszaru przyjrzano się wnioskom pochodzącym z artykułu [**H5**], w którym opracowano nowe miary bazujące na entropii do kwantyfikacji stabilności interakcji. W efekcie, w pracy [**H14**] zaproponowano nową heurystykę selekcji węzłów początkowych, która bazuje na wymienności sąsiedztwa w sieciach temporalnych. Propozycja ta wynika z faktu, że procesy dyfuzji w sieciach często mają tendencję do zamykania się w klastrach [52]. A zatem zaproponowano hipotezę, że podejście bazujące na wymienności sąsiedztwa ma szansę uzyskać wyższą skuteczność (rozumianą jako liczba aktywowanych węzłów) niż typowe podejścia bazujące na miarach sieciowych. W celu dopasowania warunków eksperymentu do uznanych modeli dyfuzji, w eksperymentach użyto modelu sieci temporalnej bazującego na oknach, dzięki czemu możliwe było wykorzystanie modelu niezależnej kaskady (ang. *independent cascade model*) [19], wykorzystywanego najczęściej także w kolejnych pracach niniejszego cyklu. Brak determinizmu modelu niestety wymaga wielokrotnego powtarzania przebiegów, jednak w eksperymentach prowadzących do innej pracy cyklu, [**H9**], udało się ustalić taką liczbę powtórzeń eksperymentu, aby wyniki były reprezentatywne. Wnioski płynące z pracy [**H14**] pokazują, że w ponad 40% zbadanych konfiguracji parametrów zaproponowana heurystyka była lepsza niż pozostałe przebadane metody. Istotnym elementem pracy było także zbadanie podobieństwa zbiorów węzłów początkowych uzyskiwanych przez poszczególne metody selekcji. W przyszłości można tę wiedzę wykorzystać do budowy bardziej złożonych heurystyk wykorzystujących różne aspekty aktywności i pozycji węzłów w sieciach.

Obie powyżej opisane prace wzbogacają zestaw metod selekcji węzłów początkowych w sieciach społecznych (odpowiedź na pytanie **P3**) i – co istotne – żadna nie niosą dużego narzutu obliczeniowego, co czyni je szczególnie przydatnymi w zastosowaniach, gdzie wymagane jest szybkie określenie zestawu węzłów początkowych.

Sekwencyjna aktywacja węzłów

Istotnym wkładem cyklu w obszar dyfuzji w sieciach społecznych jest zestaw prac poświęconych tzw. metodom sekwencyjnej aktywacji węzłów (ang. *sequential seeding*). Jak już napisano wcześniej, w typowych wyzwaniach maksymalizacji dyfuzji (np. wpływu społecznego) węzły uzyskane z procesu selekcji węzłów początkowych aktywowane są od razu. Jednak w opisanych poniżej pracach zaproponowano zupełnie inne spojrzenie na to zagadnienie – aktywacja węzłów nie musi być dokonywana natychmiast, przez co możliwe jest poprawienie całkowitego efektu dyfuzji. Opisany poniżej zestaw pięciu prac podsumowuje badania nad zaproponowaną metodą sekwencyjnej aktywacji węzłów, która początkowo została opracowana dla sieci statycznych a następnie rozbudowana o warianty dla sieci temporalnych i wielowarstwowych.

Pierwsza praca dotycząca sekwencyjnej aktywacji węzłów, [**H2**], stanowi wprowadzenie do koncepcji metody i jej wariantów oraz przebadaniu jej skuteczności w sieciach statycznych dla modelu *indepentent cascade* IC [19]. Podstawową filozofią zaproponowaną w metodzie jest wskazane w tytule pracy równoważenie prędkości aktywacji i pokrycia sieci *(„balancing speed and coverage")*, gdyż rzeczywiście w zależności od potrzeby konieczne jest odpowiednie dobranie środków w procesie dyfuzji – niekoniecznie najistotniejszy jest szybki przyrost aktywacji, być może istotniejszy jest ostateczny zasięg. Podążając za tą ideą, w pracy zaproponowano pięć wariantów opracowanej metody rozpatrujących różne scenariusze

w zależności od celu. Następnie przebadano wszystkie warianty metody na łącznie 1875 konfiguracjach piętnastu różnych sieci, parametrów modelu IC oraz metod selekcji węzłów. Wnioski płynące z eksperymentów pokazały, że w 85% przypadków zaproponowany algorytm jest lepszy niż jednoczesna aktywacja wszystkich węzłów z zestawu węzłów początkowych.

Kolejna publikacja z tego obszaru, praca [**H4**], rozszerza badania nad sekwencyjną aktywacją węzłów w kilku kierunkach. Po pierwsze, badania uzupełniono o sieci skierowane, gdyż w pracy [**H2**] rozpatrywano tylko wariant nieskierowany sieci. Po drugie, analitycznie udowodniono, że zaproponowana metoda w najgorszym przypadku może dać takie same rezultaty jak jednoczesna aktywacja węzłów, czyli nigdy nie będzie gorsza od tego podejścia. Kolejnym wkładem tej pracy jest usystematyzowanie porównania rozpatrywanych metod dla nieterministycznego modelu IC poprzez zaproponowanie podejścia określanego w pracy jako *coordinated execution* – dzięki temu możliwa jest lepsza analiza porównawcza metod na różnych sieciach dla nieterministycznych modeli dyfuzji. Istotnym wkładem pracy jest także wyznaczenie i wykazanie gdzie leży górna granica możliwości aktywacji w najlepszym koszyku węzłów początkowych i jak ma się do tego opracowana metoda. Dodatkowo wykazano, że połączenie prostej metody selekcji węzłów początkowych bazującej na stopniu węzła i połączeniu jej z sekwencyjną metodą aktywacji jest skuteczniejsze niż bardzo skuteczna, lecz obliczeniowo złożona, metoda zachłanna z jednoczesną aktywacją. W tym ostatnim kontekście badania przeprowadzono na siedmiu rzeczywistych sieciach dla różnych parametrów procesu dyfuzji i wykazano, że w ok. 92% przypadków sekwencyjna aktywacja węzłów ze stopniem węzła jako metodą selekcji jest lepsza niż algorytm zachłanny z jednorazową aktywacją. Co więcej, sekwencyjna aktywacja pozwala na zmniejszenie dystansu pomiędzy algorytmem zachłannym – uznawanym ciągle za najlepszą heurystykę – a maksymalnym możliwym zasięgiem nawet o 83%.

Ostatnią publikacją eksplorującą zagadnienie sekwencyjnej aktywacji węzłów w sieciach statycznych jest praca [**H12**]. Celem pracy było zbadanie innego aspektu związanego z sekwencyjną aktywacją węzłów, a mianowicie przeliczania rankingów do wyboru węzłów początkowych. W poprzednich dwóch pracach stosowano heurystykę selekcji węzłów jednokrotnie przed rozpoczęciem procesu aktywacji, niezależnie od metody aktywacji. W niniejszej pracy zbadano jak zmieniają się wyniki w zależności od strategii przeliczania heurystyki vote-rank [81] w trakcie procesu aktywacji. W pracy zbadano dwa obszary: (i) ogólny wpływ przeliczania miary przed aktywacją kolejnych węzłów oraz (ii) zależność pomiędzy częstotliwością przeliczania miary a efektem końcowym. Opracowane podejście zostało określone mianem *dynamic vote-rank* i skutkuje ono dalszą poprawą wyników o maksymalnie 10% względem podejścia bez zastosowania tej metody. Dodatkowo wskazano najlepsze strategie związane z poszukiwaniem kompromisu pomiędzy przeliczaniem rankingu a efektami końcowymi.

Kolejne dwie prace dotyczące sekwencyjnej aktywacji węzłów rozszerzają możliwości metody na dwa inne modele sieci: sieci temporalne [**H9**] i wielowarstwowe [**H10**]. W pierwszej pracy zaproponowano metodę sekwencyjnej aktywacji węzłów dla sieci temporalnych w wariancie okien czasowych dla – podobnie jak w poprzednich pracach – modelu dyfuzji IC, który musiał zostać dostosowany do sieci tego typu. Ponieważ praca miała odpowiadać na pytanie o skuteczność samej metody aktywacji, nie selekcji, skupiono się na porównaniu metody jednoczesnej aktywacji wszystkich węzłów i sekwencyjnej aktywacji dla heurystyki bazującej na stopniu węzła. W pracy wykazano, że dla niemal 74% przebadanych

konfiguracji metoda *sequential seeding* jest tak samo dobra lub lepsza niż metoda jednoczesnej aktywacji węzłów. W pracy zastosowano także referencyjny model losowy [17], który pokazał, że po eliminacji przyczynowości metoda sekwencyjna osiąga jeszcze lepsze wyniki niż jednoetapowa, gdyż w ponad 85% konfiguracji jest lepsza a tylko w niecałych 15% osiąga takie same lub gorsze wyniki.

Badania nad sekwencyjną aktywacją węzłów w ramach cyklu finalizuje praca [**H10**], która z kolei dotyczy sieci wielowarstwowych. W artykule zaproponowano wariant metody dla sieci wielowarstwowych i przebadano jej możliwości dla dziesięciu sieci tego typu, zarówno rzeczywistych jak i syntetycznych, znowu dla modelu IC. Oprócz klasycznego spojrzenia na wyniki zasięgu, osobnym zagadnieniem poruszonym w pracy jest spojrzenie przez pryzmat potencjalnych oszczędności budżetu dla osiągnięcia takiego samego efektu w kwestii liczby aktywacji. Wnioski z pracy są następujące: (i) w przebadanych sieciach wielowarstwowych metoda sekwencyjna zawsze jest co najmniej tak samo dobra jak metoda jednoetapowa, w 74% przypadków dała lepsze wyniki, (ii) często dla uzyskania takiego samego zasięgu w sieciach wielowarstwowych wystarczy mniejszy budżet połączony z sekwencyjną aktywacją, co wykazano dla 31.75% przypadków, (iii) obie powyższe własności jednak związane są z wydłużeniem liczby iteracji procesu, co musi być szczególnie wzięte pod uwagę w kwestii kampanii nastawionych na szybki efekt.

Powyżej opisane pięć prac cyklu poświęconych sekwencyjnej aktywacji węzłów nie wyczerpuje oczywiście zagadnienia. Obecnie w recenzji znajduje się praca będąca rozwinięciem koncepcji przedstawionych w artykule [**H12**] a także trwają przygotowania do rozpatrywania wariantu agentowego modelu Bassa w sieciach wielowarstwowych [57]. Na pewno jednak można stwierdzić, że w opisanych pracach poświęconych sekwencyjnej metodzie aktywacji węzłów udało się zbadać problem z wielu perspektyw wykazując jednocześnie możliwości i ograniczenia tej rodziny metod, co wniosło nową wiedzę w zakresie pytania **P4**.

Model sieci temporalnej i dyfuzja w sieciach wielowarstwowych

Ostatnim elementem osiągnięcia naukowego, który uzupełnia wyżej opisane metody związane z wnioskowaniem i dyfuzją w sieciach społecznych, są dwie prace, które zostały wskazane w górnej części rysunku 1. Ich lokalizacja w przedmiotowym schemacie wynika z faktu, że wnioski płynące z tych prac mogą zostać wykorzystane w metodach dotyczących wnioskowania i dyfuzji w sieciach temporalnych i wielowarstwowych na dowolnym etapie procesu.

