

**Mirosław Krzyśko<sup>1</sup>, Marcin Krzyśko<sup>2</sup>, Anna Kierczyńska<sup>1</sup>,  
Waldemar Ratajczak<sup>3</sup>**

<sup>1</sup> Uniwersytet Kaliski

MK: [m.krzyisko@uniwersytetkaliski.edu.pl](mailto:m.krzyisko@uniwersytetkaliski.edu.pl),  <https://orcid.org/0000-0001-8075-4432>

AK: [a.kierczyńska@uniwersytetkaliski.edu.pl](mailto:a.kierczyńska@uniwersytetkaliski.edu.pl),  <https://orcid.org/0000-0002-8234-7774>

<sup>2</sup> Boulie Ltd., Warszawa

[marcinkrzyisko@gmail.com](mailto:marcinkrzyisko@gmail.com),  <https://orcid.org/0009-0001-4456-4049>

<sup>3</sup> Uniwersytet im. Adama Mickiewicza w Poznaniu

Wydział Geografii Społeczno-Ekonomicznej i Gospodarki Przestrzennej

[walrat@amu.edu.pl](mailto:walrat@amu.edu.pl),  <https://orcid.org/0000-0002-0565-0761>

# Geografia współwystępowania wypadków drogowych: kolokacja zdarzeń w czasie i przestrzeni – studium przypadku Warszawy i Chicago

**Zarys treści:** Iloraz kolokacji w wersji czasoprzestrzennej (GTWCLQ – Geograficznie oraz Czasowo Ważony Iloraz Współkolokacji) ustala, czy zdarzenia jednej kategorii sąsiadują w przestrzeni i w czasie ze zdarzeniami innej kategorii częściej, niż przy rozmieszczeniu przypadkowym. W tej pracy stosujemy go do wszystkich zdarzeń drogowych w Warszawie (2016–2024; 271 706 zdarzeń) i w Chicago (2023; 109 545 zdarzeń). Skala współwystępowania okazuje się cechą rodzaju zdarzenia, nie miasta – np.: najechanie na dziurę w jezdni skupia się punktowo, potrącenie pieszego wzdłuż arterii, zderzenie boczne nie wykazuje skupień wcale. Obraz potrąceń pieszych powtarza się w obu miastach – śródmieście na poziomie losowym, natomiast wyraźne skupienia zdarzeń występują na arteriach obrzeżnych. Praca ma charakter aplikacyjny. