



Praca domowa 3

Jakub Kubacki, Antonina
Ślubowska

Cel pracy

Analiza zbiorów danych dotyczących forów internetowych pochodzących ze strony <https://archive.org/details/stackexchange>

Realizacja

Analiza danych dotyczących 20 forów obejmująca badanie różnorodności społeczności użytkowników i ich aktywności

gaming	bicycles	photo	cooking
chess	coffee	travel	bitcoin
bioinformatics	fitness	vegetarianism	astronomy
health	lifehacks	pets	literature
beer	music	parenting	sports

Różnorodność społeczności forów

Użytkownik aktywny – użytkownik o zweryfikowanym Id

Użytkownik izolowany – aktywny użytkownik, który nie wszedł w interakcję z żadnym innym członkiem forum

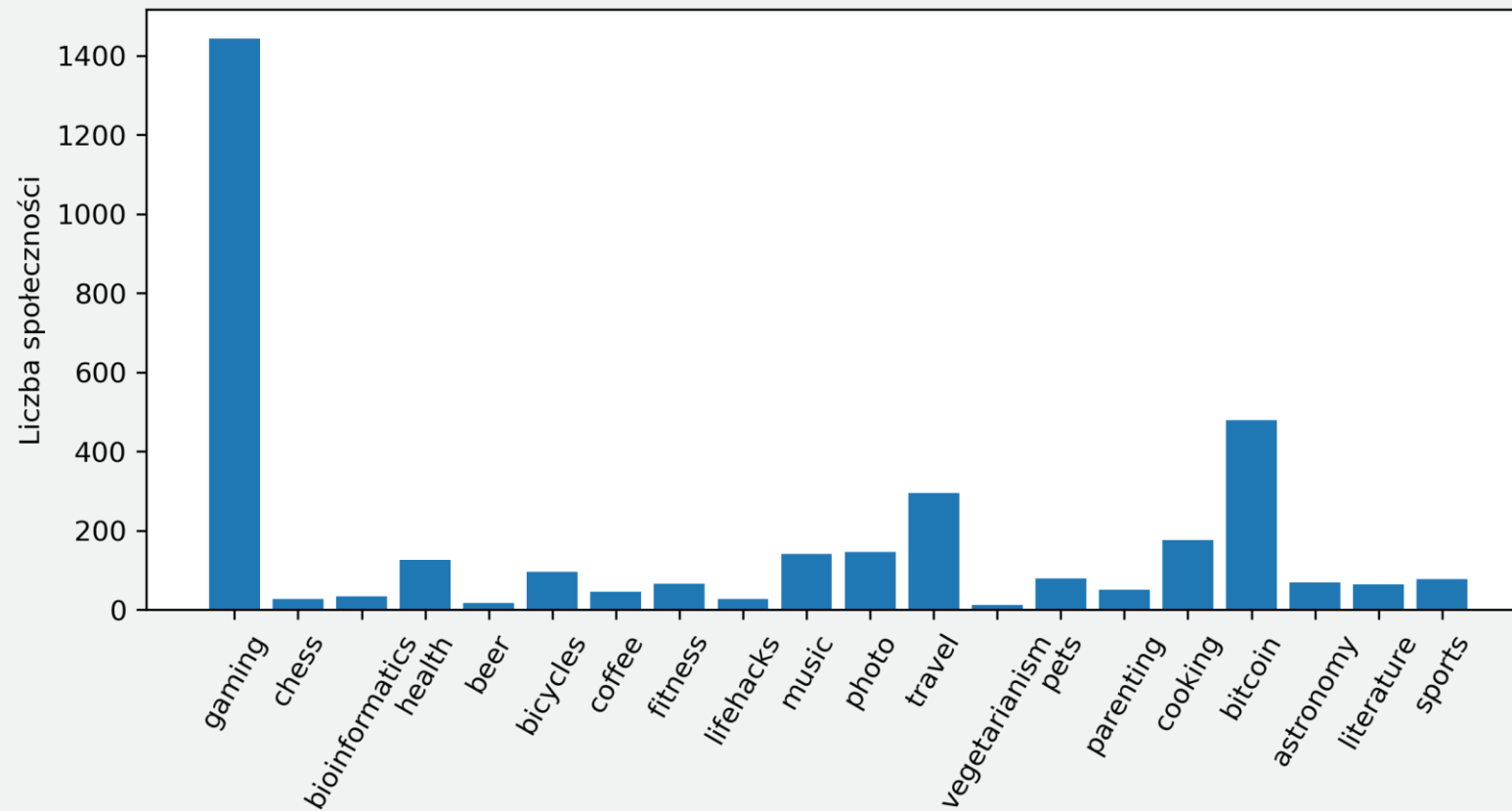
Interakcja – uzyskanie odpowiedzi na zadane przez siebie pytanie lub udzielenie odpowiedzi na pytanie innego użytkownika