W pierwszej pracy, [**H11**], odniesiono się do pytania **P5**. Pytanie to zostało sformułowane w podsumowaniu doktoratu wnioskodawcy, gdyż wiele z rozpatrywanych metod w procesach dyfuzji w sieciach bazuje na modelach zawodnych w kontekście adekwatnej reprezentacji tychże procesów. Są to często sieci statyczne, ewentualnie sieci temporalne inkrementacyjne, z oknem przesuwnym lub rozłącznym. Podobnie, wiele z modeli dyfuzji dostosowanych jest do sieci statycznych, o czym częściowo już napisano w punkcie 4.4.1. W związku z tym, że część z prac opisanych w cyklu dotyczyła procesów dyfuzji w sieciach temporalnych, zadano sobie pytanie czy możliwe byłoby wykorzystanie wiedzy o procesach kognitywnych do zaprononowania nowego modelu sieci temporalnej, który byłby pozbawiony okien i zachowywałby wspomnianą już wcześniej ścieżkę uwzględniającą czas (*time respecting path*). Dodatkowo istotnym celem było zamodelowanie funkcji zapominania ludzkiej pamięci [79], co pozwoliłoby na prostszą adaptację niektórych modeli procesów dyfuzji, takich jak np. wpływ społeczny. Efektem tych rozważań i współpracy m.in. z reprezentantami kognitywistyki i socjologii jest praca [**H11**], w której zaproponowano i przebadano model sieci temporalnej

CogSNet – ***Cog**nition-driven **S**ocial **Net**work*. Model ten wprowadza mechanizmy nawiązujące do procesów zapominania (funkcja zapominania i odcięcie), ale także proponuje element wzmocnienia przy zwiększeniu częstotliwości interakcji (ang. *reinforcement*). Model CogSNet jest modelem zapewniającym ciągłość, tzn. można policzyć wartości wag krawędzi w każdym momencie wychodząc od zbioru interakcji, a także zachowuje ścieżkę uwzględniającą czas. Aby zbadać adekwatność modelu, wykorzystano unikalny zbiór NetSense z Uniwersytetu Notre Dame [63]. Zbiór ten zawiera informacje o interakcjach pomiędzy studentami (rejestr połączeń telefonicznych i wiadomości tekstowych), ale – co istotne – studentów okresowo ankietowano, między innymi pytając o ich percepcję częstotliwości kontaktów z innymi. Tym samym można było porównać skuteczność modelu CogSNet bazującego na zdarzeniach telekomunikacyjnych z postrzeganiem przez studentów ich kontaktów z otoczeniem. Wyniki opublikowane w pracy pokazują, że zaproponowany model spośród ewaluowanych osiągnął najwyższą skuteczność mierzoną miarą Jaccarda gdy porównywano oba zbiory: sąsiadów w modelu CogSNet i listę uzyskaną od studentów. Dodatkowo skuteczność tak uzyskana była dla wartości parametrów zbieżnych z wnioskami kognitywistów dotyczącymi ludzkiej pamięci. W dalszych pracach związanych z kontynuacją badań w tym obszarze rozważa się migrację modeli dyfuzji do sieci temporalnych bazujących na zaproponowanym modelu i personalizację parametrów.

Ostatnia praca cyklu, [**H3**], jest efektem obserwacji, że bardzo ciężko o pozyskanie rzeczywistych danych związanych z procesami dyfuzji w sieciach wielowarstwowych. W związku z tym, w ramach wstępnego etapu prac badawczych nad procesami dyfuzji w sieciach wielowarstwowych podjęto się pozyskania, przeanalizowania i udostępnienia[3] bogatego zbioru danych zawierającego informacje o procesach dyfuzji w portalu timik.pl – współautorem pracy [**H3**] jest twórca tego portalu. Pracę tę można traktować jako fundament nad dalszymi badaniami dotyczącymi procesów dyfuzji w sieciach wielowarstwowych.

4.4.4 Podsumowanie wkładu osiągnięcia naukowego

W ramach osiągnięcia naukowego w dyscyplinie informatyka techniczna i telekomunikacja zaproponowano zestaw metod i algorytmów wpisujących się w zagadnienie wnioskowania i dyfuzji w sieciach społecznych. W szczególności celem prac badawczych było znalezienie odpowiedzi na pytania **P1**–**P5** przedstawione w punkcie 4.4.2. Podsumowując, cykl powiązanych tematycznie publikacji doprowadził do:

- zaproponowania podejść pozyskiwania dodatkowej wiedzy o węzłach w sieciach społecznych, niezawartej wprost w atrybutach węzłów i krawędzi sieci [**H7, H8, H13**],
- stworzenia nowych metod skuteczniejszego wyboru węzłów do aktywacji w procesach dyfuzji, co szczególnie przy połączeniu ze sposobem aktywacji może skutkować zwiększeniem całkowitego efektu procesu [**H6, H14**],
- opracowania rodziny algorytmów określanej jako sekwencyjna aktywacja węzłów – zapewniającej efektywniejszy sposób aktywacji węzłów w procesach dyfuzji oraz przebadania jej w różnych wariantach sieci społecznych (statyczne, temporalne, wielowarstwowe) – [**H2, H4, H9, H10, H12**],
- zaproponowania nowego modelu sieci temporalnej wykorzystującego wiedzę o procesach kognitywnych, umożliwiającego zastosowanie wybranych modeli dyfuzji opracowanych dla sieci statycznych w sieciach temporalnych [**H11**],

[3]Zbiór danych – dyfuzja w sieciach wielowarstwowych: https://doi.org/10.7910/DVN/V6AJRV

Tabela 3: Obszary badane w głównym osiągnięciu naukowym w rozbiciu na poszczególne publikacje cyklu.

	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14
Obszary, metody i algorytmy														
Pozyskiwanie wiedzy w sieciach	x				x		x	x					x	
Dyfuzja w sieciach społecznych		x	x	x		x			x	x	x	x		x
Selekcja węzłów początkowych						x								x
Aktywacja węzłów w procesie dyfuzji		x		x					x	x		x		
Modelowanie interakcji i dyfuzji			x		x						x			
Uczenie nadzorowane i.i.d.								x					x	
Klasyfikacja kolektywna							x							
Próbkowanie sieci	x													
Miary bazujące na entropii					x									x
Podejście ewolucyjne w selekcji węzłów						x								
Modele sieci														
Sieci statyczne	x	x		x			x	x				x	x	
Sieci temporalne					x	x			x		x			x
Sieci wielowarstwowe			x							x				
Przykłady sieci społecznych														
Klasyczne sieci społeczne	x	x	x	x	x	x			x	x	x	x		x
Sieci bazujące na łańcuchu bloków								x					x	
Sieci w organizacjach					x	x	x		x		x			x

– stworzenia nowych miar opisu dynamiki kontaktów [**H5**], pogłębienia rozumienia procesów dyfuzji w sieciach wielowarstwowych [**H3**] oraz zaproponowania algorytmu próbkowania odwzorowującego rozkłady miar sieciowych w sieciach [**H1**].

Skuteczność opracowanych metod weryfikowano dla różnych przykładów sieci społecznych: począwszy od klasycznych sieci społecznych, przez sieci organizacyjne i na sieciach bazujących na łańcuchu bloków skończywszy. Dodatkowym wymiarem był model sieci, gdyż badania przeprowadzono na sieciach statycznych, temporalnych i wielowarstwowych. Całościowe ujęcie zagadnień badanych w ramach cyklu przedstawiono w tabeli 3, gdzie dla każdej publikacji wskazano których obszarów dotyczyła oraz dla jakich modeli oraz przykładów sieci przeprowadzono badania. Warto podkreślić, że zastosowania zaproponowanych podejść wykraczają poza wymienione przykłady sieci społecznych i mogą być wykorzystane także w innych gałęziach informatyki. Przykładowo, metody opracowane w pracach dotyczących aktywacji węzłów (metody wyboru i sekwencyjnej aktywacji) mogą być z powodzeniem stosowane w sieciach technologicznych, np. komputerowych, choćby w sieciach peer to peer do dystrybucji poprawek, aktualizacji konfiguracji węzłów lub takiego rozmieszczania treści aby minimalizować czas jej propagacji w całej sieci.

Istotnym aspektem, który był podniesiony pracach [**H7, H8, H13**], jest także zagadnienie prywatności. Wyniki badań przedstawione w ww. pracach pokazują, że możliwe jest wnioskowanie o charakterze węzłów w sieciach na podstawie innych cech lub aktywności. Tym samym, mając to na uwadze, jeśli konieczne jest zachowanie prywatności, należy przedsięwziąć dodatkowe kroki zwiększające

anonimowość węzłów. W szczególności wnioski te dotyczą sieci otwartych, takich jak blockchain, gdzie możliwy jest dostęp do historii transakcji i, jak wykazały badania, można pozyskać dodatkową wiedzę niezawartą w łańcuchu bloków.

Prace przedłożone w ramach cyklu nie wyczerpują zagadnienia wnioskowania i dyfuzji w sieciach społecznych, także w zakresie badań wnioskodawcy. Ich uzupełnieniem są także niektóre prace opublikowane po doktoracie wymienione w punkcie 4.5.1, które ze względu na właśnie uzupełniający charakter zostały wyłączone z cyklu. Ponadto obecnie w recenzjach w czasopismach z IF znajdują się dwie kolejne prace: (i) rozwijająca koncepcje opracowane w pracy [**H5**], w której analizowana jest złożoność podejścia oraz proponowane są algorytmy optymalizacji czasu obliczeń, (ii) będąca kontynuacją pracy [**H14**]. W toku są kolejne badania wywodzące się koncepcyjnie z przedłożonego cyklu, w tym np. ewaluowane są metody wykorzystujące podejście agentowe do modelowania niektórych procesów dyfuzji w sieciach wielowarstwowych, np. modelu Bassa [57]. Mniejsze ujęcie elementów formalnych algorytmów w kilku pracach cyklu wynikało z interdyscyplinarnego charakteru niektórych czasopism i chęci zaprezentowania wyników naukowcom z innych dyscyplin, jednak w takich przypadkach najczęściej publikowano kod algorytmów, co także umożliwia reprodukowanie badań.

4.5 Pozostałe osiągnięcia naukowo-badawcze

4.5.1 Osiągnięcia dotyczące prowadzonych badań

W niniejszym punkcie przedstawiono prace opublikowane po uzyskaniu stopnia doktora jednak niebędące częścią głównego osiągnięcia naukowego – cyklu publikacji. Jest to łącznie dwanaście prac: cztery artykuły w czasopismach z IF, siedem referatów konferencyjnych na konferencjach międzynarodowych oraz jeden rozdział w monografii (encyklopedii). Podzielono je dwie sekcje: prace nawiązujące do cyklu oraz nienawiązujące do cyklu. Po obu wykazach prac następuje ich krótkie omówienie. Natomiast pełny wykaz prac będący dorobkiem wnioskodawcy obejmującym także okres przed uzyskaniem stopnia doktora oraz inne wystąpienia konferencyjne przedstawiono w załączniku 4 do wniosku. Numeracja poniższych prac odpowiada numeracji z bibliografii zamieszczonej na końcu niniejszego autoreferatu a wybrane publikacje stanowią załącznik 8 wniosku.

