

Międzyosobowa zgodność odpowiedzi w badaniach ankietowych

Zróźnicowanie poglądów w perspektywie
wewnętrz- i międzykrajowej

Tadeusz Krauze (Hofstra University)

Przemek Powałko (IFiS PAN)

Kazimierz M. Słomczyński (IFiS PAN, The Ohio State University)

Główne tematy

- Kontekst badania
- Zróżnicowanie opinii publicznej
 - Pojęcie
 - Pomiar
 - Struktura

Kontekst badania

- Makroskop – globalne dane na określony temat (np. zdrowie, edukacja, opinia publiczna) dla dużej liczby państw
- Dane z projektu „Wartości demokratyczne a zachowania protestacyjne: harmonizacja danych, porównywalność pomiaru i modelowanie wielopoziomowe w perspektywie międzykrajowej”
 - 1721 sondaży z lat 1966-2013
 - 142 kraje/terytoria
 - 2.2 miliony respondentów
 - średnio 228 pytań w kwestionariuszu

Różnice i zróżnicowania

- Różnice na poziomie mikro – niezgodność odpowiedzi pary respondentów (i, j) na pytanie k
- Różnice na poziomie mezo – pojęcie kontroweryjności pytania k i pojęcie odległości między respondentami (i, j)
- Różnice na poziomie makro – niezgodność między wszystkimi respondentami na wszystkie pytania (dyssensus)

Pomiar zróżnicowania (diagram przemienny)

Niezgodność odpowiedzi pary respondentów (i, j) na pytanie k

Kontrowersyjność pytania k

$$d(i, j, k) \longrightarrow d(\bullet, \bullet, k)$$

Odległość między parą respondentów (i, j)

Niezgodność wszystkich par respondentów na wszystkie pytania (dyssensus)

$$d(i, j, \bullet) \longrightarrow d(\bullet, \bullet, \bullet)$$

Wskaźnik niezgodności $d(i, j, k)$

- Definicja:

$$d(i, j, k) = \text{ind}(x(i, k) - x(j, k))$$

gdzie

$i, j = 1, \dots, m$ – numer respondenta; m – liczba respondentów

$k = 1, \dots, n$ – numer pytania; n – liczba pytań

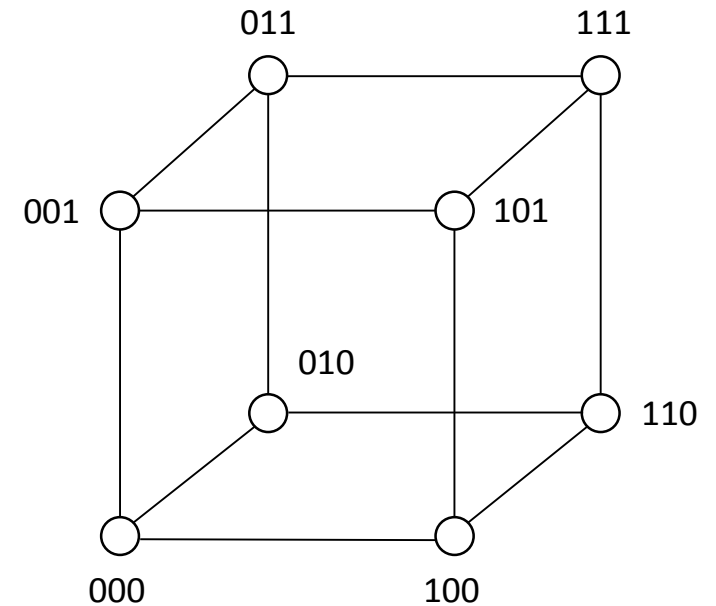
$x(i, k)$ – odpowiedź respondenta i na pytanie k

$$\text{ind}(z) = \begin{cases} 0, & \text{jeżeli } z = 0 \\ 1, & \text{jeżeli } z \neq 0 \end{cases}$$

- $d(i, j, k)$ oznacza zgodność (0) lub niezgodność (1) odpowiedzi pary respondentów (i, j) na pytanie k

Analiza odległości $d(i, j, \bullet)$

- $d(i, j, \bullet)$ spełnia aksjomaty odległości
 - Zastrzeżenie: $i \neq j \Rightarrow d(i, j, \bullet) \geq 1$
- $d(i, j, \bullet)$ są sumą jedynek ciągu zero-jedynkowego mającego n wyrazów; takie ciągi są punktami w n -wymiarowej przestrzeni Hamminga H^n



Rozkład odległości $f(r)$

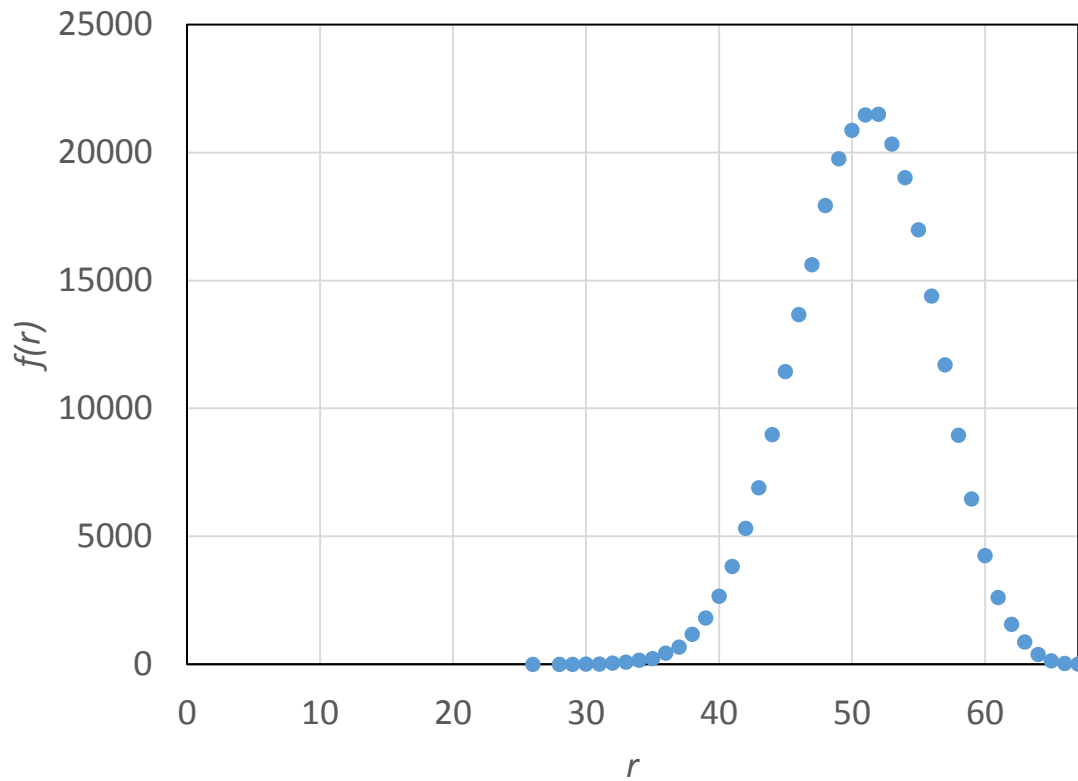
- Dla $r = 0, \dots, n$, funkcja

$$f(r) = \|(i, j): d(i, j, \bullet) = r\|$$

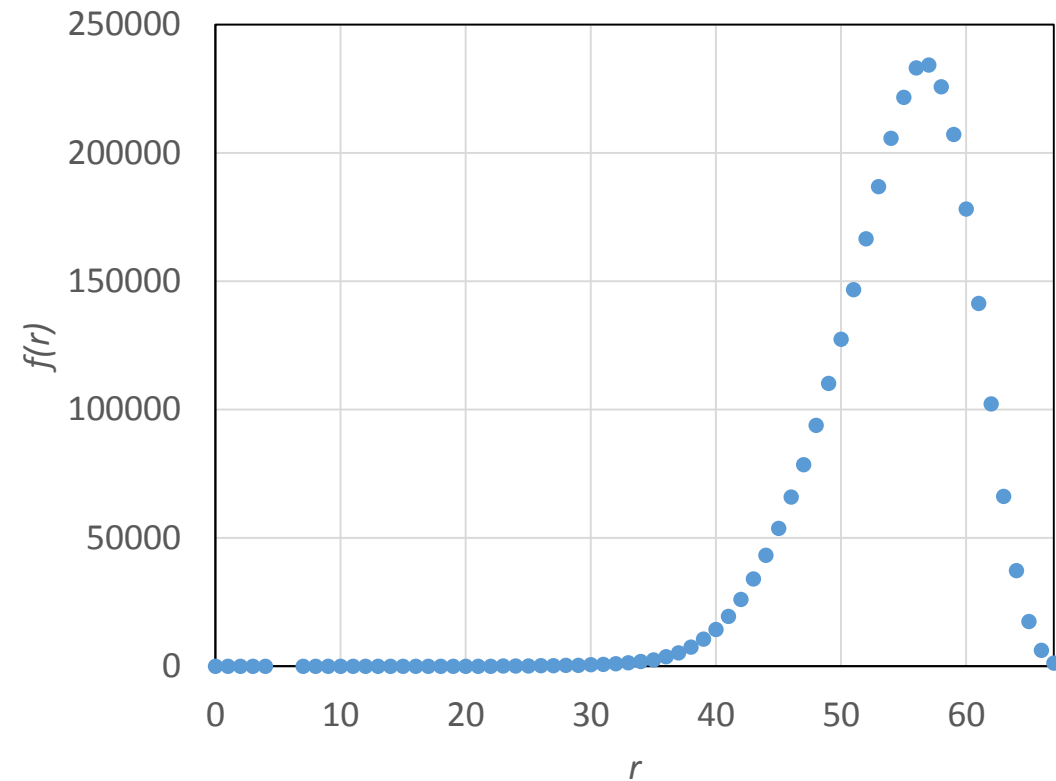
- $f(0) > 0 \implies$ NUR (*non-unique record*)
- $f(n) > 0 \implies$ POR (*polar-opposite record*)
- Po eliminacji zmiennych technicznych i metryczkowych, w zbiorze 1721 sondazy znaleziono 5893 NUR-ów (nieprecyzyjnie mówiąc: duplikatów)
- Bycie NUR-em jest relacją przechodnią, podczas gdy bycie POR-em – nie jest

Przykładowe rozkłady odległości

ESS/6 IS ($m=752, n=67$)



ESS/6 RU ($m=2484, n=67$)



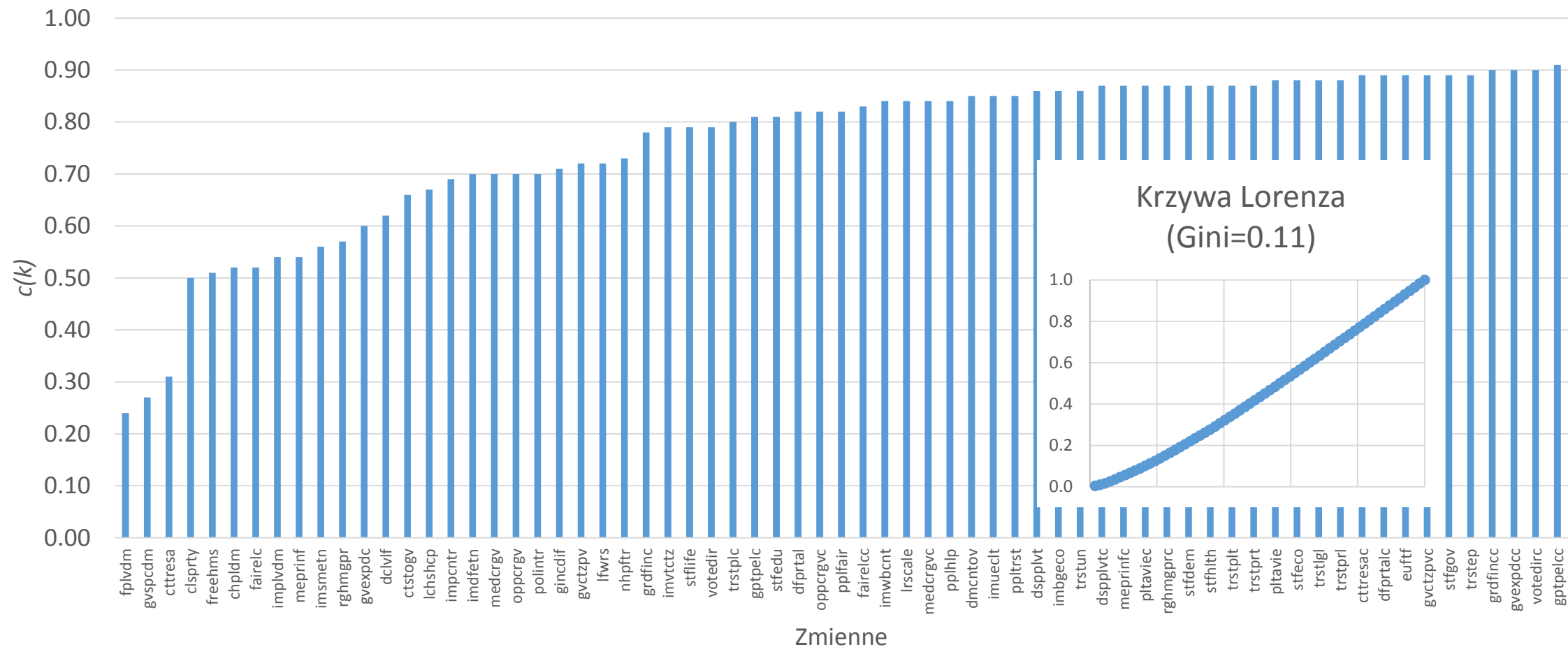
Kontrowersyjność pytań $d(\bullet, \bullet, k)$