Struktura zbiorów rozłącznych

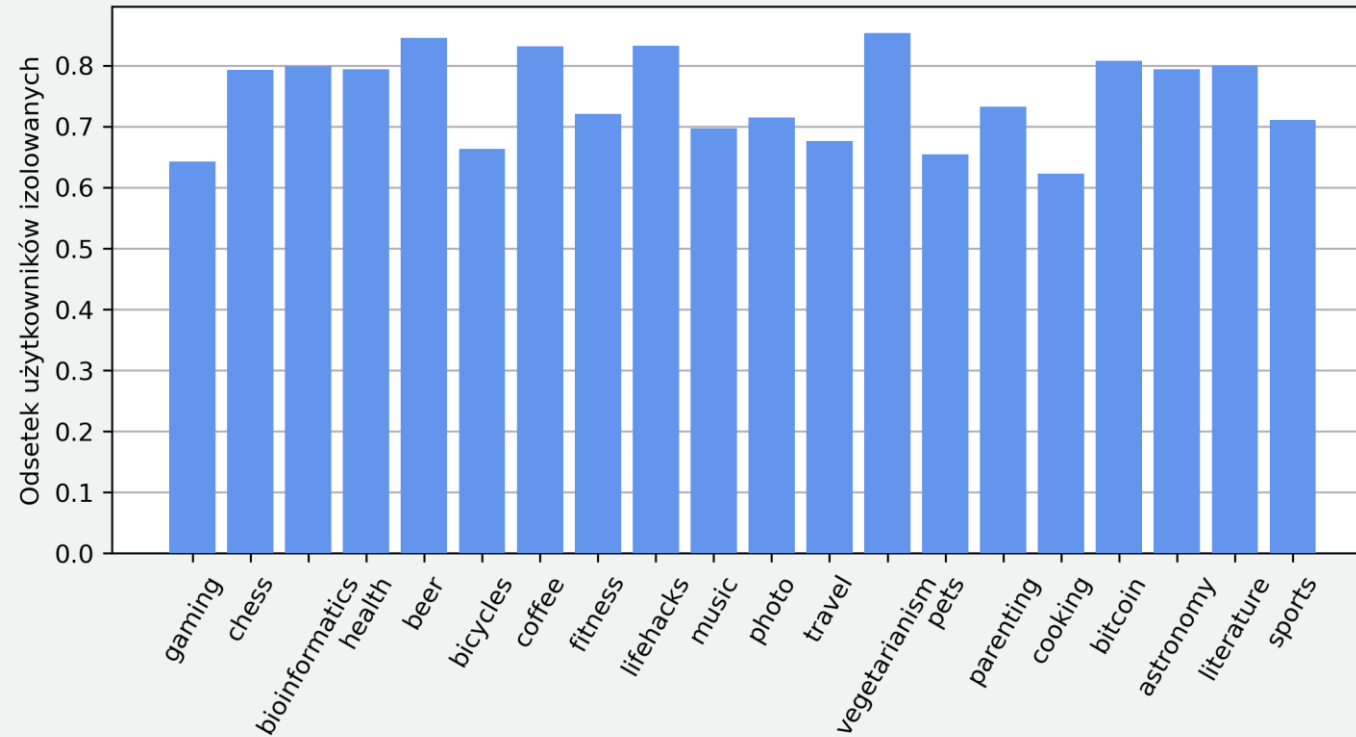
Tworzymy graf nieskierowany, którego wierzchołki symbolizują użytkowników forum, a krawędzie interakcje między nimi.

Społeczność – maksymalny spójny podgraf ww. grafu składający się z min. 2 użytkowników

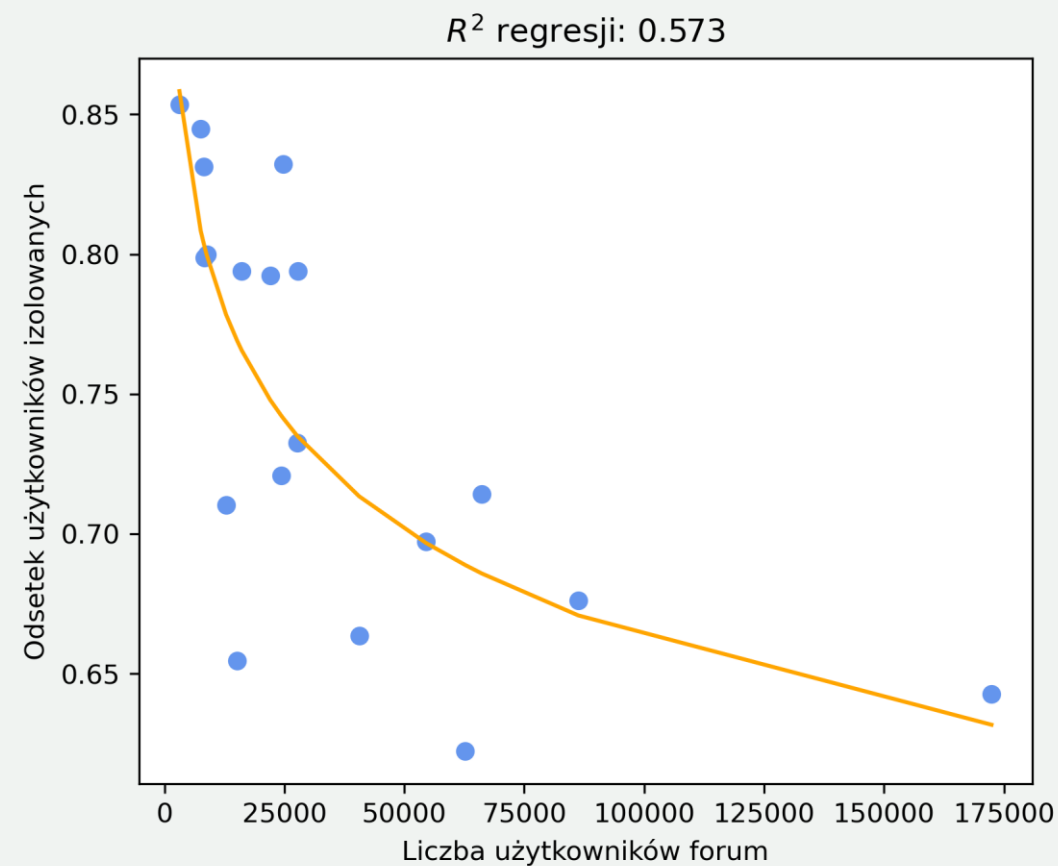
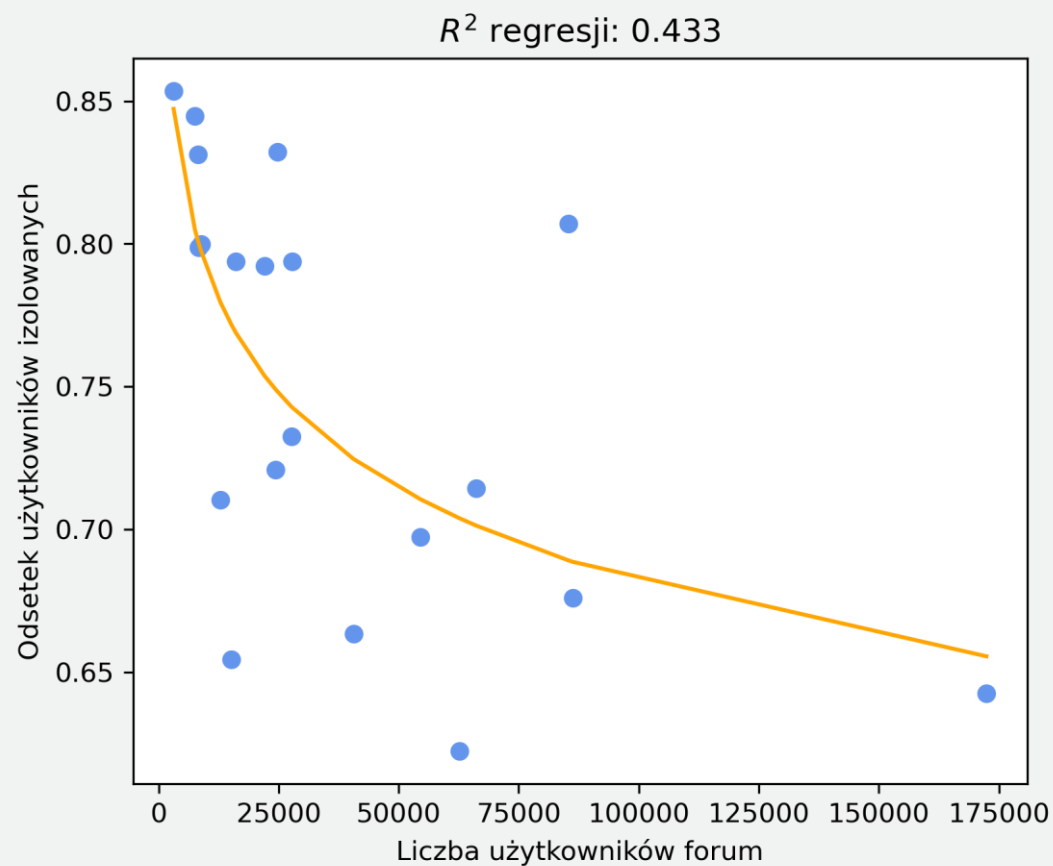
Liczba społeczności



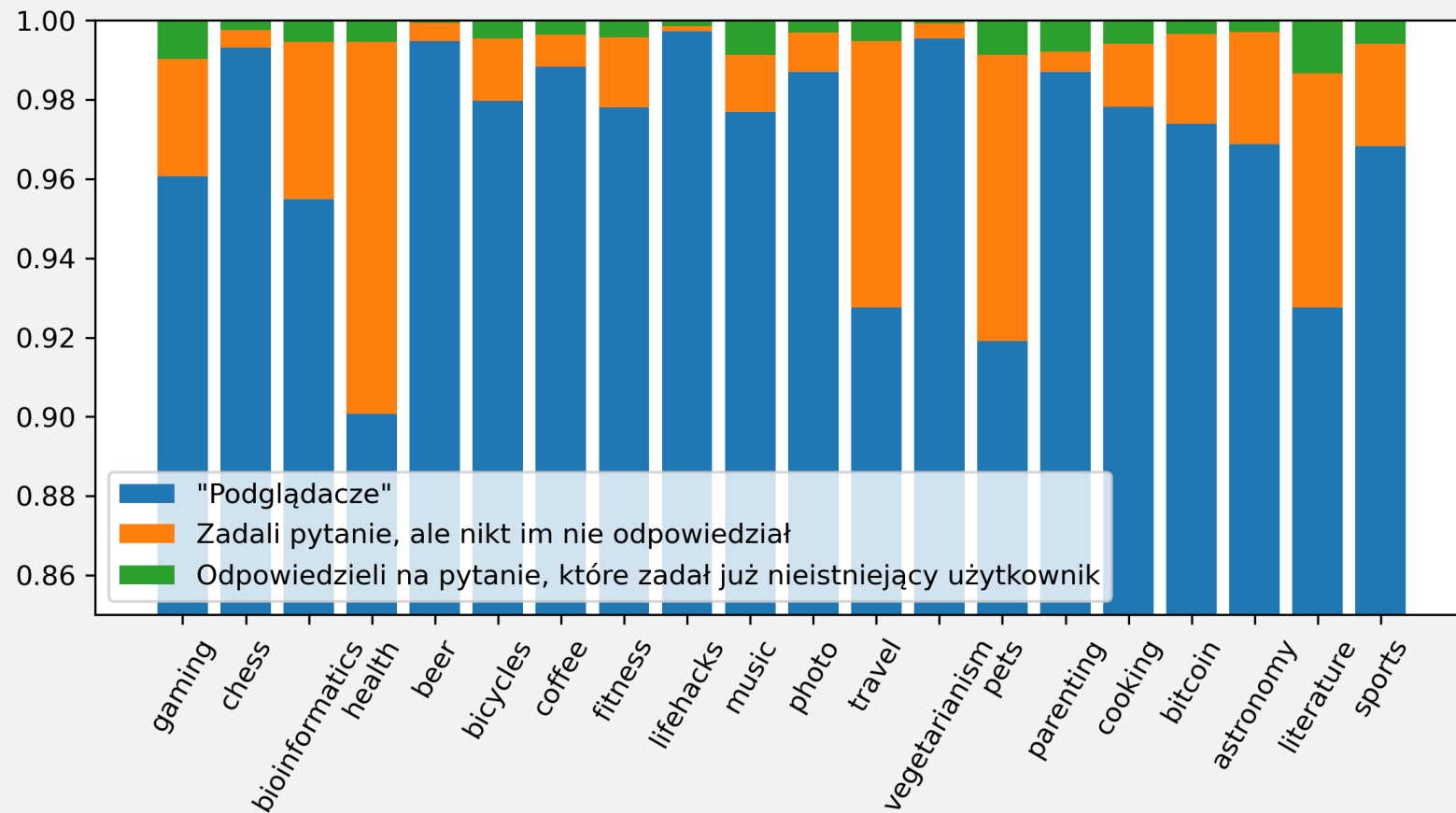
Odsetek użytkowników izolowanych



