

Tadeusz Morzy
prof. dr hab. inż.
Instytut Informatyki
Politechniki Poznańskiej
email: Tadeusz.morzy@put.poznan.pl

Poznań, 3.06.2022 r.

Ocena

dorobku naukowego i rozprawy habilitacyjnej dra inż. Mariusza Kamoli **p.t. *Metody analizy i odkrywania struktury sieci społecznościowych i technologicznych***

Niniejsza opinia została przygotowana w odpowiedzi na pismo dr. hab. inż., prof. PW, Jarosława Arabasa Przewodniczącego Rady Naukowej Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Warszawskiej w związku z powołaniem mnie przez Radę Naukową Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Warszawskiej na recenzenta w przewodzie habilitacyjnym dra inż. Mariusza Kamoli. Moja opinia, przygotowana zgodnie z obowiązującymi przepisami, dotyczy oceny osiągnięcia naukowego habilitanta dokonanej na podstawie cyklu prac stanowiących rozprawę habilitacyjną pt. *Metody analizy i odkrywania struktury sieci społecznościowych i technologicznych*, udokumentowanej cyklem publikacji powiązanych tematycznie, oraz oceny jego aktywności naukowej.

Dr inż. Mariusz Kamola jest obecnie zatrudniony w Centrum Badań i Transferu Technologii, NASK Państwowy Instytut Badawczy w Warszawie na etacie adiunkta. Stopień doktora nauk technicznych uzyskał w roku 2004 w specjalności automatyka na Wydziale Elektroniki i Technik Informacyjnych Politechniki Warszawskiej na podstawie rozprawy doktorskiej pt. *Algorithms for Optimisation Problems with Implicit and Feasibility Constraints*.

Przedmiotem niniejszej oceny jest przedłożona mi dokumentacja obejmująca:

- A. Rozprawę habilitacyjną pt. „*Metody analizy i odkrywania struktury sieci społecznościowych i technologicznych*” w postaci 10 powiązanych tematycznie następujących prac, stanowiących podstawę do oceny osiągnięcia naukowego habilitanta:

1. Kamola, M., Arabas, P., Sieci społeczne i technologiczne. Jak rozumieć, jak wykorzystać, Wydawnictwo PWN, ISBN: 978-83-01-19917-3, 2018, (**MNiSW 80 pkt.**)
2. Kamola, M., How to Verify Conway's Law for Open Source Projects, IEEE Access (7), 2019, str. 38469 - 38480 (**MNiSW 100 pkt.**)
3. Kamola, M., Sensitivity of Importance Metrics for Critical Digital Services Graph to Service Operators' Self-Assessment Errors, Security and Communication Networks, DOI: doi.org/10.1155/2019/7510809, 2019 (**MNiSW 40 pkt.**)
4. Kamola, M., Arabas, P., Network Resilience Analysis: Review of Concepts and a Countr-Level Case Study, Computer Science (15.3), str. 311 - 327, DOI: 10.7494/csci.2014.15.3.311, 2014 (**MNiSW 12 pkt.**)
5. Kamola, M., Arabas, P., Improving Time-Series Demand Modeling in Hospitality Business by Analytics of Public Event Datasets, IEEE Access (8), 2020, str. 53666 - 53677 (**MNiSW 100 pkt.**)
6. Kamola, M., Using Rate Equation for Modeling Triad Dynamics on Instagram, Proc. of the Int. Conf. on Digital Information Management (ICDIM), 2016, str. 101 – 106 (**MNiSW 15 pkt.**)
7. Kamola, M., Analysis of User Story Dynamics on a Medium-Size Social News Site, Proc of Int. Conf. on Computational Collective Intelligence, LNAI 12496, str. 97 – 109, 2020 (**MNiSW 20 pkt.**)
8. Laskowska, B., Kamola, M., Grouping compositions based on similarity of music themes, Plos ONE, DOI: 10.1371/journal.pone.0240443, 2020 (**MNiSW 100 pkt.**)
9. Kamola, M., Niewiadomska-Szynkiewicz, B., Piech, B., Reconstruction of a Social Network Graph from Incomplete Call Detail Records. Proc. of Int. Conf. on Computational Aspects of Social Networks (CASoN), 2011, DOI: 10.1371/journal.pone.0240443, (**MNiSW 15 pkt.**)
10. Kamola, M., Protecting Privacy of GPS trails by anonymization of the road graph, ACM SIG-SPATIAL Workshop on Smart Cities and Urban Analytics, 2015, DOI: 10.1145/2835022.2835033, str. 59-62

B. Inne dokumenty, w tym:

1. Autoreferat w języku polskim i angielskim
2. Wykaz informacji dotyczących dorobku naukowego, z uwzględnieniem cyklu powiązanych tematycznie publikacji.

3. Wykaz opublikowanych prac naukowych po uzyskaniu stopnia doktora.
4. Wykaz informacji osobowych oraz informacji o osiągnięciach dydaktycznych, współpracy naukowej i popularyzacji nauki.

