

# Podstawowe programy do analizy sieci

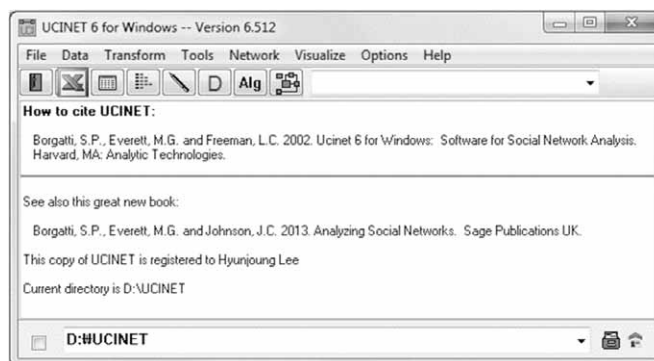
Tradycyjna analiza danych koncentruje się na atrybutach aktora. Ponieważ jednak zachowanie aktora nie zawsze jest niezależne, konieczna jest obserwacja aktorów z perspektywy ich związków z siecią. Jest takie tradycyjne koreańskie powiedzenie, zawarte w księdze *Myungshimbogam*<sup>1</sup>, które w polskiej wersji brzmi: „Pokaż mi swoich przyjaciół, a powiem ci, kim jesteś”. Przytaczamy te słowa, aby zmienić punkt widzenia z atrybutów jednego aktora na sieć. Analizę Big Data z tej perspektywy można przeprowadzać za pomocą dostępnych na rynku programów do analizy sieci. Omawiamy kilka najpopularniejszych z nich, w tym UCINET (*University of California Irvine NETWORK*), NetMiner, R, Gephi i NodeXL. UCINET i NetMiner to kompleksowe programy, w którym można stosować różne techniki analizy sieci. NodeXL oraz R to oprogramowanie do obliczeń statystycznych, natomiast Gephi służy głównie do wizualizacji. Podstawową analizę i wizualizację danych w NodeXL, można wykonać, wprowadzając dane o sieci za pomocą szablonu Excela.

## 1. UCINET

UCINET [1] to napisany przez Lintona Freemana program do różnego rodzaju analizy sieci. UCI to akronim nazwy University of California, Irvine – miejsca pracy Freemana. Program ten został później ulepszony we współpracy ze Steve'em Borgattim oraz Martinem Everettem i jest obecnie najczęściej używanym na rynku oprogramowaniem do analizy sieci. UCINET pozwala badaczowi używać różnych technik analitycznych, takich jak pomiar topograficzny z uwzględnieniem gęstości i klastrowania oraz równoważności, w tym strukturalnych, automorficznych i regularnych cech całej sieci. Ponadto za pomocą odpowiednich funkcji w UCINET można z łatwością przekształcać dane, a także wykonać analizę statystyczną związaną z analizą sieciową. UCINET można pobrać z witryny [www.analytictech.com](http://www.analytictech.com).

Interfejs zaktualizowanego niedawno programu UCINET 6 wygląda podobnie jak na rysunku 1. Po zainstalowaniu programu można za pomocą rozwijanej listy w małym oknie zobaczyć bieżący folder domyślny i otwierać dane z folderów z dokumentami. Pasek menu programu zawiera pozycje *File*, *Data*, *Transform*, *Tools*, *Network*, *Visualize*, *Options* i *Help*. Z poziomu menu *File* można wykonywać działania takie, jak zmiana domyślnego folderu lub edycja plików z danymi. Wybierając polecenie *File > Change Default Folder*, można zmienić skonfigurowany wcześniej folder domyślny, a wybór polecenia *File > Text Editor* otwiera Notatnik (w systemie Windows). W Notatniku użytkownik może z łatwością edytować pliki z danymi.

Menu *Data* zawiera wewnętrzne polecenia programu, umożliwiające otwieranie i zapisywanie plików z danymi. Za pomocą



Rys. 1. Interfejs programu UCINET 6

UCINET można nie tylko importować i analizować dane sieciowe z Excela i programów tekstowych, ale również importować dane z innych programów do analizy sieci, np. DL (*Data Language*), KrackPlot, Mage, Pajek i METIS. Co więcej, analizy można wykonywać również na podstawie danych w formatach ważonej (ang. *edge list*) i prostej (ang. *linked list*) listy relacji, co pozwala obsługiwać duże zbiory danych sieciowych. Format danych DL, kompatybilny z UCINET, opisuje język zbioru danych. Pliki DL można tworzyć w Notatniku i otwierać w programie. Format pliku DL przedstawiono w tabeli 1 (A i B)<sup>2</sup>; dl oznacza, że jest to plik dl, a rozmiar macierzy jest podany za pomocą nr (liczba wierszy) i nc (liczba kolumn). W przypadku macierzy kwadratowej rozmiary nr = 4 i nc = 4 można wyrazić jako n = 4 format = fullmatrix. Nazwy zmiennych są pokazane w etykietach, a dane można wprowadzać w sąsiedniej kolumnie. Wpisując nazwy zmiennych i dane, użytkownicy powinni zwrócić uwagę, aby nie zostawiać między nimi żadnych odstępów.

W przypadku wstawiania i edycji danych w formacie innym niż DL można użyć funkcji *Data > Data editors > Excel Matrix Editor* lub *Data > Data editors > Spreadsheet Editor (obsolete)*. Nazwę węzła można wstawić w pierwszym wierszu i kolumnie pokazanego arkusza UCINET. Polecenie *Data > Export* służy do wyboru i zapisywania danych w pożądanym formacie, a polecenie *Data > Display* pozwala na przeglądanie danych sieciowych w ramach UCINET; wykonując to polecenie, otwieramy pliki *ucinetlog* i możemy uzyskać ogólne informacje na temat wyników analizy i informacje o danych. Zapisane dane UCINET są składowane w systemie podwójnych plików: rzeczywisty plik z danymi ma rozszerzenie *##h*, a plik z informacjami o danych – rozszerzenie *##d*. W analizie danych używana jest zawartość pliku z rozszerzeniem *##h*.