Prace nawiązujące do cyklu

[29] Jarosław Jankowski, Piotr Bródka, **Radosław Michalski**, and Przemysław Kazienko. Seeds buffering for information spreading processes. In *SocInfo 2017: International Conference on Social Informatics*, vol. 10539 of *Lecture Notes in Computer Science*, pp. 628–641. Springer, Cham, 2017.
Punkty MEiN: 20, liczba cytowań: 10/9/9. Szacowany wkład własny: 25% (analiza stanu wiedzy, zaprojektowanie i wykonanie części eksperymentów, udział w opracowaniu tekstu manuskryptu).

[31] Jarosław Jankowski and **Radosław Michalski**. Increasing coverage of information spreading in social networks with supporting seeding. In *DMBD 2017: International Conference on Data Mining and Big Data*, vol. 10387 of *Lecture Notes in Computer Science*, pp. 209–218. Springer, Cham, 2017.
Punkty MEiN: 20, liczba cytowań: 4/3/3. Szacowany wkład własny: 10% (udział w opracowaniu tekstu manuskryptu).

[33] Jarosław Jankowski, **Radosław Michalski**, Piotr Bródka, and Artur Karczmarczyk. Increasing coverage of information diffusion processes by reducing the number of initial seeds. In *ASONAM 2017: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analy-*

sis and Mining, pp. 713–720. IEEE, 2017.
Punkty MEiN: 20, liczba cytowań: 1/1/–. Szacowany wkład własny: 25% (udział w przeprowadzeniu eksperymentów, analizie wyników i opracowaniu manuskryptu).

[34] Jarosław Jankowski, **Radosław Michalski**, Piotr Bródka, Przemysław Kazienko, and Sonja Utz. Adaptive survey design using structural characteristics of the social network. In *SocInfo 2015: International Conference on Social Informatics*, vol. 9471 of *Lecture Notes in Computer Science*, pp. 153–163. Springer, Cham, 2015.
Punkty MEiN: 15, liczba cytowań: 0/0/0. Szacowany wkład własny: 15% (udział w przeprowadzeniu eksperymentów, analizie wyników i opracowaniu manuskryptu).

[36] Jarosław Jankowski, Marcin Waniek, Aamena Alshamsi, Piotr Bródka, and **Radosław Michalski**. Strategic distribution of seeds to support diffusion in complex networks. *PloS one*, 13(10):e0205130, Public Library of Science, 2018.
IF: 2,776, punkty MEiN: 40, liczba cytowań: 6/6/6. Szacowany wkład własny: 5% (opracowanie założeń eksperymentu, interpretacja wyników).

[37] Tomasz Kajdanowicz, **Radosław Michalski**, Katarzyna Musiał, and Przemysław Kazienko. Learning in unlabeled networks – an active learning and inference approach. *AI Communications*, 29(1):123–148, IOS Press, 2016.
IF: 0,654, punkty MEiN: 15, liczba cytowań: 3/3/3. Szacowany wkład własny: 20% (opracowanie założeń eksperymentu, implementacja algorytmu, przeprowadzenie badań, interpretacja wyników).

[72] **Radosław Michalski** and Marcin Pieczka. Dru: Studying blockchain as a complex network. In *ASONAM 2020: Proceedings of the 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 929–932. IEEE, 2020.
Punkty MEiN: 20, liczba cytowań: 0/0/–. Szacowany wkład własny: 80% (opracowanie koncepcji, udział w implementacji, przygotowanie manuskryptu).

[77] Michał Weskida and **Radosław Michalski**. Evolutionary algorithm for seed selection in social influence process. In *ASONAM 2016: Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 1189–1196. IEEE, 2016.
Punkty MEiN: 15, liczba cytowań: 16/11/8. Szacowany wkład własny: 50% (analiza stanu wiedzy, opracowanie planu eksperymentów, analiza wyników, przygotowanie tekstu manuskryptu) – **SNAA 2016 Best Paper Award**.

Wymienione powyżej prace można traktować jako uzupełnienie cyklu w kilku obszarach. W ramach kontynuacji prac nad adaptacyjną metodą próbkowania sieci zaproponowaną w publikacji [**H1**], w pracy [**34**] zbadano jak różni się w aspekcie jakości próbkowania zaproponowana metoda próbkowania od innych podejść. W pracy zaproponowano trzy algorytmy próbkowania bazujące na miarach sieciowych z użyciem których porównano jakość próbkowania i reprezentatywność próbki. Wykazano, że w sieciach, w których obserwowane jest prawo potęgowe, dalsze próbkowanie nie jest już tak skuteczne po przekroczeniu pewnego granicznego progu dla badanych sieci.

W artykule [**37**] zbadano możliwości algorytmów uczenia relacyjnego do odkrywania wiedzy w sieciach złożonych różnego typu, co z kolei było inspiracją

do dalszych prac w zakresie algorytmu klasyfikacji kolektywnej zaproponowanego w pracy [**H7**], który został przebadany w statycznych sieciach organizacyjnych wykazując się skutecznością.

Kolejne prace, czyli [**29, 31, 33, 36**], powiązane są z sekwencyjną metodą aktywacji węzłów będącą istotną częścią cyklu publikacji. W pracach przebadano różne aspekty powiązane z tą metodą, takie jak np. wpływ buforowania (czasowego wstrzymania aktywacji) na proces dyfuzji [**29**], możliwość zwiększenia efektywności procesu poprzez odpowiedni dobór kolejnych węzłów w trakcie procesu dyfuzji [**31**], relację pomiędzy redukcją liczby węzłów aktywowanych początkowo i efektem końcowym [**33**], a także przebadano wpływ różnych strategii aktywacji w ramach medody sekwencyjnej na wynik procesu dyfuzji [**36**].

Praca [**77**] jest wyznaczeniem kierunku wykorzystania algorytmów ewolucyjnych do selekcji węzłów początkowych, który dalej eksplorowałem w pracy [**H6**]. W artykule wykazano, że możliwe jest stworzenie skutecznego algorytmu do selekcji węzłów początkowych poprzez spojrzenie na zbiór węzłów początkowych jak na osobnika w algorytmie ewolucyjnym – dzięki temu można realizować proces poszukiwania najlepszego zbioru węzłów tą metodą. W pracy wykazano, że podejście to pozwala na uzyskanie wyników, które mogą konkurować z algorytmem zachłannym, jednak natrafiono także na pewne ograniczenia (w szczególności związane z maksymalnym rozmiarem sieci), które – oprócz poruszenia wielu innych aspektów – zostały pokonane w pracy [**H6**]. Praca zdobyła nagrodę **Best Paper Award** na warsztacie SNAA 2016 przy konferencji ASONAM 2016.

Praca [**72**] proponuje możliwość spojrzenia na łańcuch bloków jak na sieć i wprowadza narzędzie – platformę Dru – opracowane w trakcie projektu, którego byłem kierownikiem. Platforma ta umożliwia proste budowanie sieci z historii transakcji a następnie przetwarzanie jej z wykorzystaniem metod nauki o sieciach. Platforma Dru była wykorzystana w pracach [**H8, H13**].

Prace nienawiązujące do cyklu

[15] Jean-Francois Ethier, Vasa Curcin, Mark M McGilchrist, Sarah N Lim Choi Keung, Lei Zhao, Anna Andreasson, Piotr Bródka, **Radosław Michalski**, Theodoros N. Arvanitis, Nikolaos Mastellos, et al. eSource for clinical trials: Implementation and evaluation of a standards-based approach in a real world trial. *International journal of medical informatics*, 106:17–24, Elsevier, 2017.
IF: 2.957, punkty MEiN: 40, liczba cytowań: 12/8/4. Szacowany wkład własny: 10% (zaproponowanie i opracowanie koncepcji informatycznych w systemie eSource, udział w opracowaniu manuskryptu).

[59] Przemysław Różewski, Jarosław Jankowski, Piotr Bródka, and **Radosław Michalski**. Knowledge workers' collaborative learning behavior modeling in an organizational social network. *Computers in Human Behavior*, 51:1248–1260, Elsevier, 2015.
IF: 2,880, punkty MEiN: 40, liczba cytowań: 40/18/13. Szacowany wkład własny: 7.5% (opracowanie założeń eksperymentu, interpretacja wyników).

[64] **Radosław Michalski**. Linear threshold model in temporal networks—Seed selection for social influence. In *ASONAM 2015: Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 922–923. IEEE, 2015.
Punkty MEiN: 15, liczba cytowań: 2/3/2. Wkład własny: 100%.

[70] **Radosław Michalski** and Przemysław Kazienko. Social network analysis in organizational structures evaluation. In Reda Alhajj and Jon Rokne, editors, *Encyclopedia of Social Network Analysis and Mining, Living Reference Work*, pp. 1–13. Springer, New York, 2017.
Punkty MEiN: 20, liczba cytowań: 0/–/–. Szacowany wkład własny: 90% (opracowanie koncepcji, badania, interpretacja wyników, udział w opracowaniu zaktualizowanego manuskryptu).

W pracy [**15**] podsumowano część wyników projektu TRANSFoRm, którego byłem wykonawcą. Praca dotyczy w szczególności aspektu interoperacyjności systemu, który integruje dane elektroniczne pacjentów (EHR) wraz z systemem wnioskowania. Interoperacyjność tę potwierdziło skuteczne wdrożenie systemu w wielu europejskich krajach.

Praca [**59**] dotyczy sieci organizacyjnych, także badanych w ramach głównego cyklu publikacji. Jednak w artykule prace badawcze dotyczyły innego aspektu organizacji, a mianowicie strategii alokacji zasobów wiedzy w celu maksymalizacji kompetencji. Dzięki zaproponowanym strategiom możliwe jest z jednej strony wnioskowanie o charakterystyce członków organizacji jak i takie ustalanie strategii aby organizacja osiągnęła pożądany poziom kompetencyjny. Fundamentem do pracy [**59**] jest zaktualizowany wpis encyklopedyczny [**70**], który proponuje analizę organizacji przez pryzmat analizy sieciowej i wprowadza techniki pozwalające na realizację tego zadania.

Publikacja [**64**] dotyczy rozważań nad modelami wpływu w temporalnych sieciach społecznych i w pewnym aspekcie sięga do koncepcji z doktoratu, stąd nie jest tutaj szerzej omawiana.

4.5.2 Opracowane zbiory danych oraz opublikowany kod źródłowy

Oprócz publikowania prac głównie w otwartym dostępie *(open access)*, staram się opracowywać i udostępniać wykorzystane zbiory danych oraz kod pozwalający na reprodukowanie wyników badań (w tym opracowane algorytmy). Na ogół wykorzystuję do tego platformy Harvard Dataverse (dane) oraz Code Ocean (kod).

Opracowane zbiory danych

- procesy rozprzestrzeniania w sieciach wielowarstwowych na platformie online – opracowany do pracy [**H3**]: https://doi.org/10.7910/DVN/V6AJRV
- komunikacja e-mailowa oraz struktura organizacyjna przedsiębiorstwa – opracowane do prac [**H7**] i [71]: https://doi.org/10.7910/DVN/6Z3CGX
- kategorie węzłów w sieci Bitcoin – niemal 9.000 etykietowanych adresów węzłów – opracowany do pracy [**H8**]: https://doi.org/10.7910/DVN/KEWU0N

Kod źródłowy algorytmów do celów reprodukowania wyników

- publikacja [**H9**]: https://doi.org/10.24433/CO.7254599.v1
- publikacja [**H10**]: https://doi.org/10.24433/CO.4337573.v1
- publikacja [**H11**]: https://doi.org/10.24433/CO.6088785.v1

Ponadto, platforma Dru do przetwarzania łańcucha bloków, którą wykorzystano w publikacjach [**H8, H13**], została także wydana na otwartej licencji i opublikowana na platformie GitHub: https://github.com/bergplace/Dru.

4.5.3 Udział w projektach badawczych

Biorę lub brałem udział w **dziesięciu wymienionych poniżej krajowych i międzynarodowych projektach badawczych finansowanych w drodze konkursów** – w jednym jako kierownik (NCN SONATA, pierwsza pozycja w wykazie) oraz w dziewięciu jako wykonawca – przedstawiam je na kolejnych stronach.

1. **Dyfuzja informacji w temporalnych sieciach społecznych**
   - finansowanie: Narodowe Centrum Nauki – program SONATA
   - numer grantu: 2015/17/D/ST6/04046
   - okres realizacji: 2016–2020 (zakończony)
   - rola: **kierownik projektu**
   - budżet projektu: 230 400 zł
2. **TRANSFoRm – Translational Research and Patient Safety in Europe**[4]
   - finansowanie: Unia Europejska, Siódmy Program Ramowy – FP7-ICT
   - numer grantu: FP7-ICT-2011-7
   - okres realizacji: 2011–2015 (zakończony)
   - rola: **wykonawca**
   - budżet projektu: € 9 726 688, budżet Politechniki Wrocławskiej 2 823 652 zł
   - pozostali główni beneficjenci projektu: King's College London (lider projektu, Wielka Brytania), Imperial College London (Wielka Brytania), University of Birmingham (Wielka Brytania), Trinity College Dublin (Irlandia), Karolinska Institute (Szwecja), NIVEL (Holandia)
3. **ENGINE – European research centre of Network intelliGence for INnovation Enhancement**[5]
   - finansowanie: Unia Europejska, Siódmy Program Ramowy – FP7-REGPOT
   - numer grantu: 316097
   - okres realizacji: 2013–2016 (zakończony)
   - rola: **wykonawca**
   - budżet projektu: € 4 731 186 (w całości dla PWr)
4. **RENOIR – Reverse EngiNeering of sOcial Information pRocessing**[6]
   - finansowanie: Unia Europejska, program Horyzont 2020 – H2020-MSCA-RISE
   - numer grantu: 691152
   - okres realizacji: 2016–2019 (zakończony)
   - rola: **wykonawca**
   - budżet projektu: € 1 314 000, budżet Politechniki Wrocławskiej € 369 000
   - pozostali główni beneficjenci i partnerzy projektu: Politechnika Warszawska (lider projektu), University of California, Davis (Stany Zjednoczone), Rensselaer Polytechnic Institute (Stany Zjednoczone), Carnegie Mellon University (Stany Zjednoczone), University of Technology Sydney (Australia), Notre Dame University (Stany Zjednoczone), Nanyang Technological University (Singapur)
5. **CARE-ARSEC: Computer Aided Reasoning in Analysing Regional Socio-Economic Complexity**[7]
   - finansowanie: Międzynarodowy Fundusz Wyszehradzki
   - numer grantu: 21810079
   - okres realizacji: 2018–2020 (zakończony)

[4] O projekcie TRANSFoRm: https://www.i-hd.eu/index.cfm/resources/ec-projects-results/TRANSFoRm
[5] O projekcie ENGINE: http://engine.pwr.wroc.pl/en
[6] O projekcie RENOIR: https://renoirproject.eu
[7] O projekcie CARE-ARSEC: https://www.globari.org/v4-care-arsec

- rola: **wykonawca**
- budżet projektu: € 43 400
- pozostali partnerzy projektu: Global Arena Research Institute (lider projektu, Czechy), Masaryk University (Czechy), Kopint-Tárki Institute (Węgry), Research Centre of the Slovak Foreign Policy Association (Słowacja)

6. **Metody uczenia maszynowego w sieciach złożonych**
   - finansowanie: Narodowe Centrum Nauki – program OPUS
   - numer grantu: 2013/09/B/ST6/02317
   - okres realizacji: 2014–2017 (zakończony)
   - rola: **wykonawca**
   - budżet projektu: 664 690 zł
7. **Wspomaganie procesów rozprzestrzeniania treści marketingowych w mediach społecznościowych**
   - finansowanie: Narodowe Centrum Nauki – program OPUS
   - numer grantu: 2016/21/B/HS4/01562
   - okres realizacji: 2017–2021 (zakończony)
   - rola: **wykonawca**
   - budżet projektu: 347 100 zł
8. **Rozprzestrzenianie się wpływu w sieciach wielowarstwowych – Multispread**[8]
   - finansowanie: Narodowe Centrum Nauki – program SONATA
   - numer grantu: 2016/21/D/ST6/02408
   - okres realizacji: 2017–2022 (w trakcie realizacji)
   - rola: **wykonawca**
   - budżet projektu: 291 400 zł
9. **Modele, metody i algorytmy obliczeniowej nauki o sieciach**
   - finansowanie: Narodowe Centrum Nauki – program OPUS
   - numer grantu: 2016/21/B/ST6/01463
   - okres realizacji: 2017–2022 (w trakcie realizacji)
   - rola: **wykonawca**
   - budżet projektu: 685 950 zł
10. **Eksploracja danych w złożonych, społecznych systemach sieciowych**
    - finansowanie: Ministerstwo Nauki i Szkolnictwa Wyższego
    - numer grantu: N 516 510839
    - okres realizacji: 2010–2013 (zakończony)
    - rola: **wykonawca**
    - budżet projektu: 375 100 zł

Przed uzyskaniem stopnia doktora realizowałem projekt 10, po uzyskaniu stopnia doktora realizowałem projekty 1, 4, 5, 7, 8, 9, natomiast projekty 2, 3 i 6 – częściowo przed, częściowo po uzyskaniu stopnia doktora.

Dodatkowo w załączniku 4 do wniosku w punkcie II.15 przedstawiłem wykaz sześciu grantów B+R, z których w dwóch brałem udział jako kierownik i czterech jako wykonawca. Z wyjątkiem jednego grantu tematyka wszystkich dotyczyła różnych aspektów analizy sieci.

[8] O projekcie Multispread: http://multispread.pwr.edu.pl

4.5.4 Współpraca międzynarodowa

Podejmowana przeze mnie współpraca międzynarodowa realizowana jest na kilku polach, głównie ma jednak charakter badawczy – poprzez wspólną pracę badawczą, projekty naukowe i publikacje. Poniżej wymieniłem wybrane instytucje międzynarodowe, z którymi jest lub było mi dane współpracować.

- Imperial College London, Wielka Brytania – współpraca w ramach projektu TRANSFoRm,
- Uppsala University, Uppsala, Szwecja – współpraca w ramach projektu ENGINE,
- King's College London, London, Wielka Brytania – współpraca w ramach projektu TRANSFoRm,
- Trinity College Dublin, Irlandia – współpraca w ramach projektu TRANSFoRm,
- Arizona State University, Tempe, AZ, Stany Zjednoczone – współorganizacja warsztatu Workshop on Social Influence, wspólne numery specjalne w czasopismach,
- University of Zurich, Zurych, Szwajcaria – współpraca badawcza,
- ETH Zurich, Zurych, Szwajcaria – współorganizacja warsztatu Workshop on Social Influence, wspólne numery specjalne w czasopismach,
- Rensselear Politechnic Institute, Troy, NY, Stany Zjednoczone – współpraca badawcza, wizyty stażowe, współpraca w ramach projektów ENGINE oraz RENOIR,
- Eberhard Karls University, Tubingen, Niemcy – współpraca w ramach projektu TRANSFoRm,
- University of California Davis, CA, Stany Zjednoczone – współpraca w ramach projektu RENOIR, staż badawczy,
- Karolinska Institutet, Szwecja – współpraca w ramach projektu TRANSFoRm,
- University of Notre Dame, Indiana, Stany Zjednoczone – współpraca w ramach projektów RENOIR i ENGINE,
- University of California Los Angeles, CA, Stany Zjednoczone – współpraca badawcza, wspólne publikacje,
- University of Dundee, Szkocja – współpraca w ramach projektu TRANSFoRm,
- University of Technology Sydney, Australia – wspólne publikacje, współpraca w ramach projektu RENOIR, staż badawczy, wspólne numery specjalne w czasopismach,
- Carnegie Mellon University, Pittsburgh, PA, Stany Zjednoczone – wspólne publikacje, współpraca w ramach projektu RENOIR,
- Masaryk University, Brno, Czechy – współpraca w ramach projektu CARE-ARSEC,
- Kopint-Tárki Institute, Budapeszt, Węgry – współpraca w ramach projektu CARE-ARSEC,
- Research Centre of the Slovak Foreign Policy Association, Bratysława, Słowacja – współpraca w ramach projektu CARE-ARSEC,
- dodatkowo, poza jednostkami stricte badawczymi, w ramach różnych projektów współpracowałem m.in. z: STA – Slovenian Press Agency (Słowenia), IOTA Foundation (Niemcy), Zcash Foundation (Stany Zjednoczone).

4.5.5 Nagrody i stypendia

Za działalność naukową otrzymałem następujące nagrody i stypendia (do 2014 roku – przed uzyskaniem stopnia doktora, od 2015 roku – po uzyskaniu stopnia doktora):

- Stypendium Ministra Nauki i Szkolnictwa Wyższego dla wybitnych młodych naukowców - w latach 2015–2018,
- Nagroda Jego Magnificencji Rektora Politechniki Wrocławskiej za w uznaniu wyróżniającego wkładu w działalność uczelni – w latach 2016 oraz 2019,
- Laur Dziekana Wydziału Informatyki i Zarządzania za wybitne osiągnięcia naukowe oraz duży wkład w rozwój badań naukowych – w roku 2019,
- Jestem laureatem programu Primus Politechniki Wrocławskiej za publikacje istotnie przyczyniające się do wzrostu punktacji dyscypliny naukowej (działanie drugie) – w roku 2020,
- Nagroda Jego Magnificencji Rektora Politechniki Wrocławskiej za uznane osiągnięcia naukowe - w roku akademickim 2011/2012 oraz 2013/2014,
- Nagroda Dziekana Wydziału Informatyki i Zarządzania Politechniki Wrocławskiej za uznane osiągnięcia naukowe – w roku akademickim 2012/2013,
- Stypendium dla najlepszych doktorantów – w latach akademickich 2012/2013, 2013/2014,
- Stypendium z funduszu własnego Politechniki Wrocławskiej dla najlepszych doktorantów – w semestrach: zima 2012 oraz lato 2013,
- Stypendium projakościowe dla doktorantów – w latach 2012/2013 i 2013/2014,
- Stypendium w ramach projektu „Przedsiębiorczy doktorant – inwestycja w innowacyjny rozwój regionu" ze środków Dolnośląskiego Urzędu Marszałkowskiego oraz Unii Europejskiej w ramach Europejskiego Funduszu Społecznego – lata 2011-2013,
- Laureat konkursu dla rozwoju młodej kadry organizowanego przez Instytut Informatyki Politechniki Wrocławskiej – lata akademickie 2011/2012, 2013/2014, 2014/2015, 2015/2016, 2016/2017,
- Stypendium „Młoda kadra" dla najlepszych doktorantów i naukowców – ze środków Unii Europejskiej w ramach Europejskiego Funduszu Społecznego – w roku 2013,
- Stypendium w ramach projektu „Grant Plus" – ze środków Dolnośląskiego Urzędu Marszałkowskiego oraz Unii Europejskiej w ramach Europejskiego Funduszu Społecznego – w latach 2013-2014.

