

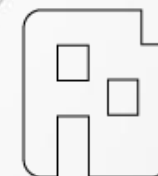
Analiza właściwości sieci współchorobowości w Polsce w oparciu o dane z kart zgonów

Autor: Gabriela Graś
Opiekun: Dr inż. Anna Chmiel



**Wydział
Fizyki**

POLITECHNIKA WARSZAWSKA



**Fizyka
Komputerowa**

FIZYKA TECHNICZNA

Plan prezentacji

1. Cel pracy
2. Omówienie medycznej terminologii
3. Postać danych
4. Konstrukcja sieci
5. Analiza właściwości sieci
6. Wizualizacja
7. Alternatywne wagi
8. Podsumowanie



Cel pracy

Zbadanie sieci współchorobowości
w Polsce na podstawie zbioru 390 tys.
kart zgonów z roku 2013.

- Analiza właściwości sieci z podziałem na płeć
- Zbadanie wpływu kodów śmieciowych



Współchorobowość

Współchorobowość (inaczej choroby współistniejące) to jednoczesne występowanie dwóch lub więcej chorób.

Na **współchorobowość** są narażone przede wszystkim **osoby starsze**. Ze względu na wydłużającą się średnią długość życia zjawisko współchorobowości jest coraz bardziej powszechne.



Międzynarodowa Klasyfikacja Chorób ICD-10

Przykładowy kod

J18.1

Kategoria

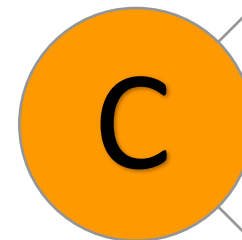
J - Choroby układu
oddechowego

Choroba

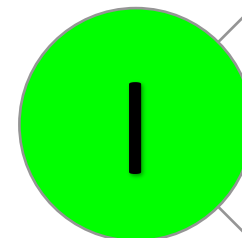
J18 - Zapalenie płuc
wywołane przez
niezidentyfikowany
czynn timer zakaźny

Podtyp choroby

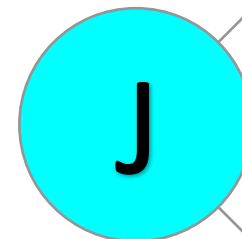
J18.1 - Płatowe
zapalenie płuc,
nieokreślone



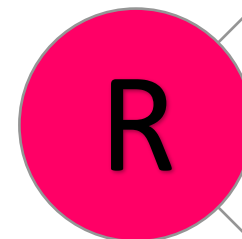
Nowotwory



Choroby układu krążenia



Choroby układu
oddechowego



Objawy, cechy chorobowe
oraz nieprawidłowe wyniki
badań klinicznych gdzie
indziej niesklasyfikowane

Kody śmieciowe

To kody zawierające niedokładne opisy chorób i stanów, przez które nie można stwierdzić właściwej przyczyny zgonu.

WHO określiło listę kodów śmieciowych podstawowych i rozszerzonych.

Przykład:

C76 - nowotwór złośliwy
umieścowań innych
i niedokładnie określonych

Przykład:

J96 - niewydolność oddechowa
niesklasyfikowana gdzie indziej



Opis zbioru danych

Plik csv z 390 tys. wierszy. Każdy wiersz zawiera dane z jednej karty zgonu:

- rodzaj karty zgonu
- płeć osoby zmarłej
- wiek w latach ukończonych
- kod wyjściowej przyczyny zgonu określony przez lekarza-kodera
- wtórne przyczyny zgonu
- bezpośrednie przyczyny zgonu

Na przykład :

Wyjściowa: cukrzyca typu 2



Wtórna: przewlekła niewydolność nerek

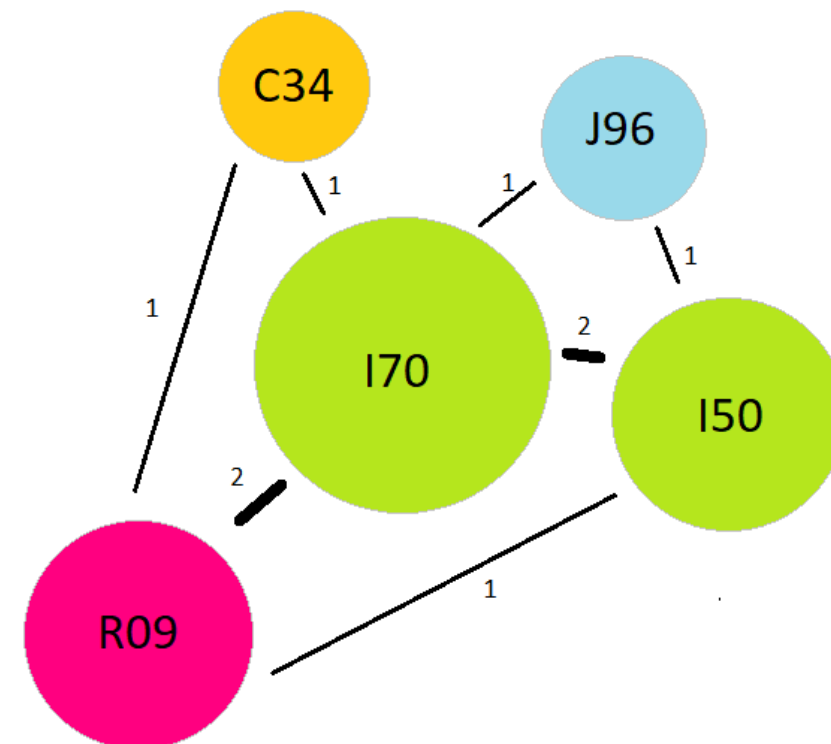


Bezpośrednia: udar niedokrwienny mózgu

Konstrukcja sieci współchorobowości

- Dla każdej pary kodów wyliczono ile osób miało je razem wpisane w kartę zgonu.
- Wszystkie kody wpisane w karcie potraktowano jednakowo, bez podziału na choroby wyjściowe, wtórne i bezpośrednie.
- Utworzono sieć ważoną, w której:
 - **wierzchołki** - kody chorób
 - **waga połączenia** - ilość osób, które chorowały na obie choroby jednocześnie.
- Minimalna waga połączenia to 1, inaczej połączenie nie istnieje.
- Usunięto wierzchołki o zerowym stopniu.

