

Wykorzystanie teorii grafów w analizie mobilności ludności w Polsce

Sylwia Filas-Przybył¹, Tomasz Klimanek^{1,2}

¹Urząd Statystyczny w Poznaniu

²Uniwersytet Ekonomiczny w Poznaniu

s.filas@stat.gov.pl

tomasz.klimanek@ue.poznan.pl, t.klimanek@stat.gov.pl

1 lipca 2025 r.

Plan prezentacji

Cel

Założenia analizy

Syntetyczna charakterystyka wykorzystanych badań

Podstawowe koncepcje teorii grafów

Pakiet *igraph*

Uzyskane wyniki

Podsumowanie

Cel

Cele prezentacji są trojakiego rodzaju:

- ▶ zaprezentowanie wybranych funkcjonalności pakietu *igraph* w  do przetwarzania macierzy pochodzących z różnych badań statystyki publicznej,
- ▶ omówienie wybranych parametrów strukturalnych grafu w kontekście interpretacji ich wielkości do pomiaru wybranych aspektów mobilności związanej z pracą, nauką bądź zmianą stałego miejsca zamieszkania
- ▶ ocena zmian miar strukturalnych grafów reprezentujących mobilność związaną z pracą w latach 2006 - 2021, mobilność związaną z edukacją w zakresie szkół ponadpodstawowych w latach 2022/23 -2023/24 oraz mobilność związaną ze zmianą miejsca pobytu stałego w ramach migracji wewnętrznych w latach 2002 - 2023.

Kontynuacja prac zaprezentowanych na konferencjach naukowych:

- ▶ **Spójność społeczna, gospodarcza i terytorialna a zrównoważony rozwój. Istota, znaczenie oraz zakres monitorowania** (2017) *Zastosowanie metod teorii grafów w przetwarzaniu, wizualizacji i analizie danych dotyczących dojazdów do pracy w Polsce*
- ▶ **Metodologia Badań Statystycznych** (2023) *Mobilność związana z pracą w ujęciu teorii grafów – analiza wybranych charakterystyk w latach 2006-2016*

Założenia badania

Wykorzystane badania statystyki publicznej:

- ▶ **Cztery edycje badania dojazdów do pracy (dane za): 2006, 2011 (NSP), 2016, 2021 (NSP)**
- ▶ **Dwie edycje badania dojazdów do szkół (dane za rok szkolny): 2022/23, 2023/24**
- ▶ **22 edycje badania migracji wewnętrznych na pobyt stały (dane lata): 2002 - 2023**

Poziom agregacji przestrzennej:

- ▶ dane o mobilności zostały zaprezentowane w ujęciu gmin:
 - ▶ miejskich (1),
 - ▶ wiejskich (2)
 - ▶ miejsko - wiejskich (3)

Założenia badania

Pozostałe uwarunkowania analizy:

- ▶ pominięte zostały przepływy między miastami (4) a obszarami wiejskimi (5) w ramach gmin miejsko - wiejskich,
- ▶ w przypadku migracji wewnętrznych na pobyt stały pominięte zostały uwzględnione w danych przepływy między delegaturami (dzielnicami) miast: Krakowa, Łodzi, Poznania, Warszawy i Wrocławia,
- ▶ w przypadku grafu opisującego dojazdy do szkół wykorzystano tylko przepływy do szkół ponadpodstawowych,
- ▶ w analizach miar strukturalnych grafów uwzględniono dwa podejścia:
 - ▶ oparte na pełnych macierzach przepływów z wykorzystanych badań,
 - ▶ oparte na macierzach przepływów ograniczonych do tych przypadków, gdzie wielkość przepływu wynosiła co najmniej 10 osób (uczniów, pracowników najemnych).