Problematyka rozprawy habilitacyjnej i dorobku

Tematyka badawcza rozprawy habilitacyjnej dra inż. Mariusza Kamoli przedstawionej w postaci cyklu prac, stanowiąca podstawę oceny jego osiągnięcia naukowego, obejmuje szereg różnych zagadnień badawczych, które habilitant określił łącznie jako „metody analizy i odkrywania struktury sieci społecznych i technologicznych”.

Podstawowym obiektem analizy są sieci złożone (społecznościowe oraz technologiczne), wykorzystywane w przedstawionych pracach do modelowania różnego typu obiektów i relacji pomiędzy nimi. Obiekty te mogą mieć różny charakter: zbiory modułów programowych i zbiory programistów, zbiory usług, zbiory użytkowników serwisu internetowego, zbiory systemów autonomicznych czy też zbiory motywów muzycznych. Powiązania pomiędzy tymi obiektami mogą reprezentować różnego typu wzajemne relacje: uczestnictwa, współpracy, lubienia, współwystępowania, itp. Generalnie, sieci złożone stanowią medium propagacji informacji, produktów oraz treści cyfrowych. Celem prowadzonych badań było opracowanie nowych algorytmów i metod analizy własności sieci wykorzystywanych do modelowania zjawisk rzeczywistych, grupowania obiektów (wierzchołków sieci), analizy charakteru dynamiki takich sieci (dodawania nowych wierzchołków, nowych połączeń w sieci czy zmiany stanu wierzchołków), opracowanie nowych metod pozyskiwania wiedzy o charakterystyce tych sieci (modelu sieci), opracowanie modeli zachowań i aktywności użytkowników na platformach społecznościowych, oraz rekonstrukcji sieci o zadanych własnościach czy w warunkach niepełnej informacji. Teza badawcza sformułowana przez Autora brzmi: **„prace badawcze, rozwojowe, edukacyjne i popularyzatorskie dotyczące sztuki pozyskiwania danych, łączenia ich w struktury sieciowe, a następnie doboru lub opracowania metod analizy są wartościowe i pożądane zarówno w kontekście naukowym jak i społecznym”**.

Zagadnienia te należą do ogólnie rozumianej problematyki związanej z modelowaniem, analizą i odkrywaniem struktury sieci złożonych i jako takie mogą z pewnością stanowić przedmiot rozprawy habilitacyjnej.

Aktualność i ważność problematyki analizy i odkrywania struktury sieci złożonych oraz zakres badań podjętych przez dra inż. Mariusza Kamolę w cyklu prac składających się na jego rozprawę habilitacyjną, jak również aktualność i ważność tematyki pozostałych jego prac wskazują, że habilitant podjął badania w zakresie aktualnej, ważnej i ciekawej problematyki,

stwarzające szansę na uzyskanie wartościowych wyników odpowiadających wymaganiom stawianym przy przyznawaniu stopnia doktora habilitowanego w naukach technicznych z zakresu informatyki.

Rozprawa habilitacyjna dra inż. Mariusza Kamoli obejmuje 10 prac. W przypadku pięciu z nich habilitant jest jedynym autorem, w pozostałych 5 przypadkach habilitant jest jednym z współautorów. Cztery prace wchodzące w skład rozprawy zostały opublikowane w czasopiśmie posiadającym IF z listy A MNiSzW [H2, H3, H5, H8] (IEEE Access, Plos ONE,), 3 prace [H6, H7, H9] ukazały się w materiałach konferencyjnych (konferencje: ICDIM, ICCCI, CASoN), jedna praca została opublikowana w ramach warsztatów konferencji ACM-SIGSPATIAL, wreszcie 1 publikacja stanowi monografię opublikowaną w wydawnictwie PWN.

Rozprawa habilitacyjna stanowiąca podstawę oceny osiągnięcia naukowego jest poświęcona w znacznej części, jak już wspomniałem, analizie i odkrywaniu struktury sieci złożonych obejmujących sieci społecznościowe i sieci technologiczne. Obejmuje ona zagadnienia: analizy i odkrywania struktury sieci społecznościowych i technologicznych, analizy odporności sieci, metody propagacji informacji w sieciach społecznościowych, pozyskiwania wiedzy o charakterystyce sieci społecznościowej (modelu sieci), analizy dynamiki sposobu rozpowszechniania treści na odbiór i zasięg procesu propagacji informacji w sieci typu „social news”, konstruowania nowych sieci i przekształcania sieci istniejących, oraz rekonstrukcji sieci przy ograniczeniach zasobowych.

Praca [H1] (Kamola, M., Arabas, P., Sieci społeczne i technologiczne. Jak rozumieć, jak wykorzystać, Wydawnictwo PWN, 2018) stanowi monografię przedstawiającą zbiór technik i technologii związanych z pozyskiwaniem, przechowywaniem, przetwarzaniem i prezentacją danych sieciowych. Monografia zawiera prezentacje podstawowych pojęć i definicji dotyczących zarówno struktur i wskaźników opisujących całą sieć złożoną jak i wskaźników opisujących pojedyncze węzły sieci. Monografia ma zarówno charakter popularnonaukowy, praktyczny jak i dydaktyczny. Uzupełniona jest również dużą liczbą praktycznych przykładów własnoręcznie stworzonego oprogramowania. Monografia składa się z dwóch części. Część pierwsza przedstawia podstawowe pojęcia odnoszące się do sieci złożonych oraz zjawiska zachodzące w tych sieciach wraz z modelami matematycznymi opisu tych zjawisk. Część druga ma charakter bardziej praktyczny i prezentuje zbiór dostępnych technologii obróbki i analizy danych, a także omówienie technologii przetwarzania danych nieustrukturyzowanych i integracji danych z różnych heterogenicznych źródeł danych.