Polecenia do transformacji danych albo skonfigurowania macierzy relacji, np. dychotomizuj (*Dichotomize*), symetryzuj (*Symmetrize*), rekoduj (*Recode*) i odwracaj (*Reverse*), są

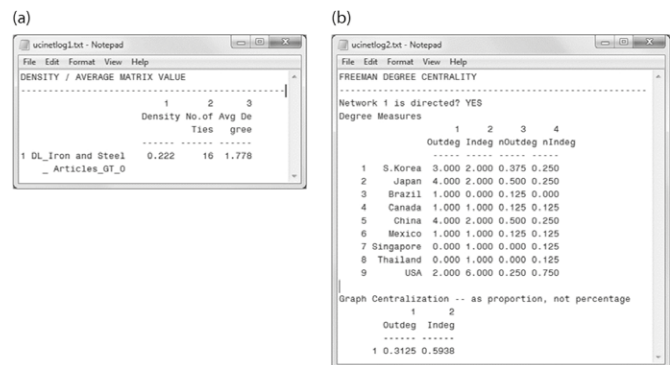
Tabela 1. Format danych DL

A. Format DL_Matrix (macierz) dl nr = 9, nc = 9 etykiety:									B. Format DL_Edgelist (ważona lista relacji) dl nr = 16 nc = 3 format = edgelist2 etykiety osadzone:		
Korea Pd.	Japonia	Brazylia	Kanada	Chiny	Meksyk	Singapur	Tajlandia	USA	dane:		
0	22 823	0	0	14 831	0	0	0	28 725	Chiny	Singapur	12 257
63 107	0	0	0	28 114	0	0	19 641	27 989	Brazylia	USA	10 851
0	0	0	0	0	0	0	0	10 851	Chiny	USA	37 296
0	0	0	0	0	0	0	0	30 789	Meksyk	USA	18 369
36 613	15 227	0	0	0	0	12257	0	37 296	Kanada	USA	30 789
0	0	0	0	0	0	0	0	18 369	Japonia	Chiny	28 114
0	0	0	0	0	0	0	0	0	Japonia	Tajlandia	19 641
0	0	0	0	0	0	0	0	0	Japonia	Korea Pd.	63 107
0	0	0	0	0	0	0	0	0	Korea Pd.	Chiny	14 831
0	0	0	45 999	0	29 782	0	0	0	Japonia	USA	27 989
									Korea Pd.	USA	28 725
									Korea Pd.	Japonia	22 823
									USA	Meksyk	29 782
									USA	Kanada	45 999
									Chiny	Japonia	15 227
									Chiny	Korea Pd.	36 614

dostępne w menu *Transform*. Polecenie *Dichotomize* pozwala zrekonfigurować dane do postaci zer i jedynek, przy czym wartość 1 oznacza, że dane spełniają określony warunek, a 0 oznacza, że warunek nie jest spełniony. Polecenie *Symmetrize* przekształca oryginalne dane, biorąc pod uwagę jedynie połączenia i ignorując kierunek. Warunki dokonywania przekształceń można podać po wybraniu danej opcji w programie UCINET.

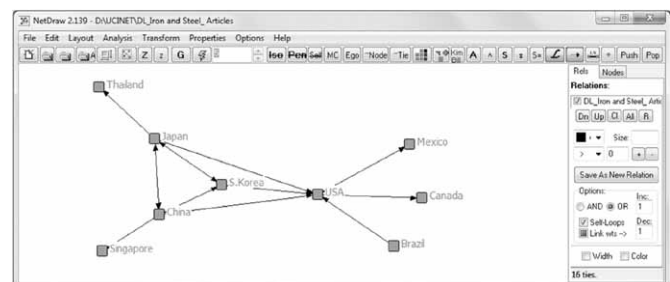
W menu *Tools* znajdują się polecenia związane z analizą danych sieciowych i analizą statystyczną. Wśród analiz dostępnych w menu *Tools* można znaleźć: skalowanie wielowymiarowe, analizę klastrow, analizę korespondencji, rozkład według wartości osobliwej, analizę czynnikową, miary podobieństwa i niepodobieństwa, a także analizę jednoczynnikową. Są również polecenia do tworzenia wykresów punktowych i dendrogramów. Na przykład polecenia *Network > Cohesion > Density > Density Overall* można użyć do ustalenia gęstości i średniego (przeciętnego) stopnia węzła na podstawie danych przedstawionych w tabeli 1 (po ich dichotomizacji). Liczba powiązań (*No. of Ties*) to liczba relacji w sieci, a *Avg Degree* to średni stopień węzła (patrz rys. 2 a). Wskaźnik centralności można uzyskać za pomocą polecenia *Network > Centrality and Power > Degree*, co pokazano na rysunku 2 b. Rezultaty analizy stopnia wychodzącego dla wyników produkcji wskazują, że Japonia i Chiny eksportują wyroby stalowe do czterech krajów, natomiast wyniki analizy stopnia wchodzącego – że Stany Zjednoczone importują z sześciu państw, co daje wskaźnik centralizacji wynoszący 0,5938. Jest to wysoki wskaźnik wśród krajów importujących z punktu widzenia takiej analizy.

Jednym z ograniczeń tego programu jest brak możliwości bezpośredniego przedstawiania graficznych wyników analizy sieciowej. Za pomocą menu *Visualize* można jednak wybrać i użyć zewnętrznych programów do wizualizacji, takich jak NetDraw, Mage albo Pajek. Po wybraniu odpowiedniego



Rys. 2. Wyniki obliczania gęstości i centralności stopnia w programie UCINET – (a) gęstość; (b) centralność stopnia

polecenia z menu *Visualize*<sup>3</sup>, otwiera się niezależne okno, w którym wybieramy dane do wizualizacji. NetDraw to program do wizualizacji, którego autorem jest Steve Borgatti, czyli jeden z twórców UCINET. Program ten zawiera różne algorytmy do dwuwymiarowej wizualizacji sieci i można go uznać za rozszerzenie programu UCINET. Na rysunku 3 pokazano uzyskaną za pomocą NetDraw wizualizację sieci eksporterów produktów stalowych dla danych przedstawionych w tabeli 1.



Rys. 3. Wizualizacja wykonana w NetDraw

Jeden ze wspomnianych programów do wizualizacji, Mage, powstał z myślą o modelowaniu molekularnym [2] i można za jego pomocą tworzyć obrazy 3D<sup>4</sup>. Z kolei Pajek, co po słowiesku oznacza „pająk”, tworzy opis struktury złożonej zaszyfrowanej sieci, wzorowany na swobodnie przecinających się pajęczych niciach. Program ten umożliwia wykonanie prostej analizy, która odszyfrowuje strukturę złożonej sieci. Wizualna interpretacja tworzona przez Pajek jest zoptymalizowana pod kątem kompatybilności z programami takimi, jak Photoshop, co ma taką zaletę, że za pomocą prostych konwersji można tworzyć grafikę o wysokiej rozdzielczości.