Dodatkowo **w 2016 roku praca *„Evolutionary algorithm for seed selection in social influence process"* uzyskała nagrodę Best Paper Award** na warsztacie SNAA 2016 organizowanym przy konferencji ASONAM - autorzy pracy: Michał Weskida i Radosław Michalski.

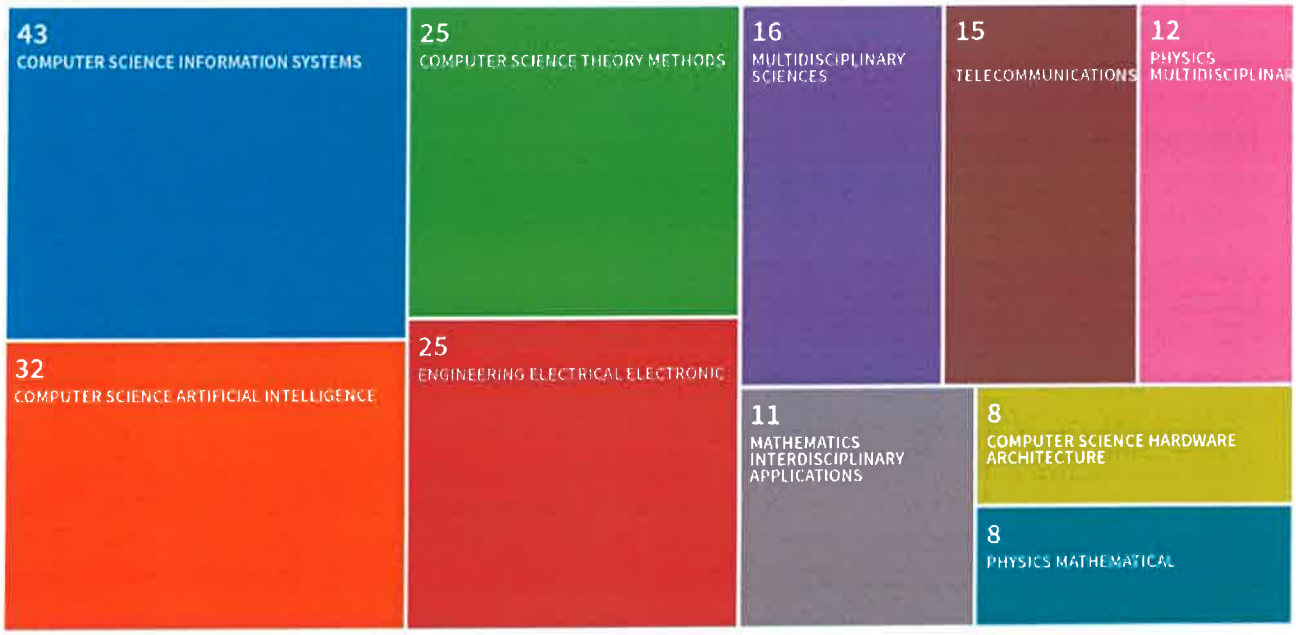
4.5.6 Podsumowanie dorobku wnioskodawcy

W tabeli 4 przedstawiłem podsumowanie mojego dorobku naukowego i relewantnych aktywności naukowych – tam, gdzie było to możliwe i zasadne, w rozbiciu na dorobek przed uzyskaniem stopnia doktora oraz po uzyskaniu stopnia doktora. Na rysunku 2 przedstawiłem obszary reprezentowane przez prace cytujące mój dorobek publikacyjny, natomiast na rysunku 3 afiliacje autorów prac cytujących. Natomiast szczegółowy wykaz moich aktywności naukowych znajduje się w wykazie osiągnięć naukowych – załączniku 4 do niniejszego wniosku.

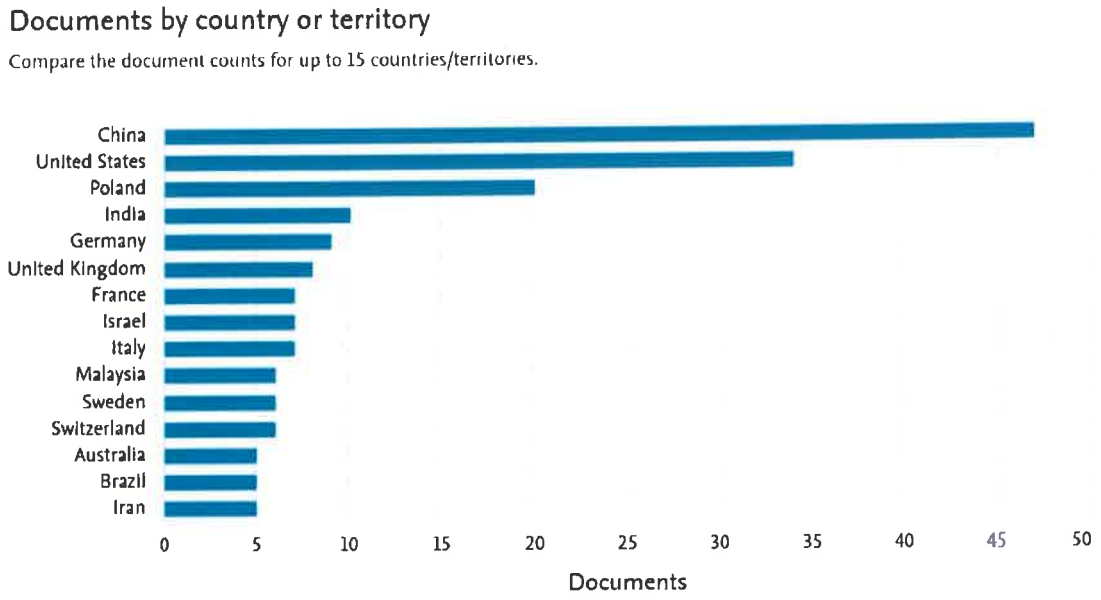
Tabela 4: Podsumowanie dorobku naukowego

Wskaźnik	Przed uzyskaniem stopnia doktora	Po uzyskaniu stopnia doktora	Łącznie
Liczba publikacji	21	26	47
Liczba publikacji indeksowanych w WoS Core C.	17	18	35
Liczba artykułów w czasopismach z IF	1	15	16
Suma współczynnika Impact Factor	0,821	47,504	48,325
Suma punktów z wykazu MEiN z roku publikacji	252	1370	1622
Liczba cytowań[1]			525/337/244
Liczba cytowań bez autocytowań wszystkich autorów[1]			345/203/160
Liczba publikacji cytowanych w WoS Core Collection			29
Indeks Hirscha według Google Scholar			12
Indeks Hirscha według Scopus			9
Indeks Hirscha według Web of Science			8
Liczba realizowanych projektów badawczych (w tym jako kierownik)			10 (1)
Liczba realizowanych projektów B+R (w tym jako kierownik)			6 (2)
Liczba wykonanych recenzji artykułów w czasopismach			78
Liczba staży badawczych (liczba różnych ośrodków)			7 (4)
Liczba zdobytych nagród i stypendiów			23

[1]w ujęciu: Google Scholar/Scopus/Web of Science, stan na dzień 23 kwietnia 2021 roku



Rysunek 2: Obszary reprezentowane przez prace, które cytują dorobek wnioskodawcy (bez autocytowań wszystkich autorów). Źródło: baza Web of Science.



Rysunek 3: Kraje afiliacji autorów prac, którzy cytują dorobek wnioskodawcy (bez autocytowań wszystkich autorów, pierwsze piętnaście krajów). Źródło: baza Scopus.

5 Informacja o wykazywaniu się istotną aktywnością naukową albo artystyczną realizowaną w więcej niż jednej uczelni, instytucji naukowej lub instytucji kultury, w szczególności zagranicznej

W tym punkcie – zgodnie z tytułem – wymieniłem ośrodki badawcze, w których prowadziłem działalność badawczą w ramach staży naukowych. Natomiast wszystkie ośrodki międzynarodowe, z którymi współpracowałem lub współpracuję, wymieniłem w punkcie 4.5.4. Potwierdzenia staży realizowanych po uzyskaniu stopnia doktora przedstawiono w załączniku 6 niniejszego wniosku – w załączniku tym także umieszczono potwierdzenie szkolenia zrealizowanego w IVEY Business School przy Western University w Kanadzie opisane w punkcie 6.5.

W ramach realizacji projektów międzynarodowych opisanych w punkcie 4.5.3 miałem możliwość realizacji wizyt stażowych na kilku uniwersytetach będących partnerami projektów - łącznie było to cztery i pół miesiąca. Najsilniejsza współpraca wynikająca ze wspólnych zainteresowań badawczych miała miejsce z **prof. Bolesławem Szymańskim z Rensselear Politechnic Institute** (Troy, NY, Stany Zjednoczone). Pierwsza – krótsza – wizyta miała miejsce po obronie doktoratu w grudniu 2014 roku, następne dwie – każda miesięczna – w maju 2015 roku oraz w lutym 2017 roku. W ramach wspólnych prac badawczych realizowanych także w trakcie powyższych wizyt opracowano i przebadano koncepcję sekwencyjnej aktywacji węzłów dla sieci statycznych [**H2, H4**] – koncepcja ta opracowana była wspólnie z dr. hab. inż. Jarosławem Jankowskim, prof. ZUT oraz innymi współautorami, obecnie w dalszym ciągu jest rozwijana, przykładowo dla sieci temporalnych [**H9**] i wielowarstwowych [**H10**].

Dodatkowo, w trakcie ostatniej wizyty na RPI w lutym 2017 roku wraz z ówczesnym doktorantem Marcinem Kulisiewiczem, którego byłem promotorem pomocniczym (doktorat obroniony w lipcu 2020 roku – z wyróżnieniem), pracowaliśmy nad koncepcją entropii jako miary dynamiki interakcji węzłów w sieciach, co doprowadziło do opublikowania pracy [**H5**]. Ponadto, w trakcie wizyt i współpracy często wracaliśmy do wyzwań związanych z modelowaniem sieci temporalnych. Rozmowy te doprowadziły do rozszerzenia grupy badawczej o kognitywistę prof. Christiana Lebiere z Carnegie Mellon University oraz socjologa prof. Omara Lizardo z University of California Los Angeles i opracowania oraz przebadania modelu CogSNet, który został opublikowany w pracy [**H11**]. Obecnie trwają prace nad personalizacją modelu oraz aproksymacją wartości stanu sieci modelowanej za pomocą CogSNet.