Praca [H2] (Kamola, M., How to Verify Conway's Law for Open Source Projects, IEEE Access (7), 2019) podejmuje problem weryfikacji prawa Conway'a w kontekście zbioru 8 projektów open-source'owych na platformie github.com. Podstawowym celem pracy było zaproponowanie metryk umożliwiających przeprowadzenie takiej weryfikacji. Wymagało to zdefiniowania reguł (miar odległości/podobieństwa) grupowania modułów programowych oraz programistów, a także sposobu określenia podobieństwa uzyskanych grup klastrów. Zdaniem Autora (patrz: Autoreferat), podstawowa kontrybucja pracy obejmuje dwa osiągnięcia: zaproponowanie algorytmu wyznaczania optymalnej liczby klastrów w procesie grupowania aglomeracyjnego (zarówno dla modułów programowych jak i programistów) oraz zdefiniowania stopnia podobieństwa strukturalnego dwóch zbiorów klastrów reprezentowanych przez odpowiednie dendrogramy. Wyniki analizy przedstawionej w pracy pokazują, że analizowane projekty nie spełniają prawa Conway'a, co zdaniem Autora wynika z charakterystyki obu sieci (sieci modułów i sieci programistów).

Praca [H3] (Kamola, M., Sensitivity of Importance Metrics for Critical Digital Services Graph to Service Operators' Self-Assessment Errors, Security and Communication Networks, 2019) podejmuje problem oceny wrażliwości wybranych indeksów (kryteriów) istotności (ważności) węzłów na błędy oszacowania wag krawędzi pomiędzy węzłami. Problem ten jest rozpatrywany w kontekście analizy współzależności serwisów (usług), które modelowano w postaci grafu skierowanego o znanej strukturze, ale z krawędziami wag podlegającymi błędom oszacowania. Krawędzie reprezentowały siłę powiązań pomiędzy usługami reprezentowanymi przez wierzchołki grafu. Do oceny istotności wierzchołków wybrano trzy wskaźniki (indeksy): Page Rank (w wersji podstawowej), Reach Centrality oraz Maximum Input. Istotność całego grafu jest reprezentowana przez pojedynczy wskaźnik liczbowy będący agregacją istotności poszczególnych węzłów grafu. Istotność poszczególnych węzłów grafu jest agregowana na kilka sposobów: wartość średnia, mediana i wartość maksymalnej istotności. W celu oszacowania wrażliwości wybranych indeksów istotności węzłów wagi krawędzi są zaburzane zgodnie z rozkładem dyskretnym jednostajnym, przy czym zdefiniowano maksymalne ograniczenie na wartość zaburzenia. Błąd oceny wskaźnika istotności całego grafu jest obliczany jako odchylenie od wartości rzeczywistej bazującej na sieci o niezaburzonych wagach lub, alternatywnie, jako rozbieżność rankingów najistotniejszych węzłów w obu wersjach sieci. Przeprowadzone badania porównywały wrażliwość różnych indeksów istotności w połączeniu z różnymi metodami agregacji na błędy w wagach krawędzi. Uzyskane wyniki pokazują, że szereg z

analizowanych kombinacji indeksów i metod agregacji zapewnia odpowiednią i podobną odporność na zaburzenia wag krawędzi i mogą być używane zamiennie.

Praca [H4] (Kamola, M., Arabas, P., Network Resilience Analysis: Review of Concepts and a Countr-Level Case Study, Computer Science (15.3), str. 311 - 327, 2014) przedstawia wyniki analizy odporności sieci infrastrukturalnej, polskiego fragmentu sieci Internet, na awarie łączy i węzłów. Analizowany model sieci bazuje na grafie połączeń systemów autonomicznych (AS). Spójność grafu jest analizowana w kontekście awarii łącza lub węzła. W pracy zaproponowano szereg wskaźników podatności na awarię względem utraconych połączeń oraz awarii pojedynczych AS. Pozwalają one na ocenę odporności AS na awarie. Wprowadzono również indeksy istotności dla łączy i węzłów (AS). Przedstawiono również lokalizację geograficzną infrastruktury sieciowej, tj. lokalizacją urządzeń dostawców, co pozwoliło na przeprowadzenie analizy odporności sieci w przypadku wystąpienia „łączonych” awarii (join failures) będących wynikiem np. celowych ataków.