W menu *Options* znajdują się polecenia<sup>5</sup> do tworzenia i edycji domyślnych folderów na oprogramowanie połączone z UCINET.

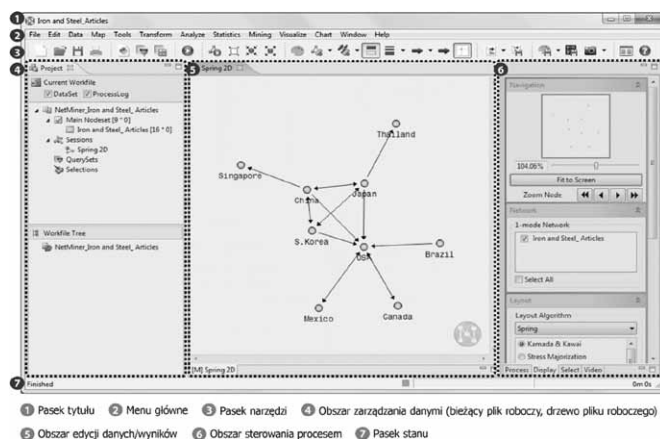
## 2. NetMiner

NetMiner [3] został stworzony przez koreańską spółkę CYRAM. Program ten nie zapisuje oddzielnych plików z analizą, a zarządzanie wynikami analiz odbywa się za pomocą pojedynczego pliku projektu. Wizualna prezentacja wyników analizy na bazie wskaźników analizy jest stosunkowo prosta, co sprawia, że program ten dość dobrze radzi sobie z analizą sieciową. Poza tym program jest dobrze przystosowany do interfejsu systemu Windows i zawiera kilka modułów analitycznych. Konfiguracja programu w wersji NetMiner 4 pozwala na łatwe wykonywanie powtarzających się działań i używanie skryptów pisanych w Pythonie, co zwiększa możliwości osoby wykonującej różne działania analityczne. Program NetMiner można pobrać z witryny <http://netminer.com>.

Na rysunku 4 pokazano środowisko robocze NetMiner 4. Na pasku tytułu (1) można zobaczyć nazwę pliku aktualnie otwartego w programie. Pliki NetMiner mają rozszerzenie .nmf (NetMiner File). Na pasku głównego menu (2) znajdują się polecenia *File, Edit, Data, Map, Tools, Transform, Analize, Statistics, Mining, Visualize, Chart, Window i Help*, które pozwalają wprowadzać i odczytywać dane, zarządzać nimi, a także analizować je i wizualizować. Na pasku narzędzi (3) widoczne są skróty do poleceń.

W menu *File* znajdują się polecenia związane z importowaniem i eksportowaniem danych. Oprócz plików tekstowych i arkuszy Excela, NetMiner potrafi importować i eksportować kilka innych kompatybilnych plików z programów do analizy sieciowej, a importowane dane mogą być w formacie ważonej lub prostej listy relacji oraz macierzy. Dane w formacie ważonej listy relacji składają się z węzła źródłowego, węzła docelowego i wagi, natomiast dane w postaci prostej listy relacji mają format: węzeł źródłowy, węzeł docelowy 1, węzeł docelowy 2, ... itd. Co więcej, pracownik naukowy może wprowadzać nowe dane bezpośrednio, wybierając kolejno polecenia *File > New > Project > Blank Project* (lub *Singleton Project*), ewentualnie *File > New > Workfile*. Polecenie *Blank Project* służy do organizowania niezależnych zbiorów danych, a polecenie *Singleton Project* pozwala bezpośrednio wprowadzać dane do zbiorów węzłów i sieci, bez tytułów wygenerowanych przez program.

Polecenia z menu *Edit* pozwalają edytować dane, a w menu *Data* znajdziemy polecenia służące do tworzenia nowych danych oraz ich importowania i eksportowania. Z poziomu



Rys. 4. Środowisko pracy w programie NetMiner 4

menu *Map* można rozwijać i związać wizualne mapy sieci oraz zmieniać kierunki w obrębie sieci. W menu *Tools* znajdziemy narzędzia do skryptów, które dają możliwość wielokrotnego wykonywania operacji, oraz kalkulatory macierzy i element tworzący zapytania. W menu *Transform* znajdują się polecenia do zmiany kierunku i wartości sieci, wyodrębniania danych węzła i łącza, rozdzielania i łączenia sieci oraz do zmiany kategorii sieci (np. zmiany sieci dwumodalnej na jednomodalną).

W menu *Analize* znajduje się kilka modułów przeznaczonych do analizy sieci. NetMiner, podobnie jak UCINET, umożliwia wykonywanie analizy gęstości i centralności stopnia. Analizę gęstości robimy, wybierając odpowiednie polecenie z menu *Analize > Properties > Network*; jej wyniki zostały pokazane na rysunku 5. Analiza centralności stopnia jest wykonywana za pomocą *Analize > Centrality > Degree*; pojawiają się wtedy dodatkowe zakładki [R] *Main*, [T] *Degree Centrality Vector*, [M] *Spring* i [M] *Concentric*, gdzie [R] oznacza raport, [T] oznacza tabelę, a [M] – mapę. Mapa obejmuje algorytm sprężynowy opracowany przez duet Kamada i Kawai [4], w którym pozycje węzłów są w przybliżeniu proporcjonalne do długości najkrótszej ścieżki i w którym przyjmuje się idealną odległość nawet dla niesąsiadujących par węzłów.

W menu *Statistics* znajdziemy polecenia do różnego rodzaju statystycznych analiz sieciowych. W tabeli 2 przedstawiono polecenia używane w programie NetMiner 4. Polecenia z menu *Mining* pozwalają wykryć wzorce w danych z sieci, a polecenia z menu *Visualize* i *Chart* to różne algorytmy do wizualizacji i tworzenia wykresów, dzięki którym sieć można przedstawić graficznie. Za pomocą menu *Window* można ukryć albo wyświetlić okna podrzędne, a polecenie *Help* otwiera menu pomocy, które zawiera m.in. ogólne informacje o programie NetMiner i aktualizacjach.