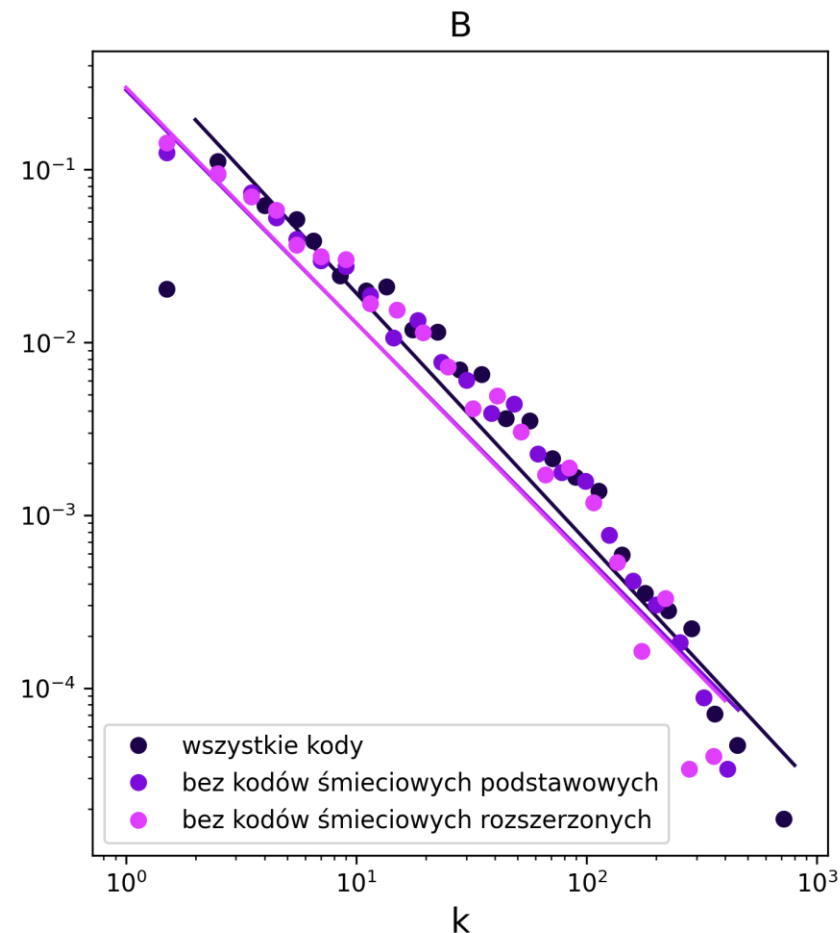
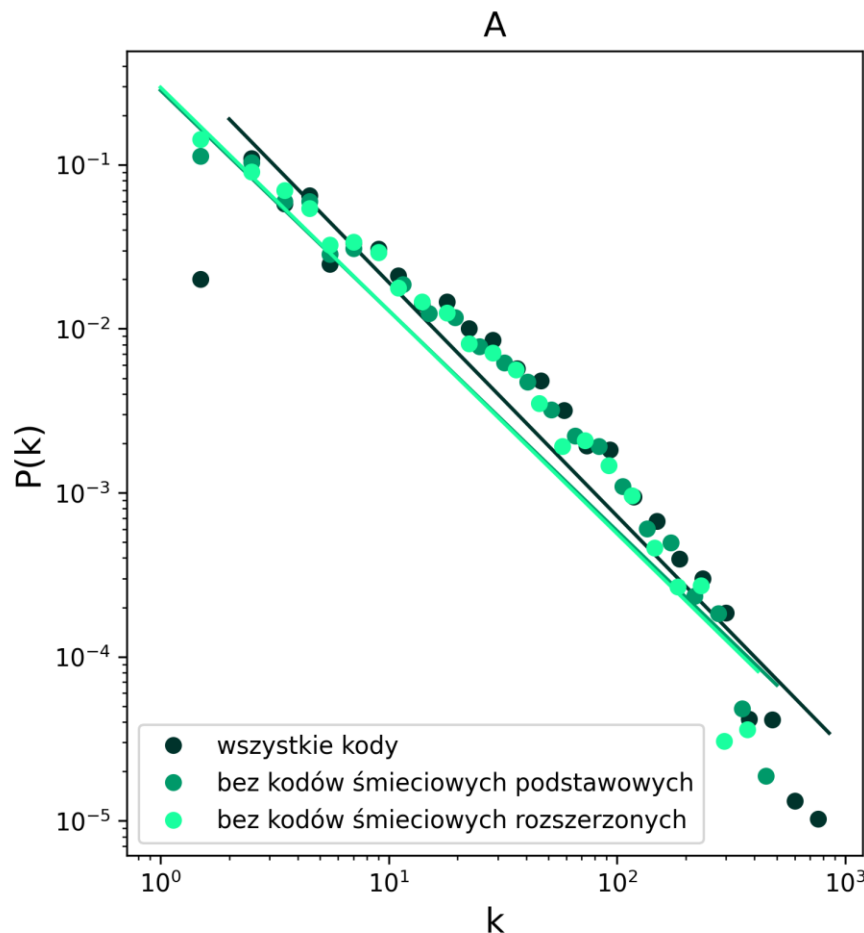
nr karty	p. wyjściowa	p. wtórna	p. bezpośrednia
1	I50	J96	I70
2	C34	R09	I70
3	I50	I70	R09



Rozmiar sieci

Rodzaj sieci	Płeć	Ilość węzłów	Ilość krawędzi
Ze wszystkimi kodami	mężczyźni	1093	23361
	kobiety	1030	21091
Bez kodów śmieciowych podstawowych	mężczyźni	986	16441
	kobiety	911	14647
Bez kodów śmieciowych rozszerzonych	mężczyźni	958	13714
	kobiety	875	12098