- Miara kontrowersyjności pytania k jest liczbą niezgodnych odpowiedzi na pytanie k wśród wszystkich par respondentów (i, j)
- Wskaźnik kontrowersyjności pytania k

$$c(k) = \frac{d(\bullet, \bullet, k)}{m(m-1)}$$

- Możliwość uporządkowania pytań ze względu na kontrowersyjność; krzywa Lorenza; indeks Giniego

Kontrowersyjność 67 pytań w ESS/6 Islandia



Maksymalny dyssensus

- Największa możliwa liczba niezgodnych odpowiedzi na pytanie k

$$D_{max}(k) = m^2 - [(s - v)u^2 + v(u + 1)^2]$$

gdzie

s – liczba możliwych odpowiedzi na pytanie k

$$m = su + v = (s - v)u + v(u + 1)$$

gdzie $0 \leq v < s$

- $D_{max}(\bullet)$ nazywamy maksymalnym dyssensusem sondażu

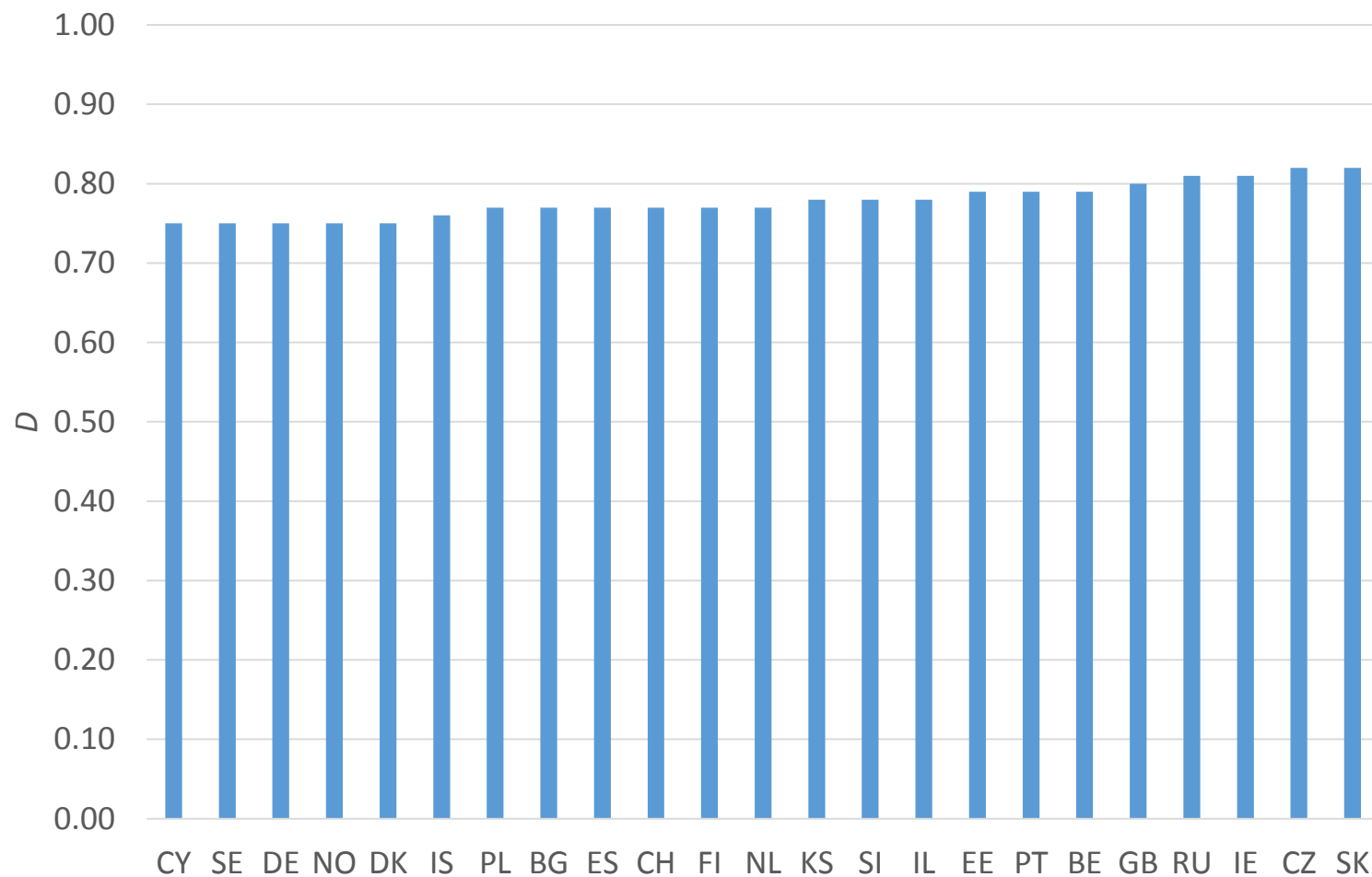
Całkowity dyssensus w sondażu $d(\bullet, \bullet, \bullet)$