Obszar zarządzania danymi (4) składa się z bieżącego pliku roboczego *Current Workfile* i drzewa pliku roboczego *Workfile Tree*. Obszar bieżącego pliku roboczego obejmuje podstawowe jednostki analizy i wizualizacji w NetMinerze, w tym zbiór danych i dziennik przetwarzania. Zbiór danych składa się z głównego zbioru węzłów, podrzędnego zbioru węzłów i danych sieci jednomodalnej oraz dwumodalnej, które służą do analizy i wizualizacji sieci. Węzeł zbiorczy w sieci jednomodalnej to główny zbiór węzłów, a zbiorczy węzeł w sieci



(a) • Output Summary

NETWORK PROPERTIES

	# of Links : O(m)	Density : O(m)	Average Degree : O(m)
Trade relationships	16	0.222	1.778

(b) [R] Main

• Output Summary

DISTRIBUTION OF DEGREE CENTRALITY SCORES

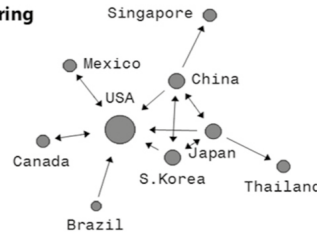
MEASURES	VALUE	
	In-Degree Centrality	Out-Degree Centrality
MEAN	0.222	0.222
STD.DEV.	0.202	0.184
MIN.	0	0
MAX.	0.75	0.5

NETWORK DEGREE CENTRALIZATION INDEX  
59.375% (IN), 31.25% (OUT)

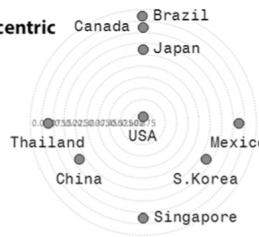
[T] Degree Centrality Vector

		1	2
		In-Degree Centrality	Out-Degree Centrality
1	Japan	0.250000	0.500000
2	S.Korea	0.250000	0.375000
3	USA	0.750000	0.250000
4	China	0.250000	0.500000
5	Brazil	0.000000	0.125000
6	Canada	0.125000	0.125000
7	Mexico	0.125000	0.125000
8	Singapore	0.125000	0.000000
9	Thailand	0.125000	0.000000

[M] Spring



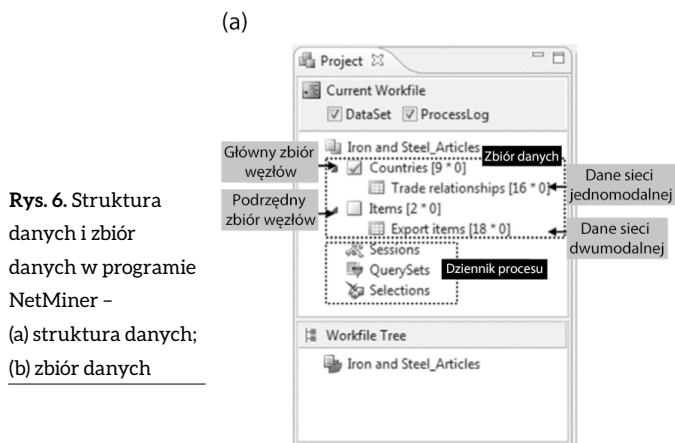
[R] Concentric



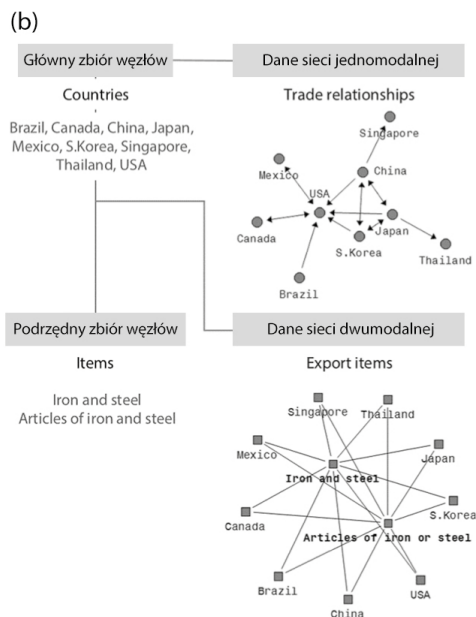
Rys. 5. Wyniki analiz gęstości i centralności stopnia w programie NetMiner - (a) gęstość; (b) centralność stopnia

Tabela 2. Menu i polecenia w programie NetMiner 4

Transform	Analyze	Clique	Role	Statistic	Visualize
Direction	Neighbor	n-Clique	SimRank	MDS	Layout
Symmetrize	Degree	n-Clan	Position	Correspondence	Drawing
Transpose	Ego-network	k-Plex	Blockmodel	Decomposition	Spring
Value	Structural hole	k-Core	Brokerage	Covariance matrix	MDS
Dichotomize	Homophily	Lambda set	Bow-tie model	Principal component	Clustered
Reverse	Assortativity	Community	Expand/collapse	Factor analysis	Layered
Normalize	Equicentrality	Cohesive block	Properties	Frequency	Circular
Recode	Subgraph	s-Clique	Network	Gini coefficient	Simple
Missing	Dyad census	Centrality	Group	Power law	Two mode
Diagonal	Triad census	Degree	Models	Descriptive	Spring
NodeSet	Triad combination	Coreness	Dyadic interaction(p 1)	Crosstabs	Chart
EgoNetwork	Connection	Closeness	ERGM (Exponential Random Graph Model) (p*)	ANOVA	Pie chart
Reorder	Shortest path	Decay	Blockmodel	Correlation	Matrix diagram
LinkSet	All path finding	Betweenness	Two mode	Autocorrelation	Area bar
Incidence	All cycle finding	Flow betweenness	Degree	Regression	Box plot
Line Graph	Dependency	R.W. betweenness	Degree centrality	Logistic regression	Scatter plot
Link reduction	Connectivity	Load	Betweenness centrality	Mining	Contour plot
Link reduction simulation	Minimum cutset	Information	Closeness centrality	Frequent subgraph	Surface plot
Matrix	Maximum flow	Eigenvector	Eigenvector centrality	Classification	Network contour plot
Vectorize	PFNet	Status	Maximum matching	Regression	Network surface plot
Layer	Topological plot	Power		Collaborative filtering	
Split	Influence	Effects		Reduction	
Merge	Accessibility	PageRank		Clustering	
Multiplex	Diffusion	Generalized PageRank		Anomaly detection	
Mode	Influence network	HITS		Text	
2-mode network	Linear threshold	Community			
Onemode network	Cohesion	Equivalence			
Main node attribute	Component	Structural			
Tree construction	Bi-component	Regular			