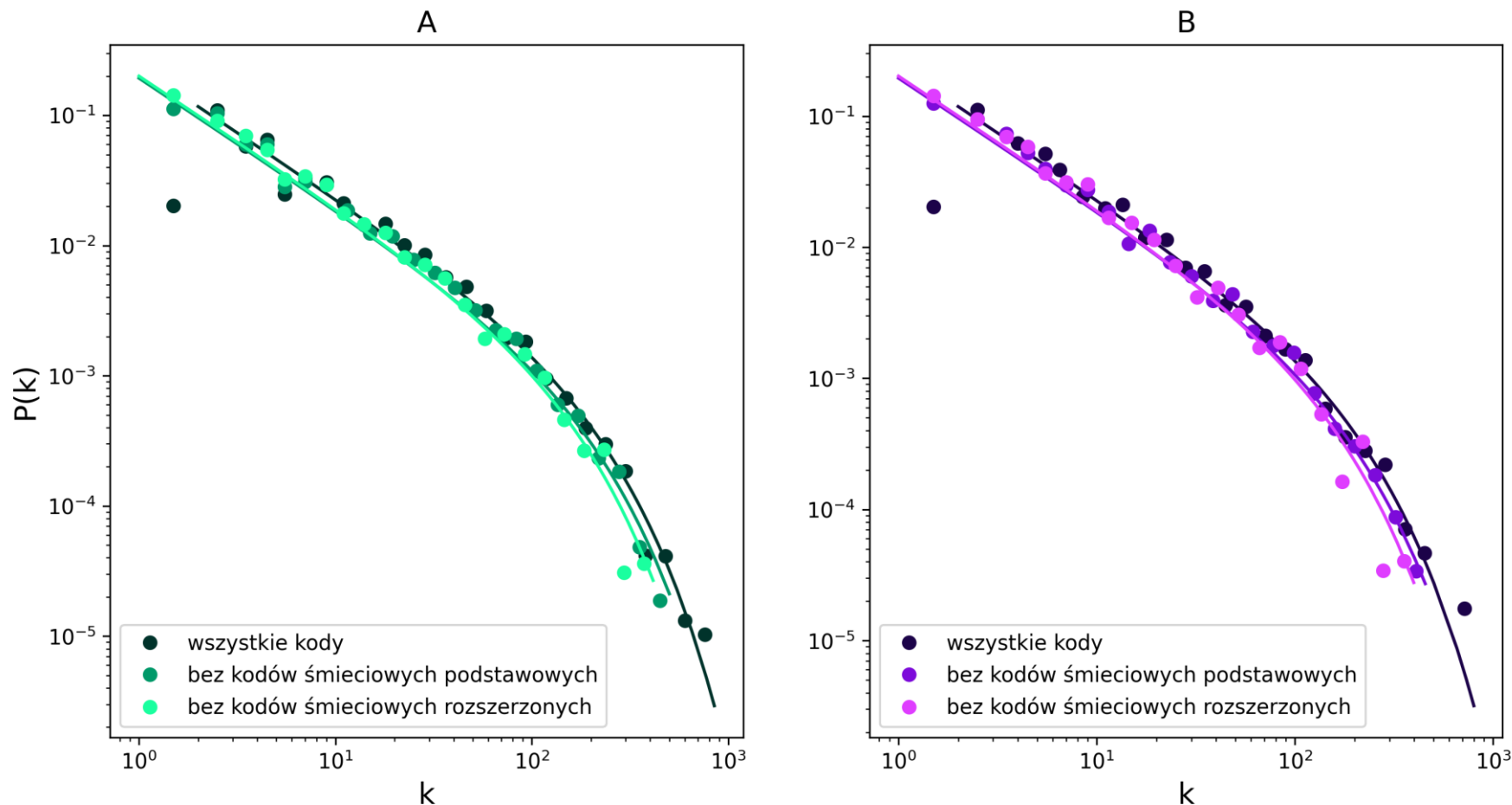
Rozkład stopni wierzchołków



Rozkład stopni wierzchołków z dopasowaniem rozkładu potęgowego dyskretnego $p(x) \sim x^{-\alpha}$
dla A – mężczyzn, B – kobiet.