Założenia badania

Wykorzystano wybrane miary strukturalne grafu:

- ▶ liczba wierzchołków,
- ▶ liczba krawędzi,
- ▶ mediana stopnia wierzchołków,
- ▶ gęstość grafu (graf skierowany, bez uwzględnienia wag),
- ▶ średni współczynnik klastrowania (graf nieskierowany, bez uwzględnienia wag),
- ▶ wzajemność (graf skierowany, bez uwzględnienia wag),

Badania dojazdów do pracy

Data publikacji: 31.01.2024

Badanie dojazdów do pracy w 2021 r. dla ogółu pracujących zostało przeprowadzone na podstawie danych zawartych w rejestrach administracyjnych, pozyskanych przez statystykę publiczną na potrzeby Narodowego Spisu Powszechnego Ludności i Mieszkań 2021 (NSP 2021).

Metodyka tego badania została opracowana na podstawie wcześniej przeprowadzonych prac związanych z realizacją zadania „Przepływy ludności związane z zatrudnieniem w 2006 r.” Tym razem dotychczasowe źródło danych dla tego tematu (rejestry podatkowe) zostało wzbogacone o zasoby innych gestorów. Pochodzące z tych źródeł informacje pozwoliły na dokonanie charakterystyki dojeżdżających do pracy według płci i wieku, a w szczególności umożliwiły dokładniejszą identyfikację terytorialną ich miejsca pracy, które w połączeniu z faktycznym miejscem zamieszkania tych osób wyznaczają kierunki przepływów związanych z pracą.



Pliki do pobrania

Archiwum



Dojazdy do pracy w świetle wyników NSP 2021. Publikacja w formacie PDF

151.79 MB



Macierz przepływów ludności związanych z zatrudnieniem w świetle wyników NSP 2021. Tablice w formacie XLSX

2.78 MB

Badania dojazdów do szkół

Dojazdy uczniów do szkół podstawowych i ponadpodstawowych w roku szkolnym 2023/2024



A A A



26.11.2024

W roku szkolnym 2023/2024 do szkół podstawowych i ponadpodstawowych dla dzieci i młodzieży w Polsce dojeżdżało 1 001 155 uczniów. Dojeżdżający stanowili 20,7% ogółu uczniów (6,1% uczniów szkół podstawowych oraz 45,0% uczniów szkół ponadpodstawowych).

Największą liczbą uczniów dojeżdżających do szkół według miejsca zamieszkania charakteryzowały się województwa: mazowieckie (14,0% ogółu dojeżdżających), wielkopolskie (10,8%) i małopolskie (10,2%).

Pliki do pobrania

Archiwum



Dojazdy uczniów do szkół podstawowych i ponadpodstawowych w roku szkolnym 2023/2024. Informacja sygnałna w formacie PDF 1.27 MB



Dojazdy uczniów do szkół podstawowych i ponadpodstawowych w roku szkolnym 2023/2024. Informacja sygnałna w formacie DOCX 6.84 MB



Macierz przepływów związanych z edukacją szkolną 2023/2024. Tablice w formacie XLSX 0.78 MB

Badania migracji na pobyt stały

Baza Demografia

► Wyniki badań bieżących

► Migracje ludności

► Migracje wewnętrzne

Tablica 2g - Migracje wewnętrzne ludności na pobyt stały według gminy poprzedniego i obecnego miejsca zamieszkania

	A	B	C	D	E	F	G	H	I	J	K	L
1	TABL. 2g.	MIGRACJE WEWNĘTRZNE LUDNOŚCI NA POBYT STAŁY WEDŁUG GMINY POPRZEDNIEGO I OBECNEGO MIEJSCA ZAMIESZKANIA.										
2	Rok :	2024	Macierz migracji wewnętrznych (gmina napływu - boczek, gmina odpływu - główka)									
3												
4	Woj.	Pow.	Gmi.	Rodz.	Kod	Gmina poprzedniego miejsca zamieszkania	0201011	0201022	0201032	0201044	0201045	0201052
5						Gmina obecnego miejsca zamieszkania	Bolesławiec	Bolesławiec	Gromadka	Nowogrodz	Nowogrodz	Osiecznica
6	02	01	01	1	0201011	Bolesławiec	0	85	9	15	29	13
7	02	01	02	2	0201022	Bolesławiec	180	0	8	#	14	5
8	02	01	03	2	0201032	Gromadka	3	5	0	#	7	#
9	02	01	04	4	0201044	Nowogrodziec	#	#	0	0	19	4
10	02	01	04	5	0201045	Nowogrodziec	15	15	#	14	0	5
11	02	01	05	2	0201052	Osiecznica	16	#	#	3	4	0
12	02	01	06	2	0201062	Warta Bolesławiecka	41	27	5	0	#	#
13	02	02	01	1	0202011	Bielawa	0	0	0	0	0	0