Rys. 6. Struktura danych i zbiorów danych w programie NetMiner – (a) struktura danych; (b) zbiór danych



dwumodalnej to podrzędny zbiór węzłów. Wyniki analizy są zapisywane w dzienniku przetwarzania. Zmiany w bieżącym pliku roboczym można zapisać jako nowy plik roboczy, a jednocześnie można zarządzać wieloma plikami roboczymi pod postacią listy w drzewie plików roboczych. Możliwość załączenia kilku plików roboczych do jednego projektu ułatwia pracę. Na rysunku 6 przedstawiono strukturę opartą na przykładowym zbiorze danych *Iron and Steel\_Articles* (artykuły z żelaza i ze stali)<sup>6</sup>. Z diagramów z danymi sieci dwumodalnej wynika, że każde z dziewięciu państw eksportuje żelazo i stal oraz wyroby z żelaza i ze stali.

Po wybraniu w bieżącym pliku roboczym jednego ze zbiorów danych aktywny staje się obszar edytowania danych w oknie edycji danych/wyników (5), w którym można bezpośrednio edytować aktywowane dane. Wyniki analizy i wizualizacji są przedstawiane w obszarze danych wynikowych. Edycja, analiza i wizualizacja danych w programie NetMiner odbywa się w jednej domenie.

Opcje analizy i wizualizacji wybiera się w obszarze sterowania procesem (6), a różne opcje stają się aktywne w zależności od wybranych poleceń analizy.

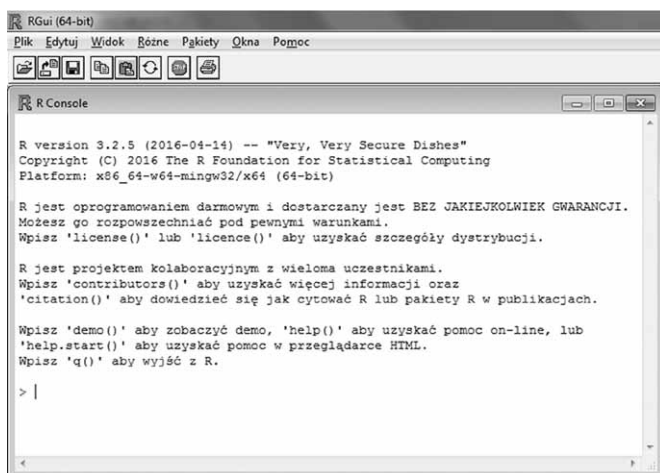
### 3. R

R to program *open source* do statystycznych obliczeń i grafiki, który opiera się na skryptach. Można go pobrać spod adresu [www.r-project.org](http://www.r-project.org). Wykorzystanie R do analizy sieci wymaga instalacji dodatkowych pakietów oprogramowania, takich jak *sna*, *igraph* i *egrm*, które powstały do realizacji konkretnych celów [5]. Obecnie w R można zainstalować 5738 pakietów, co sprawia, że należy on do najbardziej wszechstronnych dostępnych obecnie programów do analizy i wizualizacji. Na rysunku 7 widzimy ekran pojawiający się zaraz po uruchomieniu tego programu<sup>7</sup>.

W skład głównego menu R wchodzi menu *Plik*, *Edytuj*, *Widok*, *Różne*, *Pakiety*, *Okna* i *Pomoc*. Menu *Plik* zawiera polecenia, które pozwalają otwierać nowe okna ze skryptami i/lub zapisanymi skryptami. Okna *R Console*, *R Graphics* i *R Editor*

służą do wykonywania analizy w programie R. Okno *R Console* służy do wprowadzania rzeczywistego skryptu analizy, a okno *R Graphics* – do wykonywania analizy wizualizacji. Ponieważ jednak skryptu nie można edytować w oknie *R Console*, edycja i wykonywanie odbywają się w oknie *R Editor*, a program sugeruje użytkownikowi używanie właśnie tego okna. Każde aktywne okno jest zapisywane w postaci oddzielnego pliku, zaś pliki *R Console* są zapisywane z rozszerzeniem *.RData*, chociaż można je również zapisywać jako pliki tekstowe. Wyniki z aktywnego okna *R Editor* są zapisywane w plikach z rozszerzeniem *.R*.

W menu *Edytuj* znajdują się polecenia, które pozwalają edytować dane i skrypt w oknie *R Console*. Menu *Widok* pozwala wyświetlać albo ukrywać pasek narzędzi i pasek stanu, a menu *Różne* zawiera polecenia związane z wynikami okna konsoli, służące do przerywania (zatrzymywania) bieżących obliczeń, oraz opcje buforowania w trakcie wykonywania. Dzięki poleceniom z menu *Pakiety* można pobierać różne algorytmy analityczne i wykonywać je w programie R. Oprócz podstawowego programu R, istnieje wiele pakietów programistycznych, które



Rys. 7. Interfejs programu R

wymagają wcześniejszego zainstalowania. Pakiety instalujemy, korzystając z polecenia *Pakiety > Zainstaluj pakiet(y)*, które aktywuje serwer CRAN<sup>8</sup>, wybieramy właściwy kraj i postępujemy zgodnie z instrukcjami wyświetlanymi w wyskakującym oknie. W tym przypadku z listy *Pakiety* zainstalowano pakiet do analizy sieciowej o nazwie „igraph”. W menu *Okna* znajdują się polecenia, które pozwalają rozmieszczać okna, a w menu *Pomoc* mamy podręczniki użytkownika i najczęściej zadawane pytania na tematy związane z programem R.

W tabeli 3 przedstawiono wyniki analiz gęstości sieci i centralności stopnia, wykonanych za pomocą pakietu *igraph* w programie R, na danych *Iron and Steel Articles*<sup>9</sup>. Analiza danych w R wymaga napisania skryptu, który pokazano w tabeli 3. Dla większej przejrzystości w tabeli umieszczono również wizualizację sieci, chociaż w programie R znajduje się ona w osobnym oknie, *R Graphics*. Dla tego przykładu wybraliśmy algorytm rozmieszczania opracowany przez Fruchtermana i Reingolda [6]. Jest to algorytm siłowy opierający się na założeniu, że na wierzchołkach działa siła odpychająca, a na krawędziach grawitacja. Graf jest wynikiem obliczeń iteracyjnych. Innymi słowy, zakłada się, że podobnie jak w sprężynie, istniejące łącze między węzłami w odległości mniejszej niż kryterium wykazuje siłę odpychającą, natomiast łącze w odległości większej niż kryterium – siłę przyciągającą. Początkowa pozycja węzła jest ustalana losowo; gdy graf staje się stabilny, algorytm iteracyjnie określa działania pozycjonujące aż do osiągnięcia równowagi.