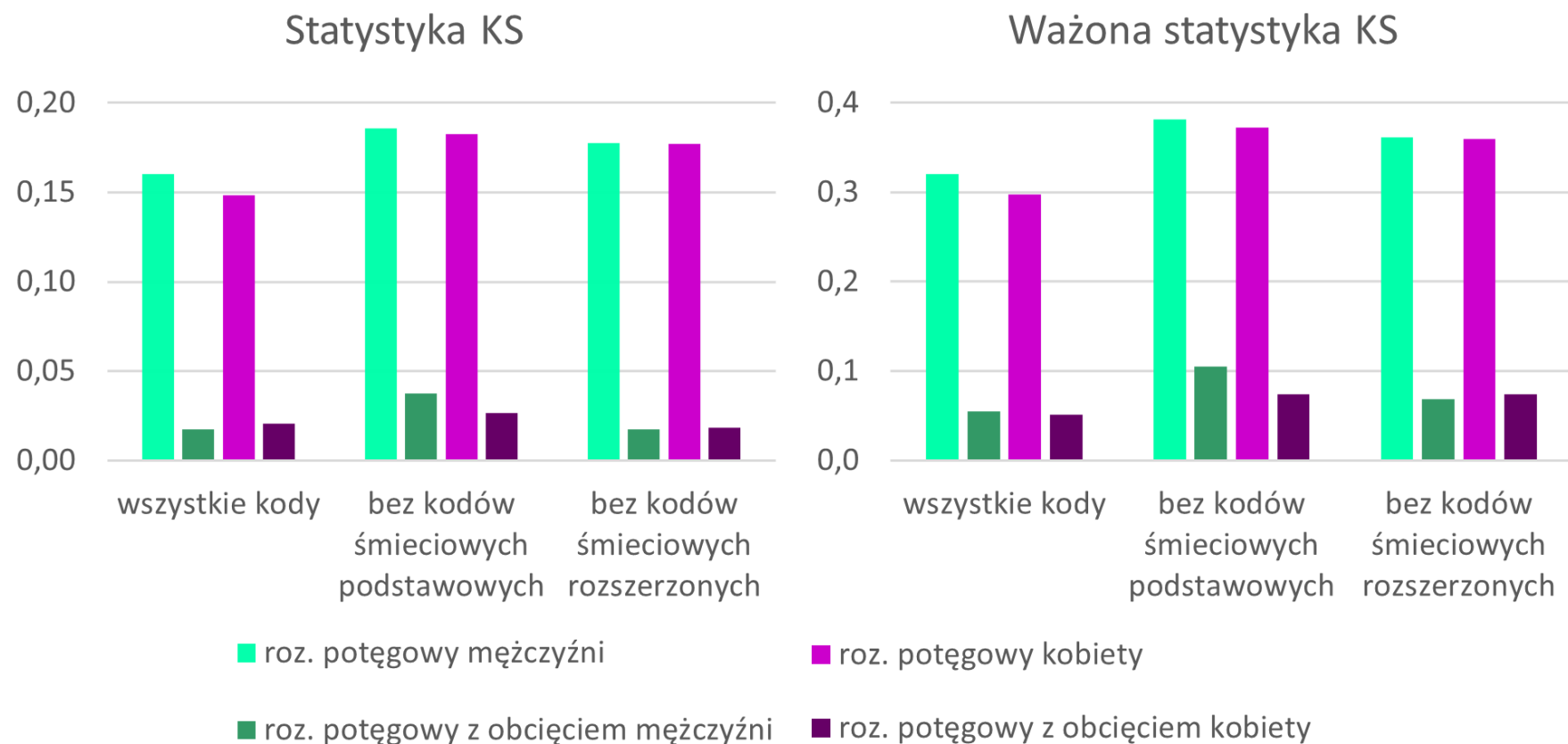
Dla wszystkich rozkładów $\hat{\alpha} \approx 1.4$

Rozkład stopni wierzchołków cd.



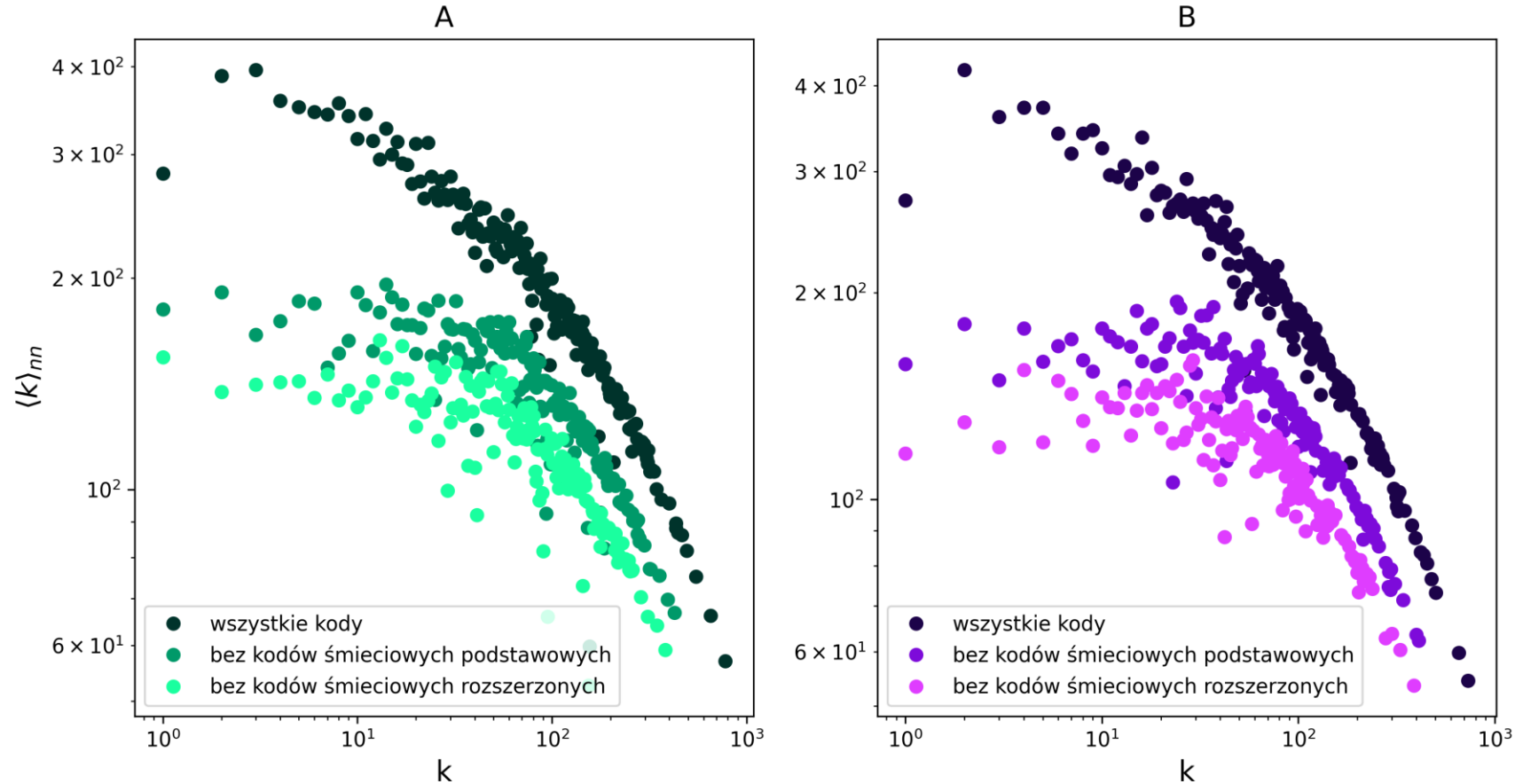
Rozkład stopni wierzchołków z dopasowaniem rozkładu potęgowego dyskretnego z wykładniczym obcięciem $p(x) \sim x^{-\alpha} e^{-\lambda x}$ dla A – mężczyzn, B – kobiet.
Dla wszystkich rozkładów $\hat{\alpha} \approx 1$, $\hat{\lambda} \approx 0.006$

Rozkład stopni wierzchołków cd.



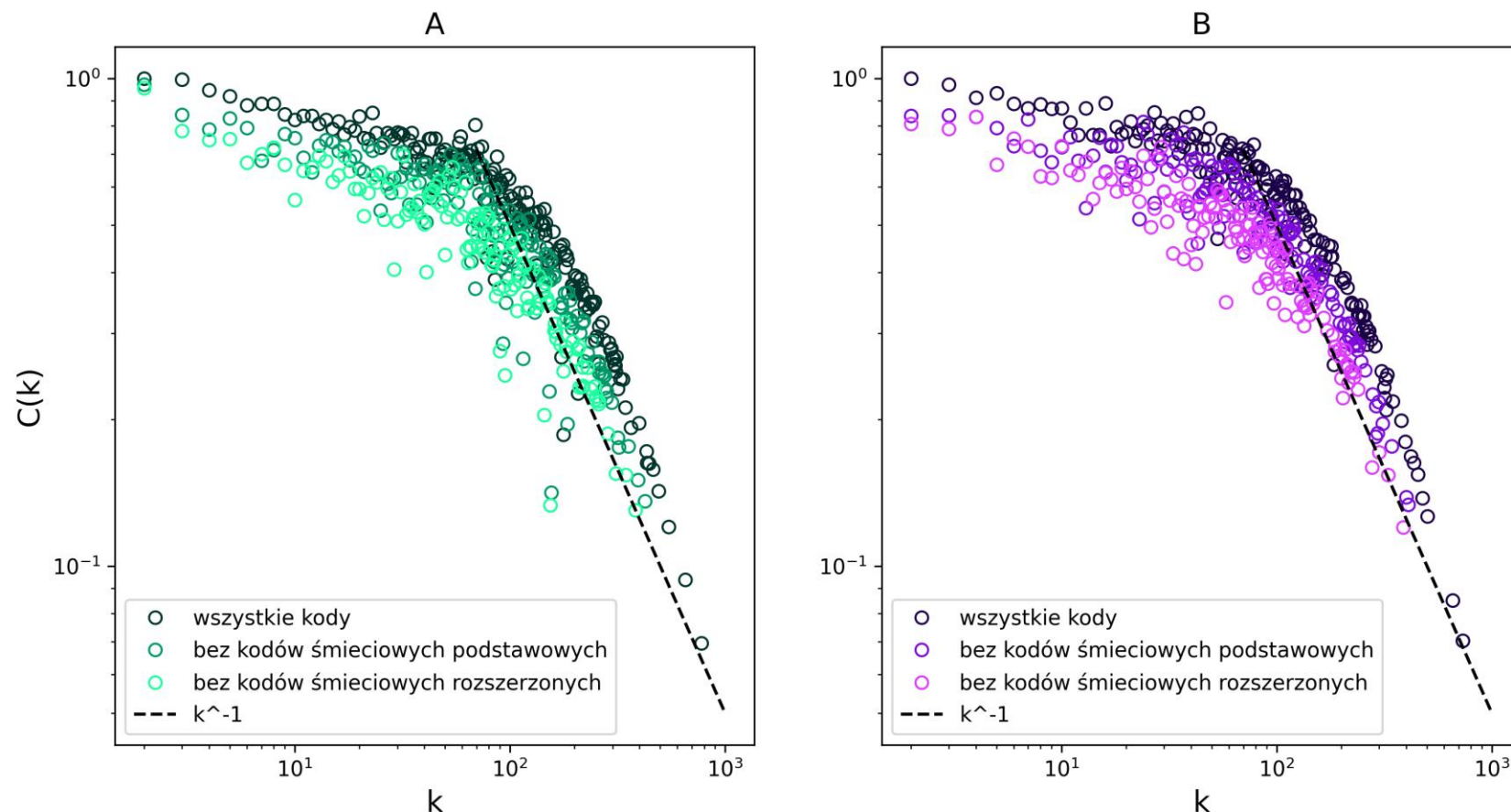
Statystyka Kołmogorowa-Smirnowa (zwykła i ważona) dla dopasowanych rozkładów: potęgowego i potęgowego z wykładniczym obciążeniem.

Zależność średniego stopnia najbliższego sąsiada węzła od jego stopnia



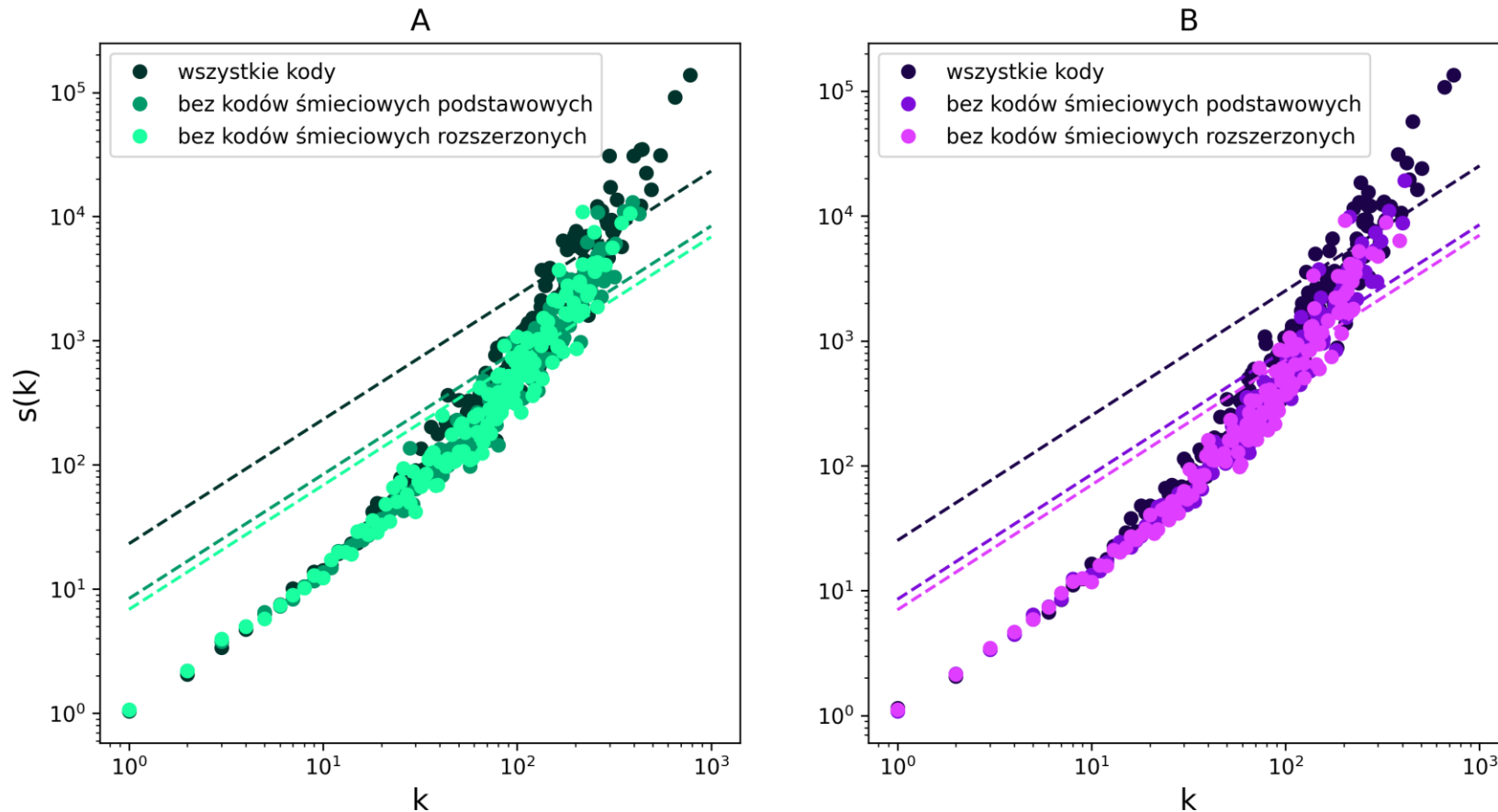
Zależności średniego stopnia najbliższego sąsiada węzła od jego stopnia dla A – mężczyzn i B – kobiet. Malejąca zależność wskazuje na dysasortatywny charakter sieci.

Zależność współczynnika gronowania od stopnia wężła



Zależność współczynnika gronowania od stopnia wężła dla A – mężczyzn i B – kobiet. Linia przerywana ma nachylenie równe -1. Zależność $C(k) \sim k^{-1}$ wskazuje na hierarchiczny charakter sieci.