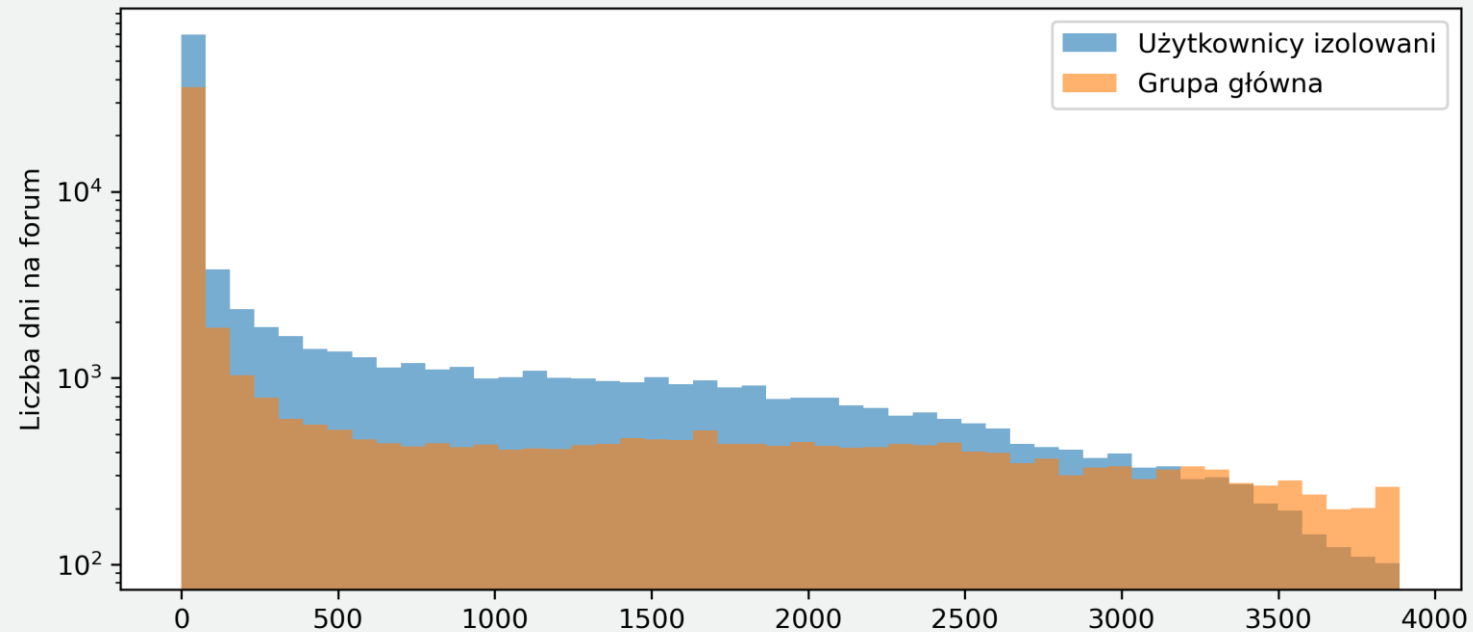
Zależność odsetka użytkowników izolowanych od liczebności forum