Definicja grafu

Definicja grafu

Graf

(ang. *graph*) jest strukturą danych składającą się z dwóch zbiorów: zbioru wierzchołków (ang. *vertices*) i zbioru krawędzi (ang. *edges*), co matematycznie zapisujemy w postaci uporządkowanej pary (tzn. takiej, gdzie istotna jest kolejność elementów tworzących tę parę).

$$G = (V, E) \quad (1)$$

$$V = \{v_1, v_2, \dots, v_n\} \quad (2)$$

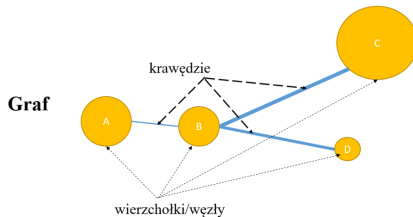
zbiór n ponumerowanych wierzchołków (ang. *vertices*)

$$E = \{e_1, e_2, \dots, e_m\} \quad (3)$$

zbiór m ponumerowanych krawędzi (ang. *edges*).

Każda krawędź jest parą¹ wierzchołków grafu połączonych tą krawędzią:

$$E = \{(u, v) : u, v \in V\} \quad (4)$$



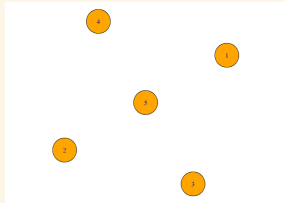
Rodzaje grafów

Graf zerowy (ang. null graph)

graf, który posiada wierzchołki, lecz nie posiada żadnych krawędzi.

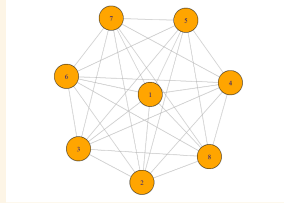
$$V = \{v_1, v_2, v_3, v_4, v_5\}$$

$$E = \{\} = \emptyset$$



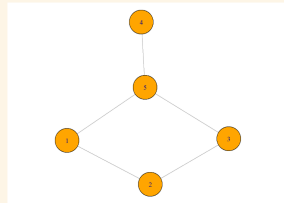
Graf pełny (ang. complete graph)

graf, w którym każde dwa wierzchołki są połączone krawędzią. Graf pełny o n wierzchołkach oznaczamy K_n .



Graf prosty

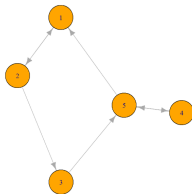
graf nie posiadający pętli² oraz krawędzi wielokrotnych (ang. simple graph lub strict graph).



Rodzaje grafów - cont.

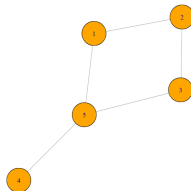
graf skierowany/digraf (ang. directed graph/digraph)

graf zawierający krawędzie skierowane¹. Na rysunku krawędzie skierowane oznaczane są strzałkami.



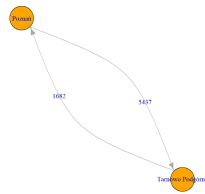
Graf nieskierowany (ang. undirected graph)

graf nie posiadający krawędzi skierowanych.



Graf ważony (ang. weighted graph)

graf posiadający dodatkowe wartości związane z krawędziami. Wartości te nazywamy wagami. Są to np. odległości z jednego miasta do drugiego, czas przejazdu lub jego koszt.

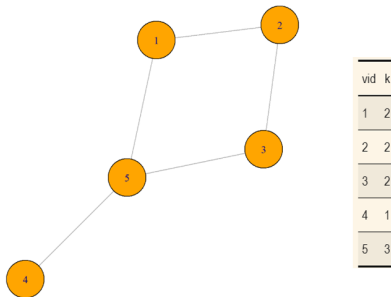


Wybrane miary strukturalne grafu

Rząd grafu (ang. graph order) - liczba wierzchołków w grafie.