W pracy [H5] (Kamola, M., Arabas, P., Improving Time-Series Demand Modeling in Hospitality Business by Analytics of Public Event Datasets, IEEE Access (8), 2020, str. 53666 - 53677) przeanalizowano możliwość poprawy predykcji zajętości pokoi hotelowych polegającej na wykorzystaniu dodatkowej informacji dotyczącej wydarzeń publicznych mogących mieć wpływ na obciążenie hoteli. Najpopularniejszym podejściem w zakresie prognozowania obłożenia pokoi w branży hotelarskiej jest wykorzystanie autoregresyjnych modeli szeregów czasowych. Jednakże, podejście to nie uwzględnia okazjonalnego wpływu wydarzeń publicznych na wzrost popytu. Celem pracy było znalezienie odpowiednich zbiorów danych zawierających informacje o wydarzeniach publicznych i, następnie, wykorzystanie tej informacji do poprawy predykcji obłożenia pokoi hotelowych. Przefiltrowane informacje o wydarzeniach zostały zamodelowane w postaci sieci słów w języku polskim. W pracy przeanalizowano dwa modele z tytułami wydarzeń jako danymi wejściowymi: jeden oparty na analizie kolokacji słów, drugi oparty na „word embeddings” przetwarzanych przez sieć neuronową. Przeprowadzone badania wykazały przewagę pierwszego z podejść. Pozwala ono na poprawę predykcji o 4% w niektóre dni, w których odbywają się istotne wydarzenia. Badania pokazały, i to autorzy uznają za najważniejsze osiągnięcie pracy, związek pomiędzy wagą (ważnością) kolokacji słów w modelu a cechami topologicznymi kolokacji w sieci słów. kolokacje leżące na peryferiach sieci słów, o mniej licznych powiązaniach, statystycznie częściej pojawiają się w tytułach wydarzeń o istotnym wpływie na obłożenie pokoi.

Praca [H6] (Kamola, M., Using Rate Equation for Modeling Triad Dynamics on Instagram, Proc. of the Int. Conf. on Digital Information Management (ICDIM), 2016, str. 101 – 106) jest poświęcona zagadnieniu ewolucji triad w grafie reprezentującym sieć społeczną. Problem analizy triad w tzw. grafach sieci społecznościowych (ang. social network graphs) był intensywnie badanych i posiada bogatą literaturę w kontekście analizy struktury i dynamiki zmian zachodzących w takich grafach, odkrywania społeczności (communities) w ramach grafów oraz ich stabilności. Artykuł jest poświęcony analizie stabilności określonego typu triady w podgrafie sieci Instagram. Obiektem analizy była dynamika domykania się triady nieprzechodniej typu 210 do triady typu przechodniego 300. Dynamika procesu domykania się tej triady była analizowana względem stopni węzłów tej triady oraz podstawowych właściwości strukturalnych sąsiedztwa triady. Autor proponuje modelowanie dynamiki tego procesu za pomocą równania kinetycznego reakcji chemicznej drugiego rzędu. Dodatkową kontrybucją pracy, poza wspomnianym powyżej wynikiem analitycznym, jest algorytm konstrukcji gęstych podgrafów oraz sam zbiór danych dotyczących użytkowników serwisu Instagram.

Praca [H7] (Kamola, M., Analysis of User Story Dynamics on a Medium-Size Social News Site, Proc of Int. Conf. on Computational Collective Intelligence, LNAI 12496, str. 97 – 109, 2020) jest poświęcona zagadnieniu propagacji informacji w sieciach społecznościowych typu „social news”. W pracy przeanalizowano zagadnienie dynamiki popularności wiadomości zamieszczanych w serwisie typu „social news”, gdzie popularność jest modelowana relacją bycia obserwowanym lub polubionym. Miarą popularności wiadomości jest liczba głosów oddanych na wiadomość przez społeczność serwisu. W pracy zaproponowano model propagacji wiadomości (jej popularności) wśród użytkowników tego typu serwisu. Zaproponowany model jest adaptacją klasycznego modelu epidemiologicznego SIR (prace Ross i Ross & Hudson) z dodaną początkową fazą liniową. Jak pisze autor, wykres popularności wiadomości przypomina odpowiedź członu inercyjnego na skok jednostkowy. Zaproponowany model został przetestowany na oryginalnych danych pobranych z serwisu wykop.pl. Jak pokazano, model dobrze dopasowuje się do rzeczywistych danych, tj. dynamiki poszczególnych wiadomości, a sieć powiązań użytkownika zamieszczającego wiadomość nie ma istotnego wpływu na dynamikę popularności wiadomości. Podobnie, nie znaleziono korelacji pomiędzy parametrami modelu a występowaniem słów kluczowych w treści wiadomości. Jak pisze autor, istnieje potrzeba uzupełnienia danych i przeprowadzenie pełniejszych i dokładniejszych danych, aby wykorzystać zaproponowany model do predykcji popularności wiadomości lub wykrywania nieprawdziwych wiadomości.

Praca [H8] (Laskowska, B., Kamola, M., Grouping compositions based on similarity of music themes, Plos ONE, 2020) jest poświęcona analizie i porównywaniu utworów muzycznych opartych na treści utworów oraz podstawowych elementach muzycznych (motywach) odpowiadających wyrazom w języku naturalnym. W pracy zaproponowano metodę grupowania utworów muzycznych wykorzystującą podobieństwo motywów muzycznych. Podstawowa idea podejścia jest porównanie notacji muzycznej do tekstu. W notacji muzycznej temat (motyw) odpowiada słowu. Im więcej podobnych motywów znajdujemy w dwóch utworach muzycznych tym wyższy jest stopień ich podobieństwa. W pracy zaproponowano zdefiniowanie motywu muzycznego w oparciu o cechy fragmentów melodii analogicznie do q-gramów dla fragmentów tekstu. Analizując utwór muzyczny, autorzy wyodrębniają wszystkie q-gramy cech charakterystycznych melodii, a następnie konstruują motywy jako zbiory wybranych q-gramów. Pozwala to na porównanie motywów jak i całych kompozycji. Do obliczenia podobieństwa motywów wykorzystano współczynnik Jaccarda, natomiast podobieństwo dwóch utworów zostało zdefiniowane jako maksimum podobieństwa par motywów pochodzących z każdego utworu. Do grupowania utworów wykorzystano hierarchiczne grupowanie aglomeracyjne. Do weryfikacji zaproponowanego podejścia przeprowadzono szereg eksperymentów grupowania utworów muzyki klasycznej oraz muzyki ludowej. Jak piszą autorzy, główną zaletą tego podejścia, jest łatwość interpretacji uzyskanych wyników i wysoka wartość muzykologiczna. Można stosunkowo łatwo, na podstawie wyników grupowania, określić czynniki decydujące o podobieństwie utworów. Wyniki pracy mogą być również wykorzystane w muzycznych systemach antyplagiatowych oraz systemach rekomendacyjnych.

W pracy [H9] (Kamola, M., Niewiadomska-Szynkiewicz, B., Piech, B., Reconstruction of a Social Network Graph from Incomplete Call Detail Records. Proc. of Int. Conf. on Computational Aspects of Social Networks (CASoN), 2011) analizowany jest problem rekonstrukcji sieci (grafu) powiązań społecznych użytkowników telefonii stacjonarnej na podstawie danych z rejestru rzeczywistych połączeń telefonicznych (CDR – call detail records). Dane z rejestru CDR są niepełne ze względu na szyfrowanie identyfikatorów klientów, stąd poszczególne połączenia można interpretować jako połączenia węzłów w grafie dwudzielnym (affiliation network) z dwoma typami wierzchołków: użytkowników i numerów. W pracy przeanalizowano szereg własności uzyskanej sieci w celu wykazania poprawności rekonstrukcji grafu, jego stabilności i przydatności. Pokazano, że zrekonstruowany graf posiada większość cech sieci złożonej charakteryzującej kontakty społeczne. Analiza dynamiki nowych klientów w tak zrekonstruowanej sieci jest zgodna z

modelem preferencyjnego dołączania dla sieci złożonych. Jak wskazują autorzy, można relatywnie niezawodnie wykorzystywać z tak zrekonstruowanego garfu do modelowania bezpośrednich relacji społecznych pomiędzy użytkownikami sieci.

W pracy [H10] (Kamola, M., Protecting Privacy of GPS trails by anonymization of the road graph, ACM SIG-SPATIAL Workshop on Smart Cities and Urban Analytics, str. 59-62, 2015) jest rozważany problem anonimizacji tras przejazdów samochodów, która pozwalałaby jednakże na wykonywanie wybranych analiz zachowania kierowców. Istotą zaproponowanego podejścia jest idea anonimizacji całego grafu tras przejazdów przy jednoczesnym zapewnieniu dokładnego odwzorowania lokalizacji położenia pojazdu na krawędziach wynikowego grafu. Lokalizacje położenia pojazdów są kodowane ich odległością względem wierzchołków grafu co pozwala na wykonywanie szeregu analiz w grafie pozbawionym szczegółowej informacji geoprzestrzennej odnośnie fizycznej lokalizacji węzłów grafu. W pracy omówiono również krótko potencjalne sposoby deanonimizacji i ewentualne środki zaradcze.

Zbiór prac składających się na rozprawę habilitacyjną Pana dra inż. Mariusza Kamoli, mimo, że dotyczy bardzo różnorodnych zagadnień, posiada jeden wspólny mianownik – podstawowym obiektem analizy są sieci o różnym stopniu złożoności: sieci modułów programowych i programistów, sieci usług, sieci systemów autonomicznych, sieci słów, sieci społecznościowe, sieci motywów muzycznych czy wreszcie sieci tras przejazdów. Stąd, w moim przekonaniu zbiór przedstawionych prac jest tematycznie spójny i jak wspomniałem, generalnie, dotyczy problematyki modelowania i analizy sieci złożonych.

Uzyskane przez habilitanta wyniki w rozprawie habilitacyjnej uważam za wartościowe i ciekawe poznawczo. Za podstawową kontrybucję habilitanta należy uznać, w moim przekonaniu, następujące wyniki: (1) całościowe opracowanie przedstawiające zbiór technik i technologii związanych z pozyskiwaniem, przechowywaniem, przetwarzaniem i prezentacją danych sieciowych uzupełnione dużą liczbą praktycznych przykładów własnoręcznie stworzonego oprogramowania, (2) badania dotyczące wrażliwości indeksów centralności (ważności) węzłów na błędy oszacowania wag krawędzi pomiędzy węzłami w sieci złożonej (w kontekście analizy współzależności serwisów) oraz propozycja nowych indeksów niezawodności, (3) zaproponowanie nowych metod oceny jakości grupowania aglomeracyjnego węzłów sieci oraz zaproponowanie nowej miary podobieństwa strukturalnego dwóch sieci w kontekście grupowania aglomeracyjnego, (4) wyniki analizy odporności sieci infrastrukturalnej na awarie łącz i węzłów oraz zaproponowanie szeregu