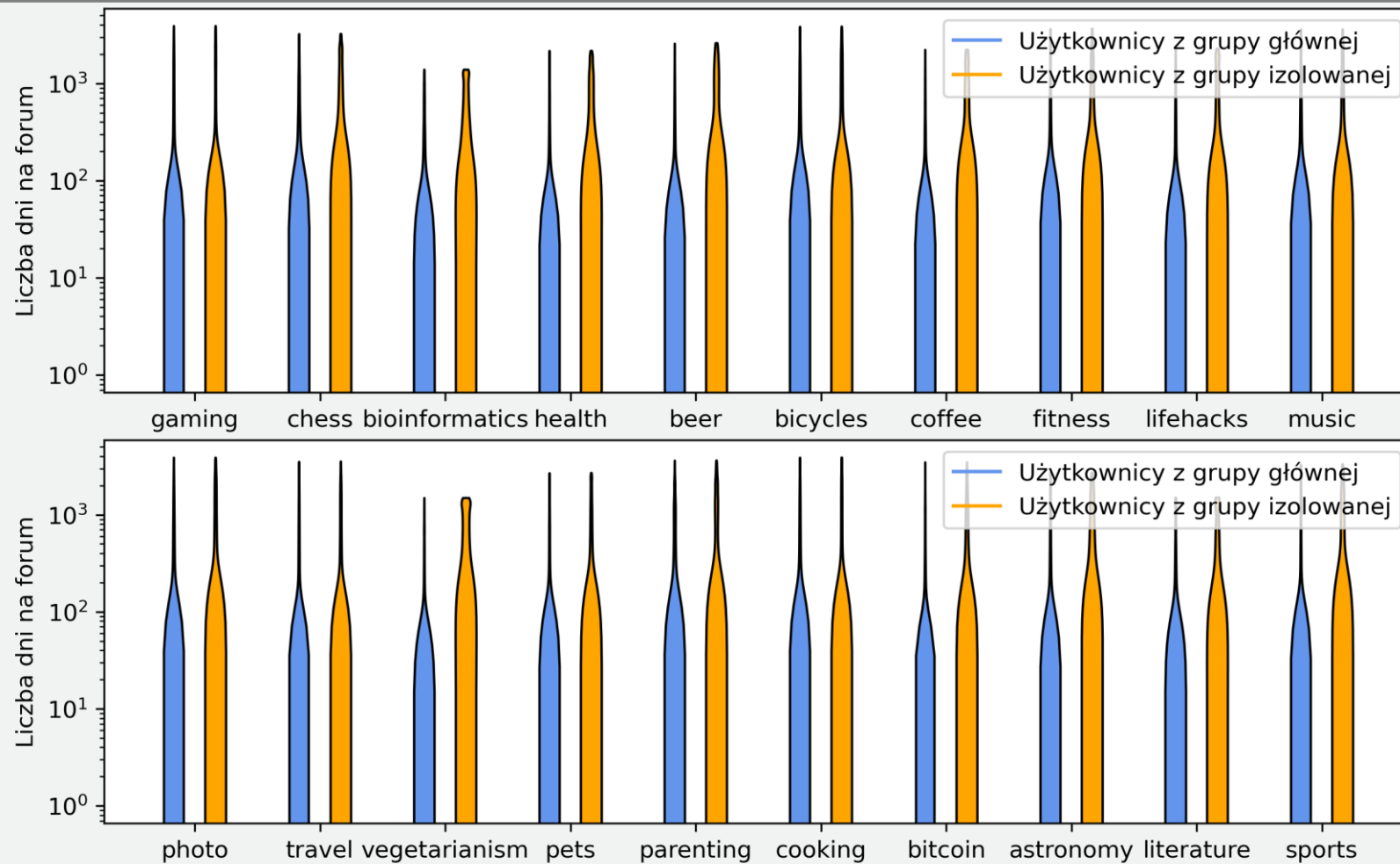
Kim są użytkownicy izolowani?



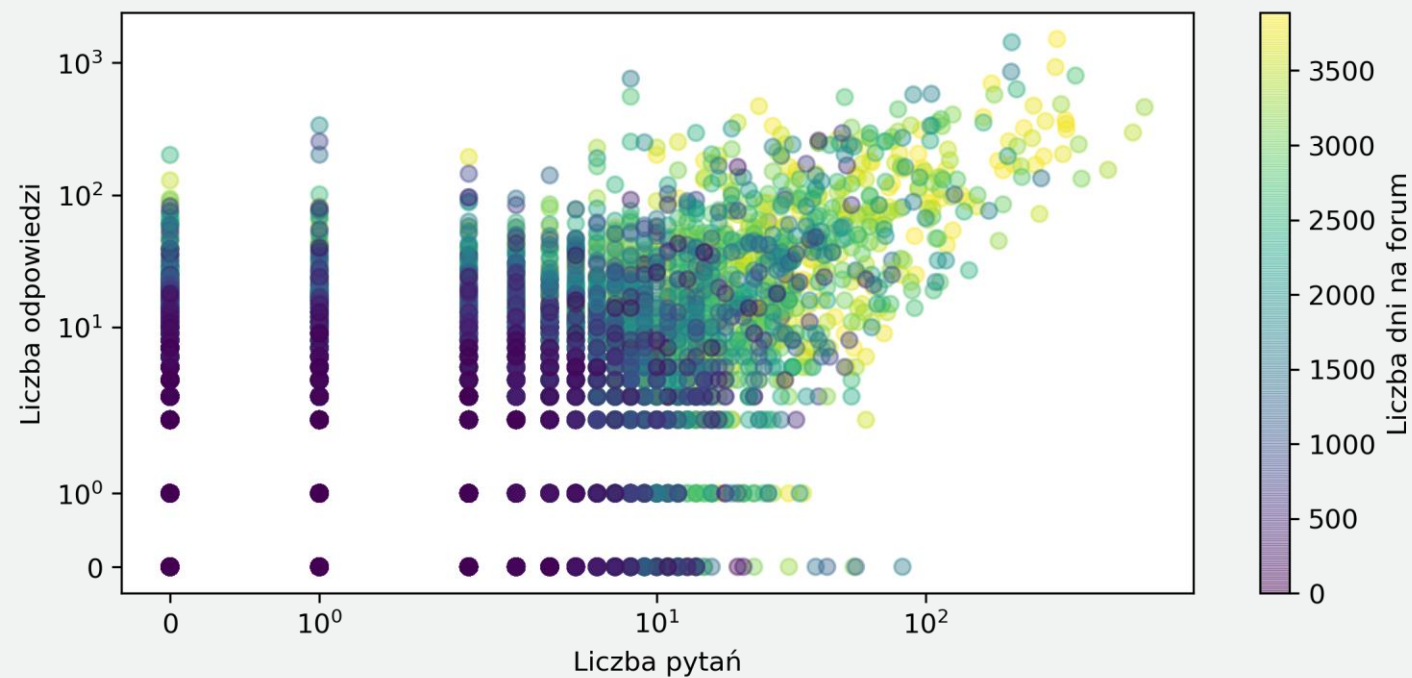
Rozkład liczby dni na forum dla użytkowników izolowanych i członków społeczności na przykładzie **gaming**



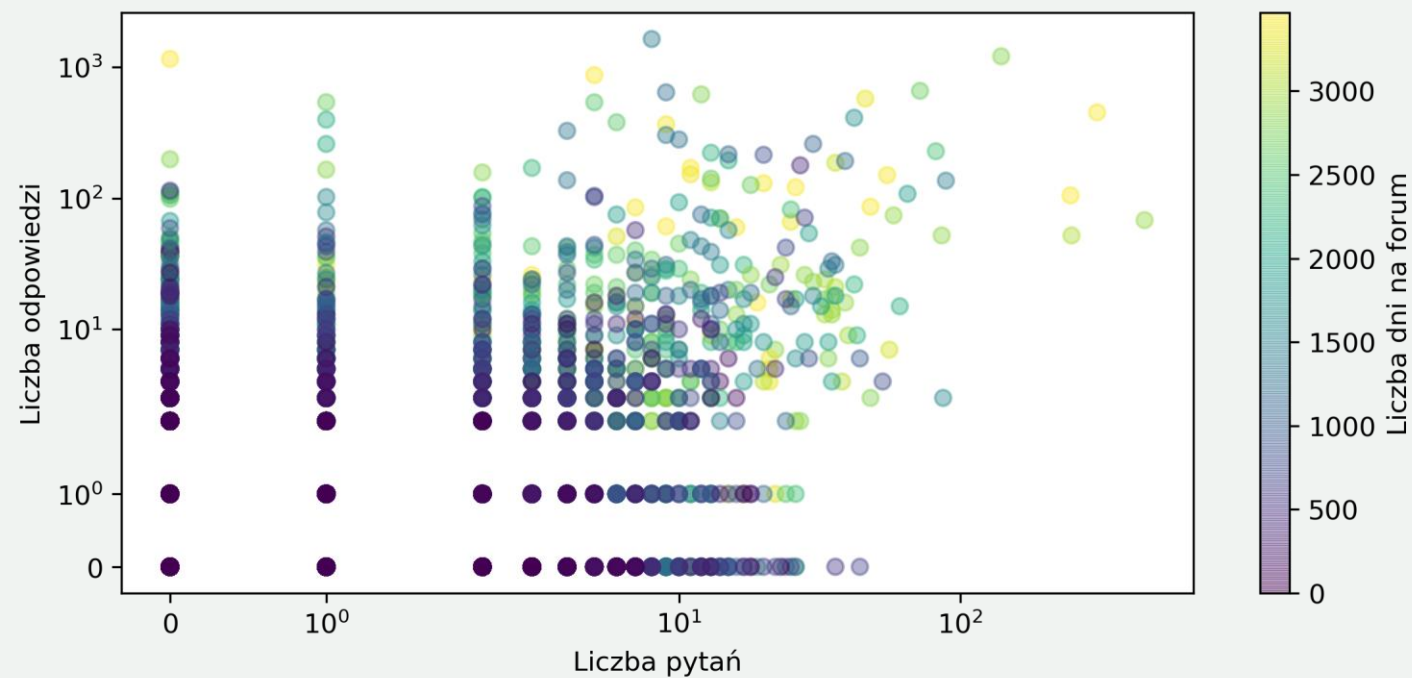
Rozkłady liczby dni na forum dla użytkowników izolowanych i członków społeczności