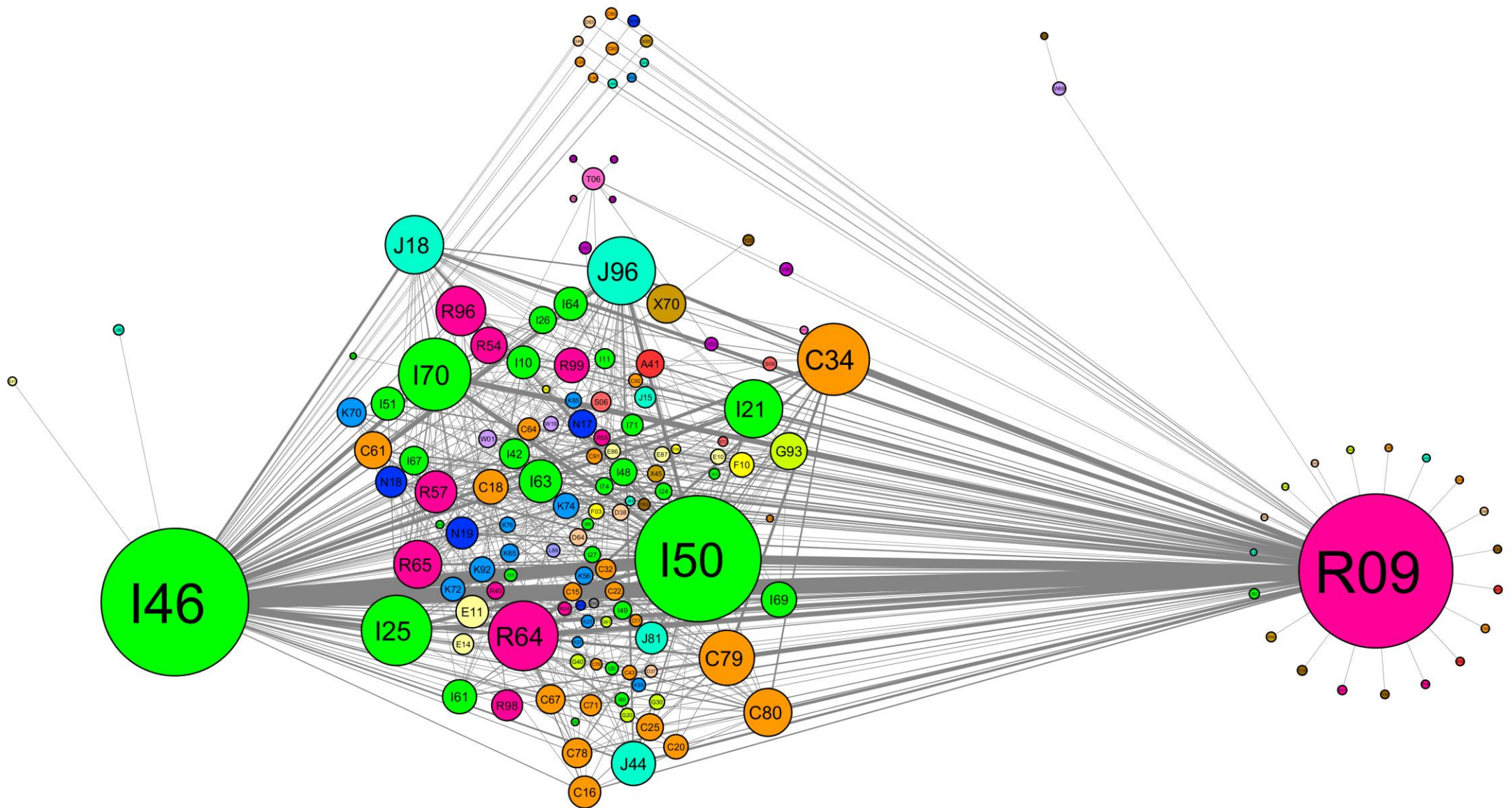
Zależność siły węzła od jego stopnia



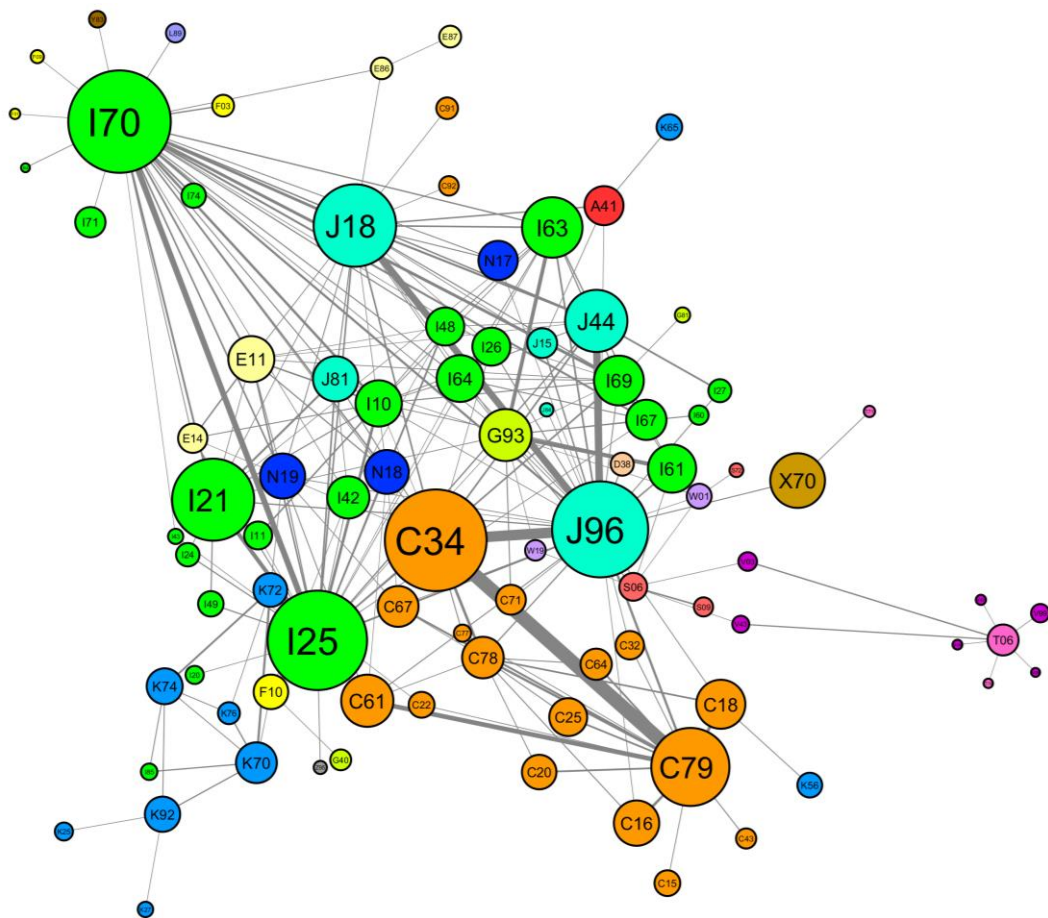
Zależność siły węzła od jego stopnia dla A – mężczyzn i B – kobiet, kolor oznacza zbiór analizowanych kodów, linia przerywana ma nachylenie równe $k \cdot \langle w \rangle$.

$s(k) \neq k \langle w \rangle$, a zatem między siłą węzła a jego stopniem występują korelacje pozytywne.

Sieć dla mężczyzn – wszystkie kody (minimalna waga połączenia równa 100)