#### 4. Gephi

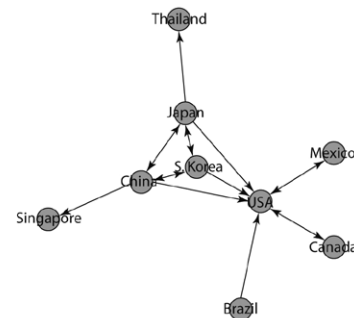
Gephi [7] to program *open source*, który został napisany przez studentów francuskiego University of Technology w Compiègne w roku 2008. Pozwala on na dynamiczną wizualizację, przeszukiwanie i analizę sieci. Dostępnych jest kilka programów do wizualizacji i analizy danych sieciowych, jednak twórcy Gephi skoncentrowali się na atrybutach węzłów i do ich przedstawienia użyli różnych kolorów i rozmiarów, co ułatwia intuicyjną analizę gołym okiem.

Możliwości Gephi nie ograniczają się do analizy i wizualizacji sieci; w programie tym można analizować sieci różnego rodzaju, w tym sieci biologiczne i komunikacyjne, topologię Internetu, sieci udostępniania plików P2P (*peer-to-peer*), oraz sieci społecznościowe, finansowe, semantyczne i organizacyjne. Oprócz tego Gephi obsługuje arkusze kalkulacyjne, pliki w formacie CSV, DL (UCINET), VNA (NetDraw), NET (Pajek), GML, GEXF, TIP, GDF i ZIP, oraz pozwala importować i eksportować kilka typów danych. Przejrzyste przedstawiane wyniki graficzne również można zapisywać w wielu formatach, w tym SVG, PDF i PNG. Gephi można pobrać spod adresu <http://gephi.github.io><sup>10</sup>, a dzięki wtyczkom można z łatwością stosować dodatkowe funkcje. Na rysunku 8 pokazano interfejs Gephi, składający się z menu głównego (1), menu zarządzania (2), obszaru sterowania procesem (3), obszaru wizualizacji (4), obszaru wyników wyjściowych analizy (5) i obszaru roboczego (6).

Znajdujące się w głównym menu (1) (*File, Workspace, View, Tools, Window, Plugins, Help*) menu *File* zawiera polecenia, które pozwalają otwierać istniejące, czyli zapisane pliki Gephi i tworzyć nowe, importować zewnętrzne dane oraz zapisywać

Tabela 3. Skrypt R i jego wyniki

```
> library(igraph)
> #Utwórz graf skierowany używając macierzy przyległości
> Iron_and_Steel_Articles<-
graph(c(5,3,6,3,4,2,4,8,4,1,4,3,2,3,7,3,1,4,1,2,2,1,2,4,1,9,3,6,1,3,3,7),directed=T)
> V(Iron_and_Steel_Articles)$name<- c("Japan", "S.Korea", "USA", "China",
"Brazil", "Canada", "Mexico", "Singapore", "Thailand")
> data
IGRAPH DN-- 9 16 - -
+ attr: name (v/c)
> plot(Iron_and_Steel_Articles, layout=layout_fruchterman_reingold)
```



```
> graph.density(Iron_and_Steel_Articles)
[1] 0.2222222
> degree(Iron_and_Steel_Articles)
Japan S.Korea USA China Brazil Canada Mexico Singapore Thailand
6 5 8 6 1 2 2 1 1
> degree(Iron_and_Steel_Articles, mode=c("in"),loops=FALSE,
normalized=TRUE)
Japan S. Korea USA China Brazil Canada Mexico Singapore Thailand
0.250 0.250 0.750 0.250 0.000 0.125 0.125 0.125 0.125
> degree(Iron_and_Steel_Articles, mode=c("out"),loops=FALSE,
normalized=TRUE)
Japan S.Korea USA China Brazil Canada Mexico Singapore Thailand
0.500 0.375 0.250 0.500 0.125 0.125 0.125 0.000 0.000
```



Rys. 8. Interfejs programu Gephi

bieżące pliki Gephi. Pliki tworzone w programie Gephi są zapisywane z rozszerzeniem *gephi*. Obszar roboczy funkcjonuje podobnie do drzewa z plikami roboczymi w programie NetMiner. Jeśli użytkownik zechce zarządzać nowymi danymi w pliku Gephi, może otworzyć nowy obszar roboczy i zaimportować do niego dane. Dane między obszarami roboczymi można przenosić za pomocą strzałek ◀▶ na obszarze roboczym<sup>11</sup> (6).



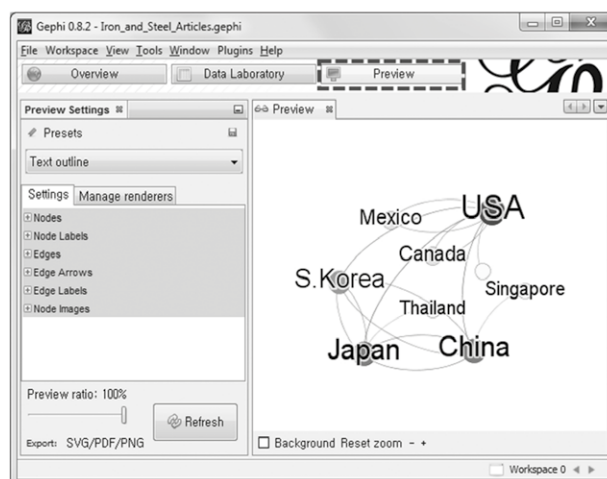
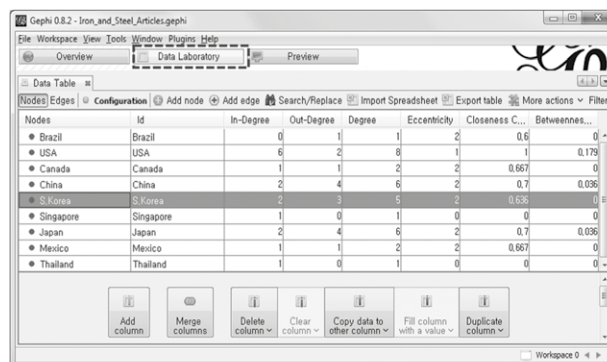
Za pomocą opcji z menu View użytkownik może otworzyć bieżące aktywne okno w trybie pełnoekranowym. Z kolei menu *Tools* zawiera polecenia pozwalające dodawać funkcje z wtyczek oraz opcje i języki konfiguracji oprogramowania. Obecnie Gephi obsługuje języki angielski, francuski, rosyjski, hiszpański, portugalski, chiński i japoński. W menu *Window* znajdują się polecenia do analizy i wizualizacji, takie jak *Clustering*, *Context*, *Data Table*, *Console*, *Filter*, *Graph*, *Hierarchy*, *Layout*, *Partition*, *Preview*, *Preview Setting*, *Ranking*, *Welcome*, *Statistics*, *Timeline*, *Configure Windows* i *Close Windows*<sup>12</sup>. Po wybraniu poleceń pojawiają się nowe karty w obszarze sterowania procesem (3), obszarze wizualizacji (4) i obszarze wyników analizy (5), z poziomu których można wykonywać dodatkowe polecenia. W menu *Plugins* znajdują się polecenia dostępne po zainstalowaniu różnych wtyczek. Z poziomu menu *Help* można rozpocząć aktualizację programu.