Rozmiar grafu (ang. graph size) - liczba krawędzi w grafie.

Rozkład stopni wierzchołków (ang. degree distribution) - rozkład liczby krawędzi incydentnych



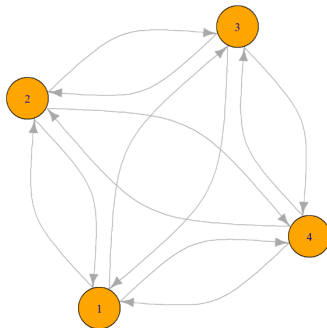
Wybrane miary strukturalne grafu - cont.

Gęstość grafu (*ang. graph density*)

Udział liczby krawędzi grafu do liczby wszystkich możliwych krawędzi grafu (graf pełny).

W przypadku grafu skierowanego liczba wszystkich możliwych krawędzi wynosi $|V| * (|V| - 1)$, a w przypadku grafu nieskierowanego $|V| * (|V| - 1)/2$.

$$D = \frac{|E|}{|V| * (|V| - 1)}$$

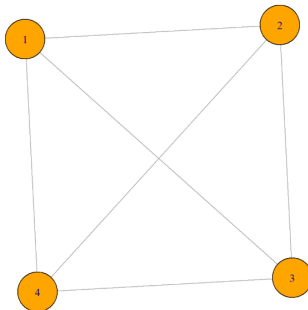


Wybrane miary strukturalne grafu - cont.

Współczynnik klastrowania (*ang. clustering coefficient*)⁵

stopień w jakim wierzchołki w grafie mają tendencję do łączenia się w klastry. Wyznacza się go jako udział trójkątów (połączonych krawędziami trójków wierzchołków) do liczby wszystkich trójków wierzchołków (połączonych i niepołączonych).

$$r = \frac{1}{|V|} \sum_{i=1}^{|V|} \frac{|e_{ij} : v_j, v_k \in N_i, e_{jk} \in E|}{|k_i * (k_i - 1)|} \quad (6)$$

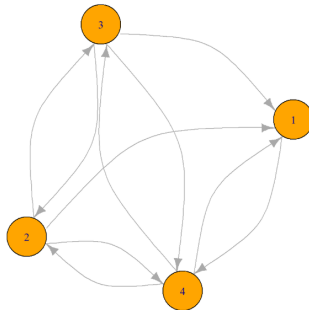


Wybrane miary strukturalne grafu - cont.

Wzajemność (ang. *reciprocity*)

stosunek liczby krawędzi skierowanych w obie strony do liczby wszystkich krawędzi skierowanych w grafie skierowanym.

$$r = \frac{|e_{ij} : e_{ij}, e_{ji} \in E|}{|E|} \quad (7)$$



Macierz przepływów - jako punkt wyjścia analizy grafowej

Numer gminy miejsca zamieszkania (i)	Numer gminy celu mobilności (j)						W_i
	1	2	3	4	...	N	
1	-	a_{12}	a_{13}	a_{14}	...	a_{1n}	W_1
2	a_{21}	-	a_{23}	a_{24}	...	a_{2n}	W_2
3	a_{31}	a_{32}	-	a_{34}	...	a_{3n}	W_3
4	a_{41}	a_{42}	a_{43}	-	...	a_{4n}	W_4
...	...	...	...	...	...	...	...
n	a_{n1}	a_{n2}	a_{n3}	a_{n4}	...	-	W_n
P_j	P_1	P_2	P_3	P_4	...	P_n	$\sum_{j=1}^n P_j = \sum_{i=1}^n W_i$

Gminy są *wierzchołkami grafu* a niezerowe przepływy między gminami to *krawędzie grafu*

Gmina miejsca zamieszkania może to być:

- ▶ gmina poprzedniego miejsca zamieszkania na pobyt stały przed migracją (migracje na pobyt stały),
- ▶ gmina zamieszkania ucznia (dojazdy do szkół),
- ▶ gmina zamieszkania pracownika najemnego (dojazdy do pracy).