Kolejnym ośrodkiem badawczym, z którym współpracuję, jest **University of Technology Sydney** (Australia). W lipcu 2019 roku w ramach projektu RENOIR spędziłem tam miesiąc realizując dwa cele związane z badaniami: wraz z prof. Katarzyną Musiał-Gabryś pracowaliśmy nad artykułem przeglądowym z zakresu modelowania wpływu w sieciach społecznych (spodziewane zgłoszenie do czasopisma w drugiej połowie 2021 roku), natomiast z dr. Marianem-Andreiem Rizoui oraz doktorantem Mateuszem Nurkiem pracujemy nad wnioskowaniem ze strumieni interakcji z pomocą procesów Hawkesa (praca w recenzji, preprint arXiv:2009.02032). W obu wyżej wymienionych zagadnieniach badawczych współpracuję także z prof. Omarem Lizardo z University of California Los Angeles.

Ponadto, w ramach projektów ENGINE i RENOIR odbyłem staże badawcze w dwóch innych ośrodkach badawczych, **Uppsala University** (Uppsala, Szwecja – prof. Matteo Magnani) oraz **University of California Davis** (Davis, CA, Stany Zjednoczone – prof. Raissa D'Souza). Wizyty te pozwoliły skonsultować założenia i dokończyć pracę nad dwiema publikacjami cyklu, odpowiednio **[H3]** oraz **[H6]**.

Jednocześnie w tym punkcie należy wspomnieć o istotnej współpracy realizowanej z ośrodkiem krajowym. W ramach wieloletniej współpracy z dr. hab. inż. Jarosławem Jankowskim, profesorem Zachodniopomorskiego Uniwersytetu Technologicznego udało nam się zaproponować, przebadać i opublikować wiele koncepcji, w tym m.in. wspomniane już prace dotyczące sekwencyjnej aktywacji węzłów. Jednym z obszarów współpracy był grant NCN OPUS kierowany przez dr. hab. Jarosława Jankowskiego, którego byłem wykonawcą.

Warto także podkreślić, że owocna współpraca ze wszystkimi powyższymi ośrodkami skutkuje także tym, że badacze z tych ośrodków chętnie dołączają do kolejnych inicjatyw badawczych, np. grantów, których jestem wnioskodawcą.

6 Informacja o osiągnięciach dydaktycznych, organizacyjnych oraz popularyzujących naukę lub sztukę

6.1 Opracowanie nowych kierunków studiów oraz kursów

1. **Brałem udział w opracowaniu dwóch programów studiów magisterskich** realizowanych na Wydziale Informatyki i Zarządzania Politechniki Wrocławskiej: **specjalności *Data Science – Danologia***[9] oraz **kierunku *Sztuczna inteligencja***[10].
2. **Brałem udział w opracowaniu programu studiów dla kierunku *Sztuczna inteligencja* na studiach inżynierskich**. Program jest gotowy od strony merytorycznej i formalnej, jednak jeszcze nie został uruchomiony.
3. **Opracowałem następujące kursy: *Architektura komputerów*** (I stopień – informatyka, w języku polskim i angielskim), ***Analiza danych złożonych*** (II stopień – Data Science) oraz ***Przetwarzanie danych złożonych*** (II stopień – Sztuczna inteligencja).
4. W ramach programu Erasmus+ Unii Europejskiej – Staff Mobility for Teaching **opracowałem autorski kurs z zakresu nauki o sieciach pt. *„Crash Course in Network Science"*** (seria wykładów i laboratoriów) oraz przeprowadziłem go w trakcie wizyt na następujących uczelniach: National University of Science and Technology MISIS (Moskwa, Rosja, 2016), Jigme Namgyel Engineering College (Deothang, Bhutan, 2017) oraz Kathmandu University (Dhulikhel, Nepal, 2019). Łączną liczbę aktywnie zaangażowanych uczestników wszystkich trzech edycji kursu szacuję na około 200 studentów. Następnie, już z Polski, dla ok. 35 z nich prowadziłem indywidualne zajęcia podnoszące zakres wiedzy w analizie danych i sieci.

[9] Witryna kierunku *Data Science*: https://studiujdatascience.pl

[10] Witryna kierunku *Sztuczna inteligencja*: https://studiuj.ai

6.2 Opieka naukowa

1. Byłem **promotorem pomocniczym w przewodzie doktorskim doktora inżyniera Marcina Kulisiewicza pt. *„Event-driven Temporal Social Networks”*** – rozprawa doktorska napisana została w języku angielskim, głównym promotorem był prof. dr hab. inż. Przemysław Kazienko. Rozprawę obroniono 14 lipca 2020 roku na Wydziale Informatyki i Zarządzania Politechniki Wrocławskiej i decyzją Rady Wydziału **została ona wyróżniona**.
2. Obecnie **jestem promotorem pomocniczym doktoranta Mateusza Nurka**, który realizuje studia doktoranckie w Szkole Doktorskiej Politechniki Wrocławskiej, w dyscyplinie *informatyka techniczna i telekomunikacja*. Temat realizowany w ramach pracy badawczej doktoranta brzmi *„Strumienie danych w sieciach społecznych”*.
3. Byłem **promotorem 19 prac inżynierskich oraz 18 prac magisterskich** studentów obronionych na kierunku informatyka Wydziału Informatyki i Zarządzania Politechniki Wrocławskiej. Dodatkowo, byłem także promotorem 11 obronionych prac na Wydziale Matematyki PWr realizowanych w tematyce uczenia maszynowego i sieci złożonych (prace magisterkie i inżynierskie) oraz 23 prac inżynierskich obronionych w Wyższej Szkole Informatyki i Zarządzania „Copernicus”.
4. W 2017 roku **założyłem i od tego czasu sprawuję opiekę nad nieformalną grupą naukową BERG – Blockchain Exploration Research Group**[11]. Celem grupy, do której należą studenci oraz absolwenci Politechniki Wrocławskiej a także inne osoby zainteresowane tematem - także spoza Polski (obecnie ok. 12 stałych członków), jest analiza łańcucha bloków (ang. *blockchain*) jako sieci złożonej z wykorzystaniem metod nauki o sieciach. W szczególności w ramach grupy staram się oferować studentom i dyplomantom możliwość rozwoju w badaniach naukowych, stąd m.in. opublikowane wspólnie artykuły naukowe z tej tematyki czy też wspólna realizacja dwóch grantów przemysłowych o charakterze badawczym – opisane w innych częściach wniosku.
5. Dla studentów pierwszego stopnia studiów zainteresowanych nabywaniem umiejętności w zakresie data science i uczenia maszynowego **oferuję udział w inicjatywie Datacamp for the Classroom**[12] i w ramach tej inicjatywy **sprawuję nad tymi studentami indywidualną opiekę w zakresie rozwoju w tych obszarach** – corocznie w programie uczestniczy ok. 70 studentów, z czego ok. 5–10 osób decyduje się na ścieżkę indywidualną, często prowadzącą do prac inżynierskich czy też magisterskich o oryginalnym charakterze badawczym.
6. **Opracowałem *„Poradnik dla dyplomantów”* ułatwiający moim i innym dyplomantom pracę nad swoimi pracami dyplomowymi**[13] – z własnych obserwacji i informacji pozyskanych od studentów znacząco podnosi on jakość opracowywanych prac dyplomowych.

[11] Witryna grupy BERG – Blockchain Exploration Research Group: https://bergplace.org
[12] Inicjatywa Datacamp for the Classroom: https://www.datacamp.com/groups/education
[13] Poradnik dla dyplomantów: https://www.ii.pwr.edu.pl/~michalski/media/poradnik.pdf

6.3 Wykłady zaproszone

Po doktoracie wygłosiłem następujące wykłady zaproszone:

1. ***„Diffusion Processes in Complex Networks – Current Challenges"*** – **Southwest University, Chongqing, Chiny** (grudzień 2015) – wykład na zaproszenie profesora Tao Jia.
2. ***„Social Influence in Temporal Social Networks"*** – **University of Zurich, Zurych, Szwajcaria** (maj 2016) – wykład na zaproszenie profesora Claudio Tessone.
3. ***„What Data do we Need to Measure Resilience of Real World Systems?"*** – **Complexity Science Hub Vienna, Wiedeń, Austria** (czerwiec 2018) – wykład na zaproszenie profesorów Franka Schweitzera i Stefana Thurnera w ramach warsztatów „Improving Resilience in Complex Networks".

Dodatkowo w ramach wizyt stażowych i krótkich wizyt badawczych, wyniki własnych badań prezentowałem w ramach seminariów wydziałowych lub seminariów poszczególnych grup badawczych na następujących uniwersytetach: **Rensselaer Polytechnic Institute** (Troy, NY, Stany Zjednoczone), **Notre Dame University** (Notre Dame, IN, Stany Zjednoczone), **University of California Davis** (Davis, CA, Stany Zjednoczone), **Uppsala University** (Uppsala, Szwecja), **University of Technology Sydney** (Sydney, Australia).

6.4 Prowadzone kursy

1. Architektura komputerów (I stopień, opracowanie, wykład i laboratorium, także w języku angielskim).
2. Analiza danych złożonych (II stopień, opracowanie, wykład i projekt).
3. Integracja systemów informacyjnych (I stopień, laboratorium).
4. Języki programowania baz danych Oracle (I stopień, laboratorium).
5. Tworzenie interfejsów użytkownika (I stopień, laboratorium, w języku angielskim).
6. Hurtownie danych (I stopień, laboratorium, także w języku angielskim).
7. Paradygmaty programowania (I stopień, ćwiczenia).
8. Organizacja systemów komputerowych (I stopień, ćwiczenia, w języku angielskim).
9. Sztuczna inteligencja i inżynieria wiedzy (I stopień, ćwiczenia).
10. Przetwarzanie danych złożonych (II stopień, opracowanie, wykład i laboratoria – kurs będzie dopiero uruchamiany).
11. Projekt naukowo-wdrożeniowy (II stopień, projekt).
12. Praca dyplomowa (prace inżynierskie i magisterskie, w języku polskim i angielskim).

6.5 Pozostałe osiągnięcia dydaktyczne, organizacyjne oraz popularyzujące naukę

1. **W 2020 roku zająłem drugie miejsce w plebiscycie na najlepszego prowadzącego na kierunku Informatyka na Wydziale Informatyki i Zarządzania Politechniki Wrocławskiej** organizowanym przez Samorząd Studencki wydziału.
2. W ramach projektu ENGINE byłem **pomysłodawcą i osobą kierującą stworzeniem platformy sci·cup**[14]. Jest to platforma konkursowa, która skupia środowisko osób zajmujących się analizą danych. Podobnie jak serwis Kaggle, pozwala na tworzenie konkursów bazujących na problemach i danych firm zewnętrznych, w których mogą brać udział wszyscy zainteresowani proponując swoje rozwiązania – więcej informacji o platformie znajduje się w załączniku 4 wniosku, punkt III.2.