W menu zarządzania (2) programu Gephi znajdują się przyciski *Overview*, *Data Laboratory* i *Preview*. Ekran przeglądania (*Overview*) to miejsce, w którym wykonywana jest większość operacji i gdzie przedstawione graficznie wyniki można przesuwać, korygować i edytować. Graficzne wyniki z okna *Overview* można zweryfikować w oknie *Preview* (podgląd), a szczegółowe informacje związane z wynikami analizy znajdziemy na ekranie *Data Laboratory*. Okno *Data Laboratory* to obszar zarządzania wizualną interpretacją danych po wybraniu polecenia *Run*. Poza tym Gephi pozwala badaczowi importować dane w formacie CSV i eksportować tabele danych. Przed wygenerowaniem wyników wizualnych elementy końcowej specyfikacji można zmodyfikować na ekranie *Preview*. Na rysunku 9 pokazano okna *Data Laboratory* i *Preview*.

W obszarze sterowania procesem (3), na karcie *Execution*, która pojawia się po wyborze polecenia *Window*, wyświetlane są opcje polecenia. Obszary sterowania procesem, wizualizacji i wyników analizy pojawiają się na ekranie podglądu w menu zarządzania (2); obszar wyników analizy pojawia się na ekranie *Data Laboratory*, a w dolnej części karty *Data Table* znajdują się polecenia, pozwalające wstawiać, modyfikować i edytować dane.

Obszar wizualizacji (4) obejmuje wyniki wizualizacji danych oraz opcje umożliwiające dostosowywanie wizualizacji. Polecenie *Windows > Ranking* tworzy na obszarze sterowania procesem nową kartę, dzięki której można określić kolor i rozmiar węzłów, krawędzi i etykiet. Gephi pozwala tworzyć wizualizacje dwu- i trzywymiarowe, a użytkownik ma do wyboru wiele różnych algorytmów rozmieszczania, w tym Fruchterman-Reingold, Force Atlas czy Yifan Hu. Wykonując algorytm rozmieszczania, można dodatkowo wybrać polecenie *Layout adjust* (dostosowywanie rozmieszczania), aby na wizualizacji wyników nie nachodziły na siebie etykiety. Niektóre algorytmy wizualizacji wymagają zainstalowania dodatkowych wtyczek.

Liczbę węzłów i krawędzi oraz cechy grafu (skierowany lub nieskierowany) można znaleźć na karcie *Context* w obszarze wyników analizy (5). Na karcie *Statistics* znajduje się trochę wyników analiz wykonywanych przez Gephi. Wyniki analizy można zweryfikować, wybierając polecenie *Run*. Wyniki analizy gęstości można zweryfikować w Gephi za pomocą polecenia *Statistics > Network Overview > Graph Density*, a analizę



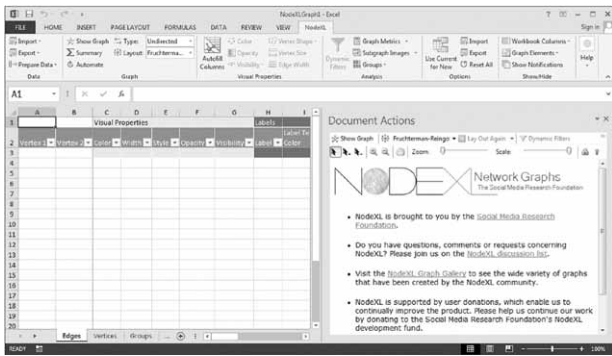
Rys. 9. Ekran laboratoryjny danych i podglądu w programie Gephi

stopnia poszczególnych węzłów weryfikujemy za pomocą poleceń znajdujących się na karcie *Data Laboratory > Data Table*. Aby przeanalizować i wizualnie zinterpretować dane *Iron and Steel\_Articles*, trzeba najpierw zaimportować te dane w formacie CSV, wybierając polecenie *Data Laboratory > Data Table > Import Spreadsheet*, pozwalające otworzyć okno, w którym można wskazać plik do zaimportowania. Trzeba wybrać opcje dotyczące separatorów, prostej i ważonej listy relacji oraz kodowania znaków. Po zaimportowaniu danych można zobaczyć wyniki graficzne na ekranach przeglądania i podglądu (*Overview* i *Preview*). Na rysunkach 8 i 9 pokazano polecenia potwierdzające wyniki analizy, dostępne odpowiednio na kartach *Overview > Statistics* i *Data Laboratory > Data Table*.

Polecenie *Timeline* (*Window > Timeline*) w Gephi umożliwia również analizę sieci dynamicznych, ponieważ pozwala użytkownikowi na dodawanie i usuwanie węzłów oraz krawędzi w funkcji czasu. Co więcej, program Gephi zapisuje bieżące środowisko i po ponownym uruchomieniu wczytuje je, więc dzięki temu użytkownik może utworzyć określone środowisko do analizy i wizualizacji.