Gmina celu mobilności może to być:

- ▶ gmina obecnego miejsca zamieszkania na pobyt stały po migracji (migracje na pobyt stały),
- ▶ gmina lokalizacji szkoły (dojazdy do szkół),
- ▶ gmina lokalizacji miejsca pracy (dojazdy do pracy).

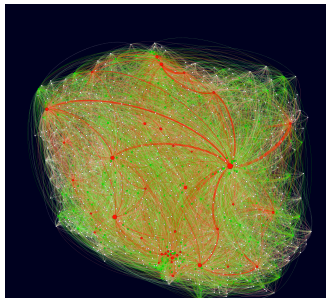
Pakiet *igraph*

Network Analysis and Visualization (Analiza sieci i wizualizacja)

Wersja 2.1.4 – 23 stycznia 2025 r., author Gabor Csardi i inni, <https://r.igraph.org/>

Syntetyczny opis:

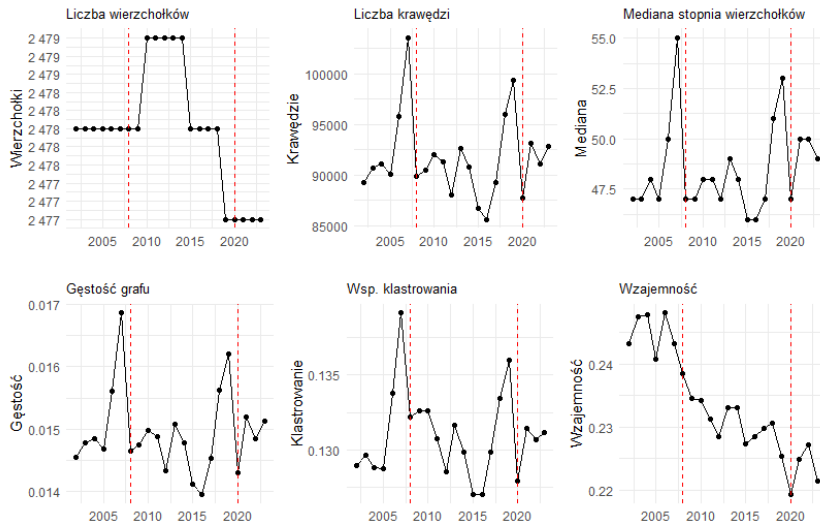
- ▶ zestaw procedur dla przetwarzania grafów i analizy sieci;
- ▶ przetwarza duże grafy;
- ▶ zawiera m.in funkcje do generowania grafów losowych i regularnych, wizualizacji grafów, przetwarzania elementów grafu oraz wyznaczania szeregu statystyk charakteryzujących miary strukturalne elementów grafu, jak i grafu jako całości.



Miary strukturalne grafu migracji wewnętrznych na pobyt stały

Rok	Liczba wierzchołków	Liczba krawędzi	Mediana stopnia	Gęstość grafu	Wsp. klastrowania	Wzajemność
2002	2478	89243	47	0.0145394	0.1289850	0.2432677
2003	2478	90715	47	0.0147792	0.1296648	0.2475418
2004	2478	91099	48	0.0148418	0.1288496	0.2477777
2005	2478	90107	47	0.0146802	0.1287840	0.2407160
2006	2478	95828	50	0.0156122	0.1338225	0.2480529
2007	2478	103505	55	0.0168630	0.1391365	0.2432734
2008	2478	89901	47	0.0146466	0.1322321	0.2384252
2009	2478	90500	47	0.0147442	0.1326486	0.2345511
2010	2479	92027	48	0.0149809	0.1326008	0.2342015
2011	2479	91353	48	0.0148712	0.1307407	0.2312555
2012	2479	88039	47	0.0143317	0.1285918	0.2284626
2013	2479	92587	49	0.0150720	0.1316366	0.2330958
2014	2479	90811	48	0.0147829	0.1298788	0.2330577
2015	2478	86696	46	0.0141245	0.1270686	0.2272586
2016	2478	85641	46	0.0139526	0.1270588	0.2284268
2017	2478	89225	47	0.0145365	0.1298383	0.2297737
2018	2478	95960	51	0.0156337	0.1334823	0.2306035
2019	2477	99377	53	0.0162035	0.1359859	0.2253184
2020	2477	87736	47	0.0143054	0.1279684	0.2193176
2021	2477	93179	50	0.0151929	0.1314546	0.2249274
2022	2477	91099	50	0.0148538	0.1306776	0.2271539
2023	2477	92789	49	0.0151293	0.1312034	0.2213579