wskaźników podatności na awarię względem utraconych połączeń oraz awarii pojedynczych węzłów, (5) dotyczące analizy ewolucji wybranego typu triady w grafie reprezentującym sieć społeczną, (6) dotyczące analizy i porównywania utworów muzycznych opartych na treści utworów oraz podstawowych elementach muzycznych (motywach), (7) dotyczące problemu rekonstrukcji sieci (grafu) powiązań społecznych użytkowników telefonii stacjonarnej na podstawie danych z rejestru rzeczywistych połączeń telefonicznych tak, aby uzyskać sieć o założonych cechach (sieć bezskalowa), oraz (8) dotyczące anonimizacji tras przejazdów samochodów (zachowania prywatności), pozwalające jednakże na wykonywanie wybranych analiz zachowania kierowców.

Opracowane metody i algorytmy, praktycznie we wszystkich przypadkach, jak wynika z przedstawionych prac, zostały zaimplementowane praktycznie i zweryfikowane w dużej mierze na danych rzeczywistych, które zostały zebrane przez habilitanta. Mają one walor nie tylko teoretyczny, ale mogą mieć również istotny walor praktyczny. Należy podkreślić również kwestię analizowanych zbiorów danych. Zebranie czy też pozyskanie szeregu zbiorów danych, niezbędnych do przeprowadzenia analizy, wymagało znacznego wkładu pracy. Analizowane zbiory danych stanowią wartość samą w sobie.

Mam dwie uwagi w odniesieniu do przedstawionego cyklu prac składających się na rozprawę habilitacyjną. Obie uwagi w jakimś stopniu dotyczą tego samego zagadnienia, które pojawia się w kilku pracach, a mianowicie hierarchicznego grupowania aglomeracyjnego (AHC).

Jednym z osiągnięć analitycznych, które jest podkreślane w autoreferacie oraz w pracy [H2], jest stwierdzenie, że autor zaproponował algorytm znajdowania optymalnej liczby klastrów (grup) w procesie grupowania hierarchicznego (dotyczy to zarówno grupowania modułów jak i programistów). Jest to zaskakujące stwierdzenie z dwóch powodów. Po pierwsze, autor stwierdza w pracy [H2], że do grupowania modułów i programistów stosuje standardową procedurę grupowania aglomeracyjnego ale proponuje specyficzną miarę odległości w procesie grupowania klastrów. Wiadomo, że algorytm hierarchicznego grupowania aglomeracyjnego nie jest algorytmem, który globalnie optymalizuje jakąś wybraną funkcję kryterialną. Jest to, generalnie, algorytm, który wykorzystuje różne miary odległości międzyklastrowej (czy też funkcje podobieństwa) lokalnie, w każdym kroku (iteracji) algorytmu do grupowania klastrów. Algorytm podejmuje „dobre” decyzje (lokalne) w odniesieniu do grupowania pary klastrów bazując na ich podobieństwie lub odległości (jakkolwiek zdefiniowanej). Niemniej, jak już decyzja o połączeniu dwóch klastrów zostanie

podjęta w którejś iteracji to nie może być już później cofnięta. Takie podejście powoduje, w ogólności, że lokalne kryterium optymalizacyjne nie jest globalnym kryterium optymalizacyjnym. W tym kontekście mam dwa pytania: (1) jak należy rozumieć optymalną liczbę klastrow, tzn. z punktu widzenia jakiego kryterium globalnego jest to optymalna liczba, oraz jakie jest kryterium globalne oceny wyniku grupowania? W pracy zdefiniowano dwie miary odległości międzyklastrowych J_{TP} i J_{TN} i stwierdzono, że ich konstrukcja powoduje, że wartości obu miar rosną monotonicznie i maleją monotonicznie. Jaki jest wynik grupowania? Analiza przedstawiona w [H2] pokazuje, że optymalne rozwiązanie to jeden klaster zawierający wszystkie obiekty. Czy to prawda?

Drugie pytanie, bardziej o charakterze dyskusyjnym, również odnosi się do grupowania aglomeracyjnego w kontekście grafów dwudzielnych. W kilka pracach pojawia się problem grupowania i praktycznie we wszystkich tych przypadkach zastosowano algorytm AHC. W wielu przypadkach zastosowany algorytm nie wpływa na charakter osiągniętych wyników (np. grupowanie utworów [H8]). Jakie cechy algorytmu AHC zdaniem Habilitanta przemawiają za jego wykorzystaniem w kontekście grafów dwudzielnych? Czy rozważano zastosowanie innych algorytmów grupowania? Klasycznym problemem, który występuje w eksploracji danych oraz systemach rekomendacyjnych jest grupowanie klientów opisanych koszykami ich zakupów. Mamy do czynienia z klientami i produktami. Modelujemy problem w postaci grafu dwudzielnego. Kluczowe znaczenie ma w tym akurat przypadku fakt, że grupujemy obiekty opisane atrybutami nominalnymi. Okazuje się, że lepsze wyniki grupowania aniżeli AHC daje zastosowanie algorytmu ROCK będącego zmodyfikowaną wersją AHC.