## 5. NodeXL

NodeXL [8] to darmowy program *open source* na bazie szablonu Microsoft Office Excel, przeznaczony do zbierania, analizowania i wizualizowania danych sieciowych. NodeXL został napisany przez Social Media Research Foundation i dzięki darowiznom ze środowiska badaczy jest nieustannie rozwijany. NodeXL importuje i eksportuje pliki z macierzami danych



Rys. 10. Interfejs programu NodeXL

w formatach UCINET, GML i Pajek, a zbiera dane sieciowe z witryn takich jak Twitter, YouTube i Flickr oraz z poczty elektronicznej. Poza tym, dzięki wtyczkom, można pobierać dane z osobistych stron Facebooka, fanpage'ów i grup na Facebooku, oraz sieci Mediawiki, Exchange Server, ONA Survey i VOSON Hyperlink. Program NodeXL można pobrać spod adresu <http://nodexl.codeplex.com/>. Jak widać na rysunku 10, po dwukrotnym kliknięciu ikony NodeXL uruchamiany jest program Excel, a menu NodeXL pojawia się jako element wstążki. W programie NodeXL dane są pokazywane w jednym obszarze, a graf widać w wydzielonej części o nazwie *Akcje dokumentu*.

W skład głównego menu NodeXL wchodzi menu *Data*, *Graph*, *Visual Properties*, *Analysis*, *Options* i *Show/Hide*. W menu *Data* znajdują się polecenia do tworzenia, wprowadzania i wysyłania danych, a w menu *Graph* polecenia do wybierania algorytmu rozmieszczania, służącego do generowania grafów w obszarze *Akcje dokumentu*. Menu *Visual Properties* zawiera polecenia pozwalające zdefiniować styl krawędzi i wierzchołków grafu, a polecenie *Autofill Columns* daje użytkownikowi możliwość jednoczesnego zdefiniowania wynikowych wartości analizy oraz rozmiaru, koloru i kształtu krawędzi oraz wierzchołków na podstawie atrybutów krawędzi. Po wybraniu z menu *Analysis* polecenia *Graph Metrics* możemy wykonywać różne analizy sieci, np. obliczyć ogólne metryki grafu, centralność stopnia, pośredniczenia i bliskości, a także zbadać klasteryzację. W menu *Options* znajdują się polecenia umożliwiające import i eksport danych na podstawie kryteriów użytkownika. W menu *Show/Hide* kryją się polecenia pozwalające badaczowi pokazywać i ukrywać tabele danych i grafy.

Gdy używamy NodeXL do analizy i wizualizacji danych z handlu artykułami z żelaza i ze stali, w dolnej części ekranu znajduje się arkusz *Vertices* (wierzchołki) z wynikami analizy dla poszczególnych węzłów. Arkusz *Overall Metrics* zawiera ogólne cechy sieci uzyskane w trakcie analizy. Aby zweryfikować wynikowy graf, użytkownik może w okienku *Akcje dokumentu* odświeżyć graf (*Refresh Graph*) albo za pomocą rozwijanego menu zmienić algorytm rozmieszczania i kliknąć *Layout Again* (rozmieść ponownie).

## Przypisy

- 1 *Myungshimbogam* – pochodząca z XIV w. koreańska księga cytatów i powiedzeń, pełniąca rolę elementarza za czasów dynastii Chosun (przyp. red.).

- 2 Jako przykładu użyto danych dotyczących handlu produktami stalowymi (patrz tabela 1).
- 3 W aktualnej wersji programu UCINET dostępne są opcje: Netdraw, Pajek i Draw (przyp. tłum.).
- 4 Nie ma możliwości bezpośredniego wejścia do programu Mage z poziomu UCINET, można jednak z niego skorzystać, eksportując dane (przyp. tłum.).
- 5 Chodzi tu o Helper Applications (przyp. tłum.).
- 6 Jako przykładu użyto danych dotyczących handlu produktami stalowymi (patrz tabela 5.1).
- 7 Pokazany zrzut przedstawia polską wersję programu.
- 8 R to oprogramowanie tworzone przez wielu pracowników naukowych i programistów, a przeznaczone do niego pakiety analityczne można pobrać z serwera lustrzanego CRAN (*Comprehensive R Archive Network*), który znajduje się na stronie domowej R. Po wybraniu na serwerze CRAN państwa lub regionu można pobrać program R i pakiety w odpowiednim języku.
- 9 Jako przykładu użyto danych dotyczących handlu produktami stalowymi (patrz tabela 1).
- 10 Informacje w książce są oparte na ostatniej wersji programu Gephi, którą w chwili pisania tych słów jest 0.8.2.
- 11 W aktualnie dostępnej wersji 0.9.1 GUI interfejs graficzny wygląda nieco inaczej, więc część opisów programu może być już nieaktualna (przyp. tłum.).
- 12 W najnowszej wersji są to kolejno: Context, Data Table, Filters, Graph, Layout, Output, Preview, Preview Settings, Appearance, Welcome, Statistics, Timeline, Confgure Window, Reset Windows, Close Window, Document Groups (przyp. tłum.).

## Literatura

- [1] BORGATTI S.P., EVERETT M.G., FREEMAN L.C.: *Ucinet 6 for Windows: Software for Social Network Analysis*, Harvard, „Analytic Technologies”, MA 2002.
- [2] RICHARDSON D.C., RICHARDSON J.S.: *The kinemage: a tool for scientific communication*, „Protein Science”, 1(1)/1992.
- [3] Cyram, NetMiner 4, Cyram Inc, Seoul, 2014.
- [4] KAMADA T., KAWAI S.: *An algorithm for drawing general undirected graphs*, „Information Processing Letters” 31(1)/1989.
- [5] BUTTS C.T.: *Network: a package for managing relational data in R*, „Journal of Statistical Software” 24(2)/2008.
- [6] FRUCHTERMAN J.T., REINGOLD E.: *Graph drawing by force-directed placement*, „Software Practice and Experience”, 21/1991.
- [7] BASTIAN M., HEYMANN S., JACOMY M.: *Gephi: an open source software for exploring and manipulating networks*, International AAAI Conference on Weblogs and Social Media, 2009.
- [8] SMITH M., MILIĆ-FRAYLING N., SHNEIDERMAN B. ET AL.: *NodeXL: A Free and Open Network Overview, Discovery and Exploration Add-in for Excel 2007/2010, 2010*, <http://nodexl.codeplex.com/> from the Social Media Research Foundation (dostęp: 22 maja 2015).

Fragment pochodzi z książki: *Big Data w przemyśle*  
Lee Hyunjung, Sohn Il  
Wydawnictwo Naukowe PWN, 2019