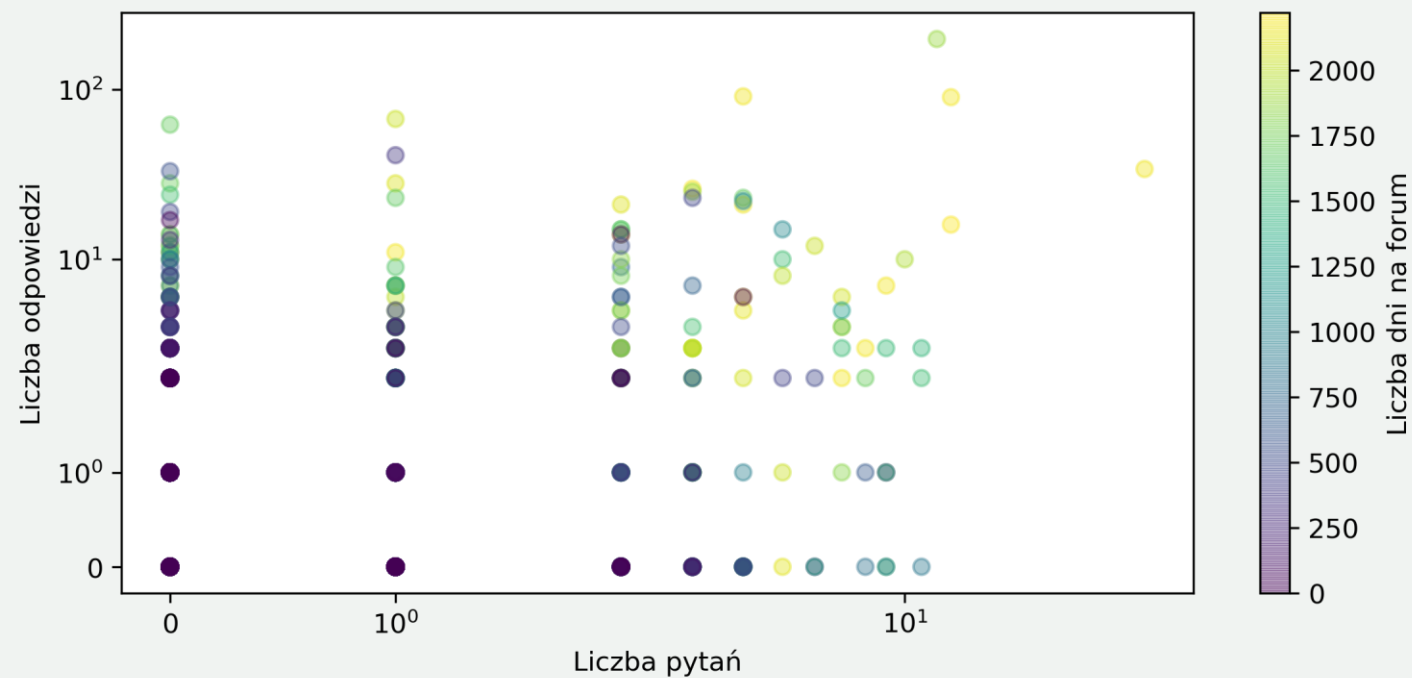
Zależność pomiędzy liczbą pytań, a liczbą odpowiedzi danego użytkownika na przykładzie **gaming**



Zależność pomiędzy liczbą pytań, a liczbą odpowiedzi danego użytkownika na przykładzie **bitcoin**



Zależność pomiędzy liczbą pytań, a liczbą odpowiedzi danego użytkownika na przykładzie **coffee**



Podsumowanie