Powyższe uwagi (pytania) mają drugorzędny charakter w stosunku do oceny merytorycznej rozprawy habilitacyjnej i nie wpływają na moją pozytywną ocenę rozprawy habilitacyjnej jako oryginalnego osiągnięcia naukowego.

Ocena dorobku naukowego nie objętego rozprawą habilitacyjną, tzn. ocena aktywności naukowej dra inż. Mariusza Kamoli

Prace dra inż. Mariusza Kamoli nie wchodzące w skład habilitacji obejmują szerszą tematykę i zakres niż ta podjęta w rozprawie habilitacyjnej. Dr inż. Mariusz Kamola jest autorem (współautorem) 45 publikacji (w tym 10 prac złożonych jako rozprawa habilitacyjna) po uzyskaniu stopnia doktora. W chwili składania dokumentów przez habilitanta wskaźniki

dotyczące IF, indeksu Hirscha i cytowań były następujące: (1) suma wskaźników IF dla osiągnięcia naukowego wynosi 11,5, (2) łączna suma wskaźników IF z uwzględnieniem pozostałych publikacji wynosi 20,5 (3) łączna suma punktów za osiągnięcie naukowe wynosi 482, (4) łączna suma punktów MNiSzW za wszystkie publikacje po doktoracie wynosi 904 (przed doktoratem - 36), (4) liczba publikacji ujętych w bazie WoS – 20, Scopus – 28 i Google Scholar 43, (5) liczba cytowań wg. WoS – 57 (bez autocytowań), liczba cytowań wg. Google Scholar – 160 (135 bez autocytowań) , (6) indeks Hirscha wg. WoS $H=3$, indeks Hirscha wg. Scopus $H=5$, indeks Hirscha wg. Google Scholar $H = 6$. Przytoczone powyżej dane syntetyczne wskazują, że dorobek dra inż. Mariusza Kamoli jest odbierany pozytywnie przez środowisko naukowe. Zdecydowana większość cytowań to cytowania po roku 2015. W bazie Google Scholar najczęściej cytowanymi pracami dra inż. Mariusza Kamoli są prace: *E Niewiadomska-Szyrkiewicz, A Sikora, P Arabas, M Kamola, M Mincer, J. Kołodziej, Dynamic power management in energy-aware computer networks and data intensive computing systems, Future Generation Computer Systems 37, 284-296, 2014* (65 cytowań wg. GS) oraz *M. Kamola, P. Arabas, Shortest path green routing and the importance of traffic matrix knowledge, 24th Tyrrhenian Int. Workshop on Digital Communications-Green ICT, 2013, IEEE Explore* (14 cytowań wg. GS).

Z 45 prac opublikowanych po doktoracie 9 prac zostało opublikowanych w czasopismach naukowych znajdujących się w bazie JCR, 6 publikacji to rozdziały w monografiach, 16 publikacji to recenzowane artykuły spoza bazy JCR. Pozostałe 12 publikacje to publikacje, które ukazały się w materiałach konferencji lub warsztatów naukowych zarówno międzynarodowych jak i krajowych. Te dane wskazują na aktywną działalność naukową i publikacyjną dra inż. Mariusza Kamoli.

Jak wspomniałem już wcześniej, prace dra inż. Mariusza Kamoli, nie wchodzące w skład ocenianego osiągnięcia naukowego, obejmują szerszą tematykę i zakres niż te ujęte w ramach tego osiągnięcia. Znacząca część tych prac [P1 – P11], co wynika z charakteru instytucji zatrudniającej dra inż. Mariusza Kamolę, dotyczy inżynierii ruchu w sieciach pakietowych i wiąże się z zagadnieniami jakości, ekonomii i energooszczędności usług przesyłu danych. Do tego nurtu zaliczyłbym również prace związane z badaniem i rozwojem rynku domenowego – instytucja zatrudniająca Habilitanta jest operatorem krajowym rejestru DNS. W ramach tej problematyki Habilitant zajmował się analizą zapytań do rejestru WHOIS oraz opracowaniem wyszukiwarki wolnych nazw domen w domenie .pl, a także wykorzystaniem DNS jako rejestru urządzeń inteligentnych w IoT [P16 – P19]. Habilitant zajmował się również zagadnieniem analizy danych dotyczących ruchu pojazdów i

sterowania obiektami przemysłowymi (tj. fragmentem sieci gazowej). To drugie zagadnienie stanowi w pewnej części kontynuację badań prowadzonych wcześniej w ramach pracy doktorskiej. Ponadto, dr inż. Mariusz Kamola posiada w swoim dorobku prace dotyczące: cyberbezpieczeństwa, analizy danych pochodzących z różnych heterogenicznych źródeł danych oraz narzędzi i algorytmów przetwarzania równoległego i rozproszonego [P20-P23].