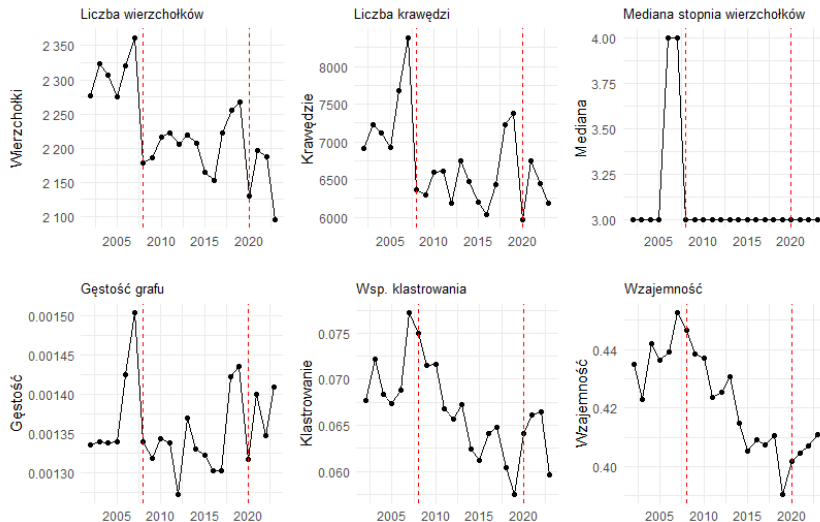
Miary strukturalne grafu migracji wewnętrznych na pobyt stały



Miary strukturalne grafu migracji wewnętrznych na pobyt stały (10 osób i więcej)

Rok	Liczba wierzchołków	Liczba krawędzi	Mediana stopnia	Gęstość grafu	Wsp. klastrowania	Wzajemność
2002	2276	6919	3	0.0013363	0.0676466	0.4348818
2003	2323	7230	3	0.0013404	0.0721803	0.4229482
2004	2307	7124	3	0.0013391	0.0683141	0.4421053
2005	2275	6929	3	0.0013394	0.0673289	0.4363599
2006	2321	7673	4	0.0014250	0.0688405	0.4390473
2007	2361	8375	4	0.0015031	0.0771818	0.4524801
2008	2179	6360	3	0.0013401	0.0749972	0.4467698
2009	2186	6299	3	0.0013188	0.0714719	0.4384563
2010	2217	6602	3	0.0013438	0.0715978	0.4370919
2011	2222	6606	3	0.0013386	0.0668349	0.4237069
2012	2206	6193	3	0.0012732	0.0656381	0.4253165
2013	2220	6751	3	0.0013704	0.0672578	0.4305997
2014	2207	6480	3	0.0013310	0.0624355	0.4148472
2015	2165	6199	3	0.0013231	0.0612154	0.4053503
2016	2153	6039	3	0.0013034	0.0640756	0.4090061
2017	2223	6437	3	0.0013032	0.0647655	0.4073021
2018	2256	7234	3	0.0014220	0.0603987	0.4106864
2019	2268	7379	3	0.0014352	0.0574914	0.3904277
2020	2130	5973	3	0.0013172	0.0641042	0.4017836
2021	2196	6748	3	0.0013999	0.0661103	0.4046628
2022	2187	6446	3	0.0013483	0.0664566	0.4071164
2023	2096	6191	3	0.0014099	0.0596286	0.4108933