- $d(\bullet, \bullet, \bullet)$ = suma kontrowersyjności wszystkich pytań w sondażu
- $d(\bullet, \bullet, \bullet)$ = suma odległości wszystkich par respondentów
- Znormalizowana wartość dyssensusu całkowitego

$$D = \frac{1}{m(m-1)n} d(\bullet, \bullet, \bullet)$$

- Przykładowo w ESS/6 dla Islandii $D=0.76$, dla Rosji $D=0.81$

Całkowity dyssensus - przykład ESS/6 ($n=67$)



Odległość a kontrowersyjność

Twierdzenie:

W sondażu mającym m respondentów i n pytań, „przeciętna” odległość wszystkich respondentów jest równa „przeciętnej” kontrowersyjności wszystkich pytań.

Obie te wielkości są równe całkowitemu dyssensusowi D .

Pojęcie struktury zróżnicowania

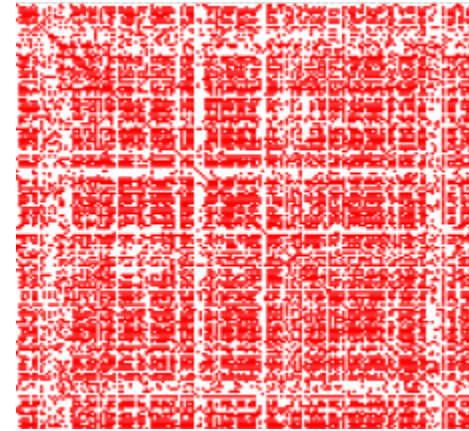
- Struktura jako szczególny zbiór otoczeń (sąsiedztw) respondenta
- Otoczenie – zbiór respondentów o podobnych poglądach
- Stopień podobieństwa poglądów – promień otoczenia

Struktura zróżnicowania

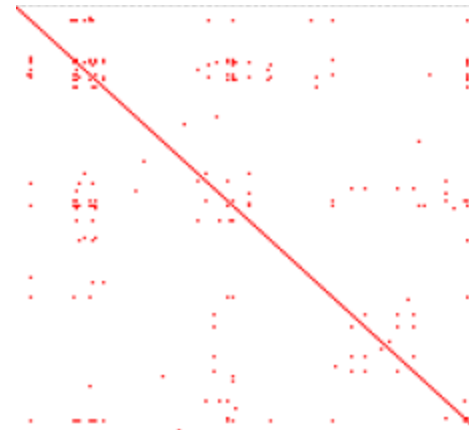
- Otoczenie (sąsiedztwo) o promieniu r respondenta i

$$N(i, r) = \{j\}: d(i, j, \bullet) \leq r$$

- Problem pokrycia zbioru (*set covering problem*)
- Przykładowa reprezentacja macierzy wymiaru $m \times m$: jedynki oznaczone na czerwono