Podsumowując, dorobek naukowy dra inż. Mariusza Kamoli, stanowiący podstawę do oceny jego aktywności naukowej, jest bardzo różnorodny i dotyczy zarówno problematyki metod analizy i odkrywania struktury sieci społecznościowych i technologicznych (tematyka rozprawy habilitacyjnej), telekomunikacji (inżynieria ruchu w sieciach pakietowych, analiza rynku domenowego), sterowania (sterowanie obiektami przemysłowymi) oraz cyberbezpieczeństwa. Generalnie, dorobek naukowy Pana dra inż. Mariusza Kamoli oceniam pozytywnie. Spełnia on, w moim przekonaniu, kryterium istotnej aktywności naukowej, na co składa się w szczególności wartość merytoryczna i oryginalność prac.

Działalność dra inż. Mariusza Kamoli na rzecz środowiska naukowego i jego pozycja w środowisku naukowym

Dr inż. Mariusz Kamola wykonał szereg recenzji dla międzynarodowych czasopism naukowych, m. in., Int. Jour. Appl. Math. Comput., Foundations of Comp. And Decision Sciences, MDPI Sensors, MDPI Electronics, Journal of Telecommunications and Information Technology. Był również recenzentem prac zgłaszanych na konferencje międzynarodowe.

Ważnym aspektem działalności dra inż. Mariusza Kamoli jest jego udział jako wykonawcy bądź kierownika w realizacji programów międzynarodowych (2 projekty: 7PR i 5PR) i krajowych projektów badawczych finansowanych m. in. przez NCBiR, EFRR, NCN, KBN i MNiSzW. Uczestniczył i uczestniczy nadal w realizacji projektów wykonywanych w NASK.

Dr inż. Mariusz Kamola uczestniczył aktywnie w pracach zespołów eksperckich i konkursowych, m. in. uczestniczył w opracowaniu strategii informatyzacji Urzędu m. Warszawy oraz strategii rozwoju sztucznej inteligencji dla Ministerstwa Cyfryzacji. Po uzyskaniu stopnia doktora intensywnie współpracował i współpracuje z przedsiębiorstwami sektora gospodarczego uczestnicząc w realizacji szeregu projektów jako kierownik lub wykonawca projektu.

Otrzymał szereg nagród, m.in. nagrodę naukową I stopnia za współautorstwo książki *Programowanie równoległe i rozproszone* (wydawnictwo Politechniki Warszawskiej, 2009), nagrodę zespołową III stopnia Rektora PW za osiągnięcia dydaktyczne w roku akademickim 2014/2015 i 2016, oraz nagrodę zespołową III stopnia Rektora PW w dziedzinie informatyki oraz automatyki i robotyki za udział w projekcie 5 Programu Ramowego.

Ocena dorobku dydaktycznego i organizacyjnego

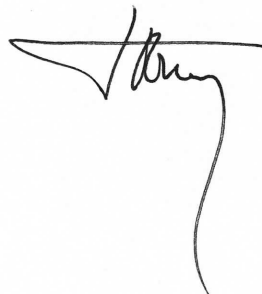
Dr inż. Mariusz Kamola w trakcie swojej pracy zawodowej prowadził i prowadzi nadal zajęcia (wykłady, ćwiczenia, laboratoria, projekty) na I i II stopniu studiów z następujących przedmiotów: Programowanie zdarzeniowe, Systemy operacyjne, Programowanie obiektowe, Programowanie w języku C, Techniki analizy sieci społecznościowych, Sieci i sterowanie systemami, Sterowanie sieciami komputerowymi, Podstawy obliczeń równoległych i rozproszonych, Sieci komputerowe. Prowadzi również zajęcia na studium podyplomowym Zarządzanie Zasobami IT. Jest promotorem ponad 40 prac dyplomowych (18 prac magisterskich i 23 prace inżynierskie).

Działalność organizacyjna dra inż. Mariusza Kamoli obejmuje pełnienie szeregu funkcji administracyjnych, w tym: zastępca kierownika Pracowni Sterowania Siecią w NASK (2010 – 2016), Kierownika zespołu NASK w projektach ECONET i Platforma budowy usług multimedialnych, kierownik techniczny zespołu PW w projekcie QOSIPS.

Wniosek końcowy

Podsumowując, uważam, że zarówno osiągnięcie naukowe dra inż. Mariusza Kamoli, ocenione na podstawie zgłoszonego przez habilitanta zbioru 10 powiązanych tematycznie publikacji „*Metody analizy i odkrywania struktury sieci społecznościowych i technologicznych*”, jak i jego aktywność naukowa, oceniona na podstawie prac nie wchodzących w skład rozprawy habilitacyjnej oraz udziału jako wykonawcy bądź kierownika w realizacji szeregu programów międzynarodowych i krajowych projektów badawczych, spełniają wymogi stawiane przez odpowiednie przepisy o stopniach i tytule naukowym. W związku z powyższym wnoszę o przyjęcie przez Radę Wydziału Elektroniki i Technik Informacyjnych Politechniki Warszawskiej przedłożonej rozprawy habilitacyjnej, w postaci

cyklu powiązanych publikacji, i dopuszczenie dra inż. Mariusza Kamoli do dalszych etapów postępowania habilitacyjnego.

A handwritten signature in black ink, consisting of a stylized 'M' followed by a long, sweeping vertical stroke.