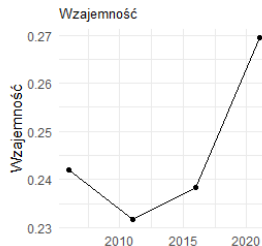
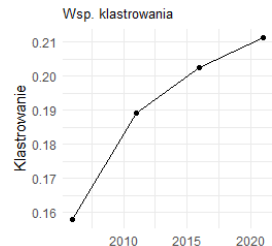
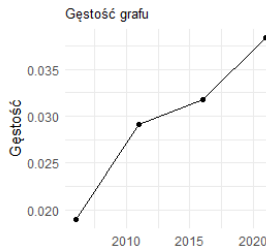
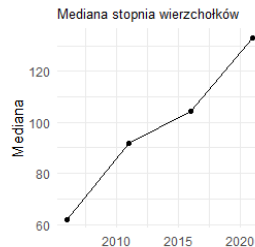
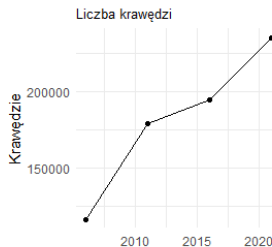
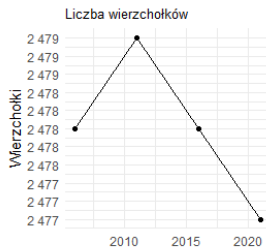
Miary strukturalne grafu migracji wewnętrznych na pobyt stały (10 osób i więcej)



Miary strukturalne grafu dojazdów do pracy

Rok	Liczba wierzchołków	Liczba krawędzi	Mediana stopnia	Gęstość grafu	Wsp. klastrowania	Wzajemność
2006	2478	116567	62.0	0.0189910	0.1579928	0.2420965
2011	2479	179044	92.0	0.0291462	0.1890527	0.2317218
2016	2478	194794	104.5	0.0317357	0.2026692	0.2384038
2021	2477	235089	133.0	0.0383315	0.2112106	0.2694090

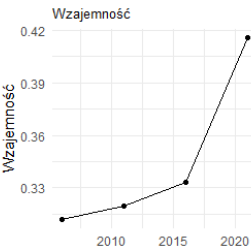
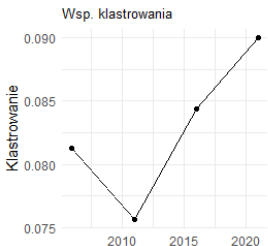
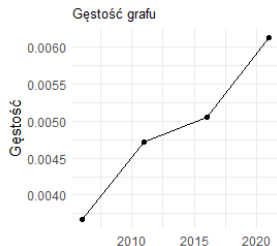
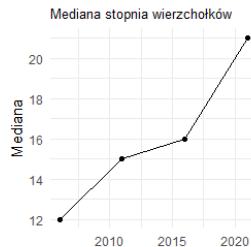
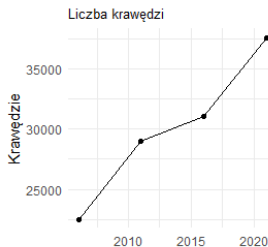
Miary strukturalne grafu dojazdów do pracy



Miary strukturalne grafu dojazdów do pracy (10 osób i więcej)

Rok	Liczba wierzchołków	Liczba krawędzi	Mediana stopnia	Gęstość grafu	Wsp. klastrowania	Wzajemność
2006	2476	22512	12	0.0036736	0.0812803	0.3119617
2011	2478	28958	15	0.0047178	0.0756530	0.3194528
2016	2478	31040	16	0.0050570	0.0843377	0.3329245
2021	2477	37537	21	0.0061204	0.0899704	0.4156432

Miary strukturalne grafu dojazdów do pracy (10 osób i więcej)



Miary strukturalne grafu dojazdów do szkół ponadpodstawowych

Rok	Liczba wierzchołków	Liczba krawędzi	Mediana stopnia	Gęstość grafu	Wsp. klastrowania	Wzajemność
2022	2477	20419	9	0.0033293	0.1314571	0.1197083
2023	2477	20947	9	0.0034154	0.1381245	0.1224413