[14]Platforma sci·cup: https://scicup.com

3. Jestem **członkiem Rady Konsultacyjnej Wydziału Informatyki i Zarządzania Politechniki Wrocławskiej** (od roku akademickiego 2019/2020).
4. W ramach projektu ENGINE **stworzyłem nowe laboratorium sieci złożonych i analiz dużych zbiorów danych**, które zapewnia wydajne i skalowalne środowisko dla analiz dużych danych, w tym do modelowania i wizualizacji. Jest ono wykorzystywane do analiz sieci złożonych (w tym społecznych), analizy tekstu, uczenia maszynowego czy przetwarzania strumieni sygnałów. Wartość projektu laboratorium wyniosła €490 000.
5. Jestem **recenzentem prac w Ogólnopolskim konkursie na najlepsze prace magisterskie z informatyki**[15] organizowanym przez Polskie Towarzystwo Informatyczne (edycje w latach 2015, 2018 oraz 2020).
6. Jestem **dolnośląskim regionalnym koordynatorem Konkursu GEEK – Gry Eksperymentalne Edukacyjne Komputerowe**[16] polegającego na tworzeniu dedykowanych gier komputerowych wspomagających nauczanie przedmiotów (niekoniecznie informatyki) w szkołach podstawowych i średnich - konkurs adresowany właśnie do uczniów szkół na tych dwóch poziomach. Organizatorem konkursu jest Polskie Towarzystwo Informatyczne (2021).
7. Jestem **pomysłodawcą oraz *Chairem* warsztatu poświęconego analizie wpływu społecznego w sieciach społecznych - Workshop on Social Influence**[17]. Warsztat organizowany jest od 2014 roku, a oprócz mnie *Chairami* są prof. Paulo Shakarian z Arizona State University (Tempe, AZ, Stany Zjednoczone) oraz dr hab. inż. Jarosław Jankowski, prof. Zachodniopomorskiego Uniwersytetu Technologicznego w Szczecinie. Do tej pory odbyło się sześć edycji warsztatu, w większości przy międzynarodowej konferencji ASONAM.
8. W ramach podnoszenia kompetencji z zarządzania projektami i ich komercjalizacji, brałem udział w: **(i) programie Transformation.Doc** realizowanym przez Ministerstwo Nauki i Szkolnictwa Wyższego – celem programu było wzmocnienie kompetencji jego uczestników w zakresie przedsiębiorczości i umiejętności miękkich niezbędnych do współpracy z gospodarką. W ramach tego programu w 2015 roku odbyłem szkolenie w Ivey Business School przy University of Western Ontario (Londyn, Ontario, Kanada) w 2015 roku – potwierdzenie udziału umieszczam w załączniku 6 do wniosku; **(ii) szkoleniu z zarządzania projektami badawczymi** prowadzonym przez organizację Vitae z Wielkiej Brytanii w ramach projektu SKILLS Fundacji na rzecz Nauki Polskiej.
9. Jednocześnie dbam o to, aby **wyniki badań i własne doświadczenie mogły służyć także pozostałym obszarom funkcjonowania społeczeństwa**. Przykładem tego może być rozpoczęcie w Polsce w 2017 roku popularyzacji niemieckiego projektu niskokosztowych czujników smogu – Sensor.Community (poprzednia nazwa – Luftdaten)[18] – po rozwinięciu projektu w Polsce, obecnie na bazie gromadzonych przez czujniki danych prowadzę wraz z dyplomantami szereg analiz z zakresu korelacji i predykcji poziomu smogu w polskich miastach. Biorę także udział w hackathonach związanych z danymi miejskimi, np. Code for Freedom w 2014 roku w Gdańsku czy HackWro#6 w 2015 roku we Wrocławiu – **w tym ostatnim zwyciężyło rozwiązanie zaproponowane i zakodowane przez zespół, którego byłem liderem**.

[15] Witryna internetowa konkursu PTI na najlepsze prace magisterskie z informatyki: https://kpm.pti.org.pl
[16] Witryna internetowa konkursu GEEK: https://mlodzi.pti.org.pl
[17] Witryna internetowa warsztatów Workshop on Social Influence: https://wosinf.org
[18] Projekt Sensor.Community: https://sensor.community/pl/

7 Inne informacje dotyczące kariery zawodowej wnioskodawcy

Wykaz wszystkich aktywności naukowych zawierający także poszerzone informacje dotyczące niektórych aktywności wymienionych w tym załączniku oraz rozbicie na osiągnięcia przed i po uzyskaniu stopnia doktora znajduje się w załączniku 4 do niniejszego wniosku.

Radosław Michalski

(podpis wnioskodawcy)

Bibliografia

[1] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, APS, 2002.

[2] Alberto Aleta and Yamir Moreno. Multilayer networks in a nutshell. *Annual Review of Condensed Matter Physics*, 10:45–62, Annual Reviews, 2019.

[3] Alex Arenas, Alberto Fernandez, and Sergio Gomez. Analysis of the structure of complex networks at different resolution levels. *New journal of physics*, 10(5):053039, IOP Publishing, 2008.

[4] I Artico, I Smolyarenko, V Vinciotti, and EC Wit. How rare are power-law networks really? *Proceedings of the Royal Society A*, 476(2241):20190742, The Royal Society Publishing, 2020.

[5] Aritra Banerjee, Michael Clear, and Hitesh Tewari. Demystifying the role of zk-snarks in zcash. In *2020 IEEE Conference on Application, Information and Network Security (AINS)*, pp. 12–19. IEEE, 2020.

[6] Baruch Barzel and Albert-László Barabási. Universality in network dynamics. *Nature physics*, 9(10):673–681, Nature Publishing Group, 2013.

[7] Pavlos Basaras, George Iosifidis, Dimitrios Katsaros, and Leandros Tassiulas. Identifying influential spreaders in complex multilayer networks: A centrality perspective. *IEEE Transactions on Network Science and Engineering*, 6(1):31–45, IEEE, 2017.

[8] Song Bian, Qintian Guo, Sibo Wang, and Jeffrey Xu Yu. Efficient algorithms for budgeted influence maximization on massive social networks. *Proceedings of the VLDB Endowment*, 13(9):1498–1510, VLDB Endowment, 2020.

[9] Stefano Boccaletti, Vito Latora, Yamir Moreno, Martin Chavez, and D-U Hwang. Complex networks: Structure and dynamics. *Physics reports*, 424(4-5):175–308, Elsevier, 2006.

[10] Piotr Bródka, Jarosław Jankowski, and **Radosław Michalski**. Sequential seeding in multilayer networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 31(3):033130, AIP Publishing, 2021.

[11] Anna D Broido and Aaron Clauset. Scale-free networks are rare. *Nature communications*, 10(1):1–10, Nature Publishing Group, 2019.

[12] Yi Chen, Xiaolong Wang, Junzhao Bu, Buzhou Tang, and Xin Xiang. Network structure exploration in networks with node attributes. *Physica A: Statistical Mechanics and its Applications*, 449:240–253, Elsevier, 2016.

[13] Manlio De Domenico, Clara Granell, Mason A Porter, and Alex Arenas. The physics of spreading processes in multilayer networks. *Nature Physics*, 12(10):901–906, Nature Publishing Group, 2016.

[14] Paul Erdős and Alfréd Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60, 1960.

[15] Jean-Francois Ethier, Vasa Curcin, Mark M McGilchrist, Sarah N Lim Choi Keung, Lei Zhao, Anna Andreasson, Piotr Bródka, **Radosław Michalski**, Theodoros N. Arvanitis, Nikolaos Mastellos, et al. eSource for clinical trials: Implementation and evaluation of a standards-based approach in a real world trial. *International journal of medical informatics*, 106:17–24, Elsevier, 2017.

[16] Agata Fronczak and Piotr Fronczak. *Świat sieci złożonych: Od fizyki do Internetu.* Wydawnictwo Naukowe PWN, 2009.

[17] Laetitia Gauvin, Mathieu Génois, Márton Karsai, Mikko Kivelä, Taro Takaguchi, Eugenio Valdano, and Christian L Vestergaard. Randomized reference models for temporal networks. *arXiv preprint arXiv:1806.04032*, 2018.

[18] Edgar N Gilbert. Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, JSTOR, 1959.

[19] Jacob Goldenberg, Barak Libai, and Eitan Muller. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing letters*, 12(3):211–223, Springer, 2001.

[20] Michael Golosovsky. Power-law citation distributions are not scale-free. *Physical Review E*, 96(3):032306, APS, 2017.

[21] Amit Goyal, Francesco Bonchi, Laks VS Lakshmanan, and Suresh Venkatasubramanian. On minimizing budget and time in influence propagation over social networks. *Social network analysis and mining*, 3(2):179–192, Springer, 2013.

[22] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 855–864, 2016.

[23] Daniel Gruhl, Ramanathan Guha, David Liben-Nowell, and Andrew Tomkins. Information diffusion through blogspace. In *Proceedings of the 13th international conference on World Wide Web*, pp. 491–501, 2004.

[24] Petter Holme. Rare and everywhere: Perspectives on scale-free networks. *Nature communications*, 10(1):1–3, Nature Publishing Group, 2019.

[25] Petter Holme and Jari Saramäki. Temporal networks. *Physics reports*, 519(3):97–125, Elsevier, 2012.

[26] Petter Holme and Jari Saramäki. *Temporal network theory*, vol. 2. Springer, 2019.

[27] Shankar Iyer and Lada A Adamic. The costs of overambitious seeding of social products. In *International Conference on Complex Networks and their Applications*, pp. 273–286. Springer, 2018.

[28] Jarosław Jankowski, Piotr Bródka, Przemysław Kazienko, Bolesław K. Szymański, **Radosław Michalski**, and Tomasz Kajdanowicz. Balancing speed and coverage by sequential seeding in complex networks. *Scientific reports*, 7(1):1–11, Nature Publishing Group, 2017.

[29] Jarosław Jankowski, Piotr Bródka, **Radosław Michalski**, and Przemysław Kazienko. Seeds buffering for information spreading processes. In *SocInfo 2017: International Conference on Social Informatics*, vol. 10539 of *Lecture Notes in Computer Science*, pp. 628–641. Springer, Cham, 2017.

[30] Jarosław Jankowski, Bolesław K. Szymański, Przemysław Kazienko, **Radosław Michalski**, and Piotr Bródka. Probing limits of information spread with sequential seeding. *Scientific reports*, 8(1):1–9, Nature Publishing Group, 2018.

[31] Jarosław Jankowski and **Radosław Michalski**. Increasing coverage of information spreading in social networks with supporting seeding. In *DMBD 2017: International Conference on Data Mining and Big Data*, vol. 10387 of *Lecture Notes in Computer Science*, pp. 209–218. Springer, Cham, 2017.

[32] Jarosław Jankowski, **Radosław Michalski**, and Piotr Bródka. A multilayer network dataset of interaction and influence spreading in a virtual world. *Scientific data*, 4(1):1–9, Nature Publishing Group, 2017.