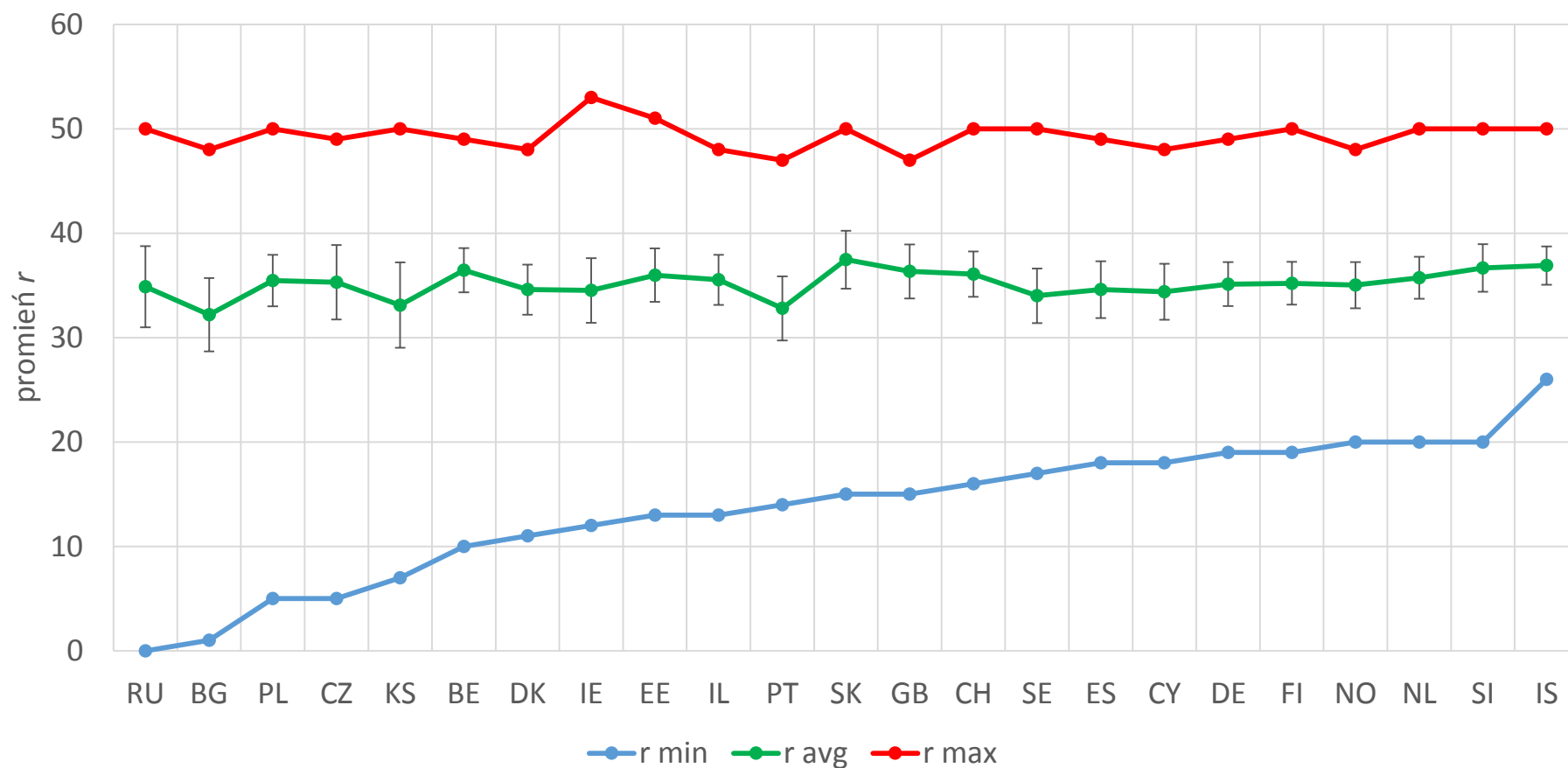


ESS/6 Islandia
 $r=50$



ESS/6 Islandia
 $r=37$

MIN, AVG, MAX promień r w 23 krajach ESS/6



Dziękujemy, zapraszamy do dyskusji.