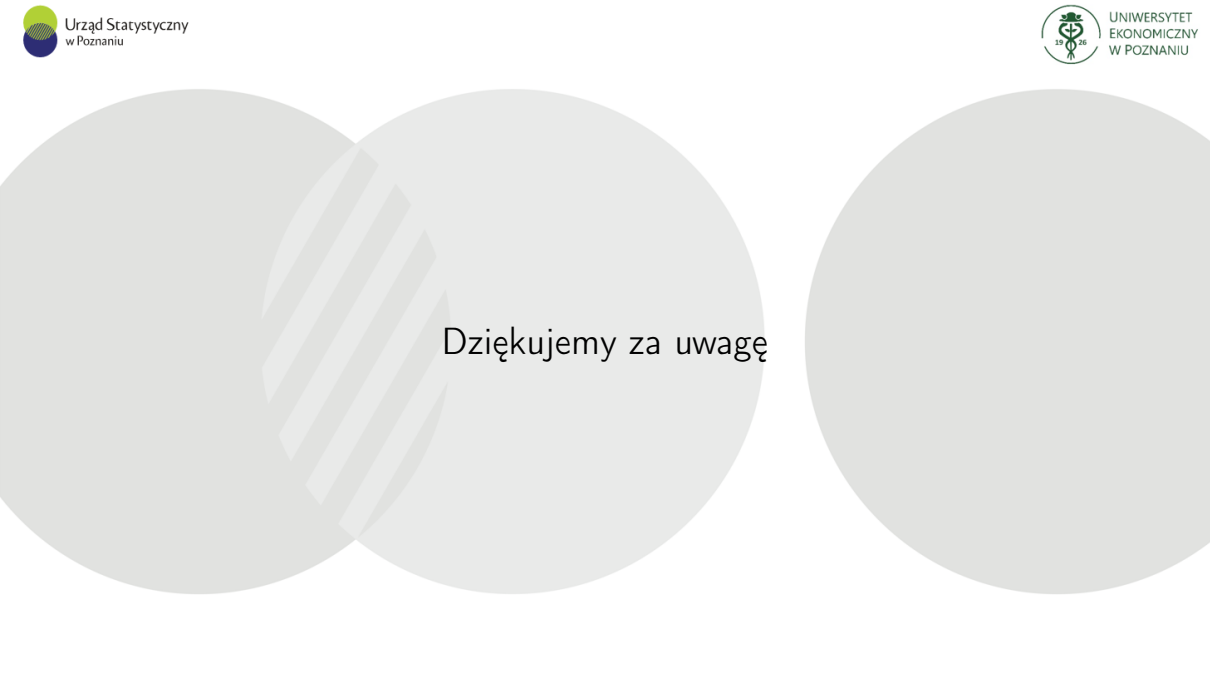
Rok	Liczba wierzchołków	Liczba krawędzi	Mediana stopnia	Gęstość grafu	Wsp. klastrowania	Wzajemność
2022	2477	6466	3	0.0010543	0.1869253	0.0963038
2023	2476	6647	3	0.0010847	0.1922863	0.1001324

Tabela: Dla dojeżdżających 10 i więcej uczniów

Podsumowanie

Wnioski (kierunki dalszych prac):

- ▶ Statystyki stosowane do opisu grafu, można z powodzeniem wykorzystywać do syntetycznego opisu i wnioskowania o złożonych sieciach powiązań, takich jak np. dojazdy do pracy, dojazdy do szkół, migracje na pobyt stały;
- ▶ Potwierdzają one zwiększanie się mobilności Polaków, w tym sensie, że następuje gęstnienie sieci przepływów między miejscem zamieszkania a miejscem celem mobilności
- ▶ zwiększa się udział przepływów w obydwu kierunkach (wzajemność) w przypadku dojazdów do pracy, jednakże w migracjach na pobyt stały udział przepływów w obydwu kierunkach w latach 2008-2020 malał. Dopiero w ostatnich latach widać pewne odbicie tej proporcji;
- ▶ Zastosowane podejście teorii grafów wydaje się być ciekawym kierunkiem poszukiwań badawczych (będzie ono kontynuowane i rozwijane w obszarze innych badań z większym naciskiem na uwzględnienie wag krawędzi i wierzchołków grafu).



Dziękujemy za uwagę