[33] Jarosław Jankowski, **Radosław Michalski**, Piotr Bródka, and Artur Karczmarczyk. Increasing coverage of information diffusion processes by reducing the number of initial seeds. In *ASONAM 2017: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 713–720. IEEE, 2017.

[34] Jarosław Jankowski, **Radosław Michalski**, Piotr Bródka, Przemysław Kazienko, and Sonja Utz. Adaptive survey design using structural characteristics of the social network. In *SocInfo 2015: International Conference on Social Informatics*, vol. 9471 of *Lecture Notes in Computer Science*, pp. 153–163. Springer, Cham, 2015.

[35] Jarosław Jankowski, **Radosław Michalski**, Piotr Bródka, Przemysław Kazienko, and Sonja Utz. Knowledge acquisition from social platforms based on network distributions fitting. *Computers in Human Behavior*, 51:685–693, Elsevier, 2015.

[36] Jarosław Jankowski, Marcin Waniek, Aamena Alshamsi, Piotr Bródka, and **Radosław Michalski**. Strategic distribution of seeds to support diffusion in complex networks. *PloS one*, 13(10):e0205130, Public Library of Science, 2018.

[37] Tomasz Kajdanowicz, **Radosław Michalski**, Katarzyna Musiał, and Przemysław Kazienko. Learning in unlabeled networks – an active learning and inference approach. *AI Communications*, 29(1):123–148, IOS Press, 2016.

[38] Fariba Karimi and Petter Holme. Threshold model of cascades in empirical temporal networks. *Physica A: Statistical Mechanics and its Applications*, 392(16):3476–3483, Elsevier, 2013.

[39] Márton Karsai, Mikko Kivelä, Raj Kumar Pan, Kimmo Kaski, János Kertész, A-L Barabási, and Jari Saramäki. Small but slow world: How network topology and burstiness slow down spreading. *Physical Review E*, 83(2):025102, APS, 2011.

[40] David Kempe, Jon Kleinberg, and Éva 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, 2003.

[41] Mikko Kivelä, Alex Arenas, Marc Barthelemy, James P Gleeson, Yamir Moreno, and Mason A Porter. Multilayer networks. *Journal of complex networks*, 2(3):203–271, Oxford University Press, 2014.

[42] Marcin Kulisiewicz, Przemysław Kazienko, Bolesław K. Szymański, and **Radosław Michalski**. Entropy measures of human communication dynamics. *Scientific reports*, 8(1):1–8, Nature Publishing Group, 2018.

[43] Timothy La Fond and Jennifer Neville. Randomization tests for distinguishing social influence and homophily effects. In *Proceedings of the 19th international conference on World wide web*, pp. 601–610, 2010.

[44] Christine Largeron, Pierre-Nicolas Mougel, Reihaneh Rabbany, and Osmar R Zaïane. Generating attributed networks with communities. *PloS one*, 10(4):e0122777, Public Library of Science, 2015.

[45] Jundong Li, Kewei Cheng, Liang Wu, and Huan Liu. Streaming link prediction on dynamic attributed networks. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pp. 369–377, 2018.

[46] Xiao Fan Liu, Xin-Jian Jiang, Si-Hao Liu, and Chi Kong Tse. Knowledge discovery in cryptocurrency transactions: A survey. *IEEE Access*, 9:37229–37254, IEEE, 2021.

[47] Liye Ma, Ramayya Krishnan, and Alan L Montgomery. Latent homophily or social influence? an empirical analysis of purchase within a social network. *Management Science*, 61(2):454–473, INFORMS, 2015.

[48] Naoki Masuda and Renaud Lambiotte. *Guide To Temporal Networks, A*, vol. 6. World Scientific, 2020.

[49] Tsuyoshi Murata and Hokuto Koga. Approximation methods for influence maximization in temporal networks. In *Temporal Network Theory*, pp. 345–368. Springer, 2019.

[50] Jennifer Neville and David Jensen. Collective classification with relational dependency networks. In *Workshop on Multi-Relational Data Mining (MRDM-2003)*, p. 77, 2003.

[51] Mateusz Nurek and **Radosław Michalski**. Combining machine learning and social network analysis to reveal the organizational structures. *Applied Sciences*, 10(5):1699, MDPI, 2020.

[52] Jungseul Ok, Youngmi Jin, Jinwoo Shin, and Yung Yi. On maximizing diffusion speed in social networks: impact of random seeding and clustering. In *The 2014 ACM international conference on Measurement and modeling of computer systems*, pp. 301–313, 2014.

[53] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *Reviews of modern physics*, 87(3):925, APS, 2015.

[54] Patryk Pazura, Kamil Bortko, Jarosław Jankowski, and **Radosław Michalski**. A dynamic vote-rank based approach for effective sequential initialization of information spreading processes within complex networks. In *ICCS 2020: International Conference on Computational Science*, vol. 12137 of *Lecture Notes in Computer Science*, pp. 638–651. Springer, Cham, 2020.

[55] Pierre Perrault, Jennifer Healey, Zheng Wen, and Michal Valko. Budgeted online influence

maximization. In *International Conference on Machine Learning*, pp. 7620–7631. PMLR, 2020.

[56] René Pfitzner, Ingo Scholtes, Antonios Garas, Claudio J Tessone, and Frank Schweitzer. Betweenness preference: Quantifying correlations in the topological dynamics of temporal networks. *Physical review letters*, 110(19):198701, APS, 2013.

[57] William Rand and Roland T Rust. Agent-based modeling in marketing: Guidelines for rigor. *International Journal of research in Marketing*, 28(3):181–193, Elsevier, 2011.

[58] Everett M Rogers. *Diffusion of innovations*. Simon and Schuster, 2010.

[59] Przemysław Różewski, Jarosław Jankowski, Piotr Bródka, and **Radosław Michalski**. Knowledge workers' collaborative learning behavior modeling in an organizational social network. *Computers in Human Behavior*, 51:1248–1260, Elsevier, 2015.

[60] Ingo Scholtes, Nicolas Wider, René Pfitzner, Antonios Garas, Claudio J Tessone, and Frank Schweitzer. Causality-driven slow-down and speed-up of diffusion in non-markovian temporal networks. *Nature communications*, 5(1):1–9, Nature Publishing Group, 2014.

[61] Claude E Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, Nokia Bell Labs, 1948.

[62] Qihao Shi, Can Wang, Jiawei Chen, Yan Feng, and Chun Chen. Post and repost: A holistic view of budgeted influence maximization. *Neurocomputing*, 338:92–100, Elsevier, 2019.

[63] Aaron Striegel, Shu Liu, Lei Meng, Christian Poellabauer, David Hachen, and Omar Lizardo. Lessons learned from the netsense smartphone study. *ACM SIGCOMM Computer Communication Review*, 43(4):51–56, ACM New York, NY, USA, 2013.

[64] **Radosław Michalski**. Linear threshold model in temporal networks—Seed selection for social influence. In *ASONAM 2015: Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 922–923. IEEE, 2015.

[65] **Radosław Michalski**. Analysing and predicting the adoption of anonymous transactions in cryptocurrencies. In *BIS 2020: International Conference on Business Information Systems*, vol. 394 of *Lecture Notes in Business Information Processing*, pp. 132–144. Springer, Cham, 2020.

[66] **Radosław Michalski**, Daria Dziubałtowska, and Piotr Macek. Revealing the character of nodes in a blockchain with supervised learning. *IEEE Access*, 8:109639–109647, IEEE, 2020.

[67] **Radosław Michalski**, Jarosław Jankowski, and Piotr Bródka. Effective influence spreading in temporal networks with sequential seeding. *IEEE Access*, 8:151208–151218, IEEE, 2020.

[68] **Radosław Michalski**, Jarosław Jankowski, and Patryk Pazura. Entropy-based measure for influence maximization in temporal networks. In *ICCS 2020: International Conference on Computational Science*, vol. 12140 of *Lecture Notes in Computer Science*, pp. 277–290. Springer, Cham, 2020.

[69] **Radosław Michalski** and Przemysław Kazienko. Maximizing social influence in real-world networks – the state of the art and current challenges. In Dariusz Król, Damien Fay, and Bogdan Gabryś, editors, *Propagation Phenomena in Real World Networks*, vol. 85 of *Intelligent Systems Reference Library (Janusz Kacprzyk and Lakhmi C. Jain, Series Editors)*, pp. 329–359. Springer, Cham, 2015.

[70] **Radosław Michalski** and Przemysław Kazienko. Social network analysis in organizational structures evaluation. In Reda Alhajj and Jon Rokne, editors, *Encyclopedia of Social Network Analysis and Mining, Living Reference Work*, pp. 1–13. Springer, New York, 2017.

[71] **Radosław Michalski**, Sebastian Palus, and Przemysław Kazienko. Matching organizational structure and social network extracted from email communication. In *BIS 2011: International Conference on Business Information Systems*, vol. 87 of *Lecture Notes in Business Information Processing*, pp. 197–206. Springer, Berlin Heidelberg, 2011.

[72] **Radosław Michalski** and Marcin Pieczka. Dru: Studying blockchain as a complex network. In *ASONAM 2020: Proceedings of the 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 929–932. IEEE, 2020.

[73] **Radosław Michalski**, Bolesław K. Szymański, Przemysław Kazienko, Christian Lebiere, Omar Lizardo, and Marcin Kulisiewicz. Social networks through the prism of cognition. *Complexity*, 2021:4963903, Wiley/Hindawi, 2021.

[74] Gerald L Thompson. Hamiltonian tours and paths in rectangular lattice graphs. *Mathematics Magazine*, 50(3):147–150, Taylor & Francis, 1977.

[75] Thomas W Valente, Kathryn Coronges, Cynthia Lakon, and Elizabeth Costenbader. How correlated are network centrality measures? *Connections (Toronto, Ont.)*, 28(1):16, NIH Public Access, 2008.

[76] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442, Nature Publishing Group, 1998.

[77] Michał Weskida and **Radosław Michalski**. Evolutionary algorithm for seed selection in social influence process. In *ASONAM 2016: Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 1189–1196. IEEE, 2016.

[78] Michał Weskida and **Radosław Michalski**. Finding influentials in social networks using evolutionary algorithm. *Journal of Computational Science*, 31:77–85, Elsevier, 2019.

[79] John T Wixted and Ebbe B Ebbesen. On the form of forgetting. *Psychological science*, 2(6):409–415, SAGE Publications Sage CA: Los Angeles, CA, 1991.

[80] Soon-Hyung Yook, Hawoong Jeong, and Albert-László Barabási. Modeling the internet's large-scale topology. *Proceedings of the National Academy of Sciences*, 99(21):13382–13386, National Acad Sciences, 2002.

[81] Jian-Xiong Zhang, Duan-Bing Chen, Qiang Dong, and Zhi-Dan Zhao. Identifying a set of influential spreaders in complex networks. *Scientific reports*, 6:27823, Nature Publishing Group, 2016.

[82] Lu Zhong, Chao Gao, Zili Zhang, Ning Shi, and Jiajin Huang. Identifying influential nodes in complex networks: a multiple attributes fusion method. In *International conference on active media technology*, pp. 11–22. Springer, 2014.