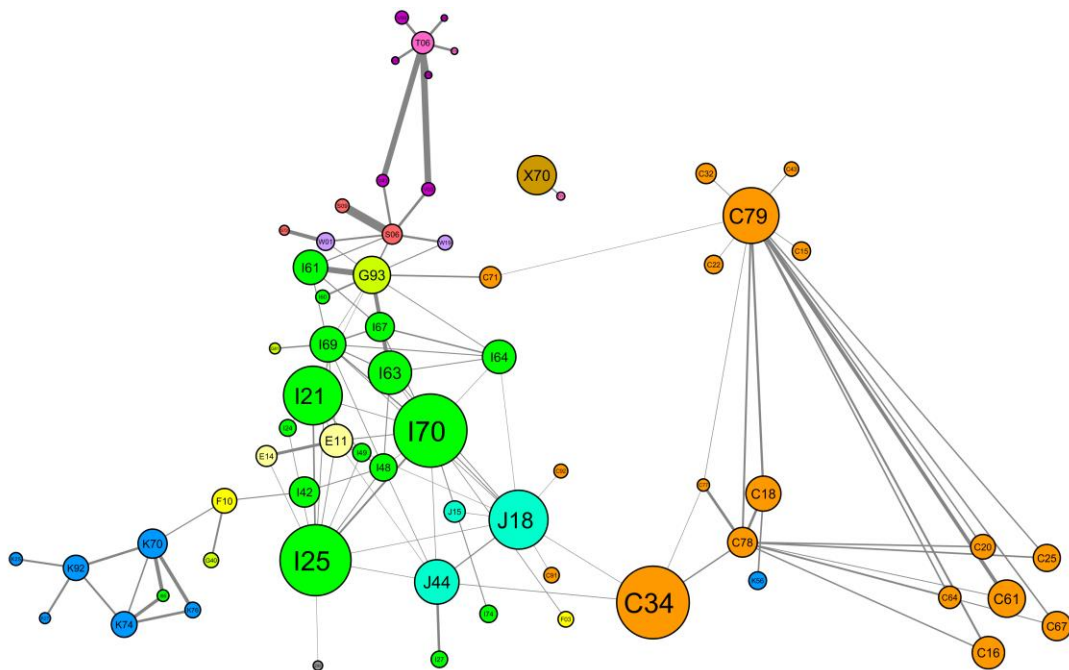


Sieć dla mężczyzn – bez kodów
śmieciowych podstawowych
(minimalna waga połączenia równa 100)

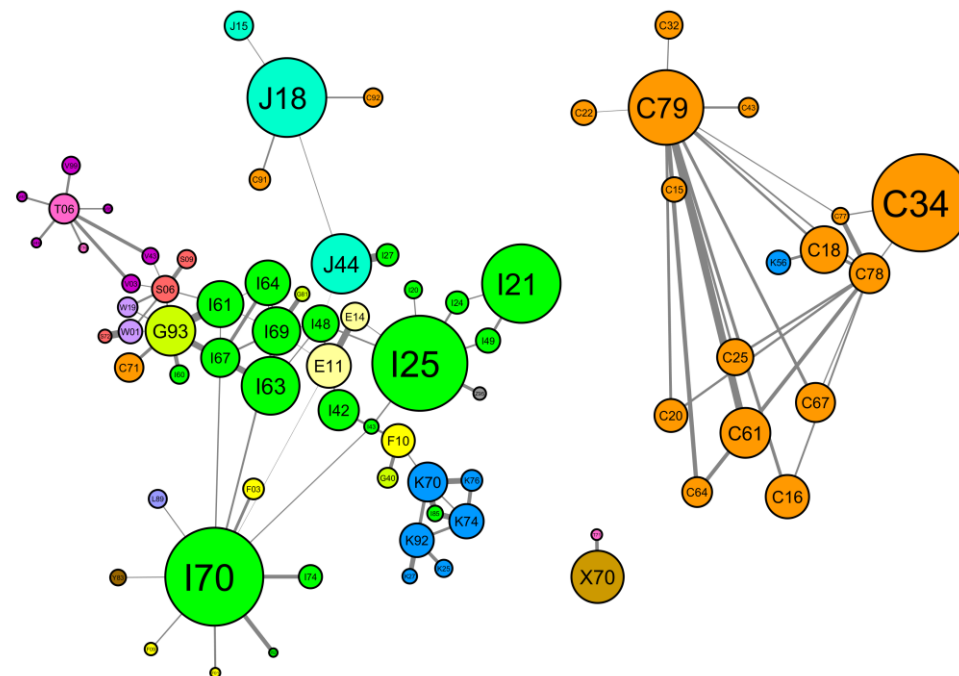


Alternatywne wagi: waga unormowana i skorygowany współczynnik phi

Analizowano sieci z alternatywnymi wagami – **wagą unormowaną** i **skorygowanym współczynnikiem phi ϕ^*** .



Sieć dla mężczyzn z minimalną wagą połączenia równą 100 i minimalną unormowaną wagą połączenia równą 0.01, bez kodów śmieciowych rozszerzonych



Sieć dla mężczyzn z minimalną wagą połączenia równą 100
i skorygowanym współczynnikiem phi połączenia równym 2,
bez kodów śmieciowych rozszerzonych

Podsumowanie i wnioski

- Cechy sieci współchorobowości:
 - Potęgowy rozkład stopni wierzchołków z wykładniczym obcięciem.
 - Dysasortatywny i hierarchiczny charakter.
 - Między stopniem a siłą węzła występują korelacje pozytywne.
 - Kody chorób śmieciowych są jednymi z największych hubów w sieci współchorobowości
- Waga unormowana i skorygowany współczynnik phi pozwalają porównać siłę połączeń między chorobami mimo ich różnego rozpowszechnienia. Są ze sobą silnie skorelowane.

Dziękuję za uwagę



Wagi połączeń

Waga zwykła w :

$$w_{ij} = N_{ij}$$

Waga unormowana w_{norm} :

$$w_{norm_{ij}} = \frac{N_{ij}}{N_i + N_j - N_{ij}}$$

Skorygowany statystyczny współczynnik phi ϕ_{ij}^* :

$$\phi_{ij} = \frac{N_{ij}N - N_jN_i}{\sqrt{N_i(N - N_i)N_j(N - N_j)}}$$

$$\tilde{\phi}_{ij} = \frac{\phi_{ij}}{\langle \phi \rangle_i}$$

$$\phi_{ij}^* = \max\{\tilde{\phi}_{ij}, \tilde{\phi}_{ji}\}$$

Oznaczenia:

N_{ij} – liczba kart w których współwystępują kody chorób i oraz j .

N_i – liczba kart w których występuje kod choroby i

N_j – liczba kart w których występuje kod choroby j

N – liczba kart

ϕ_{ij} – współczynnik phi ϕ

$\tilde{\phi}_{ij}$ - skierowany skorygowany statystyczny współczynnik phi

$\langle \phi \rangle_i$ – średni współczynnik phi dla połączeń wychodzących z węzła i , przy liczeniu średniej brano pod uwagę tylko $\phi_{ij} > 0$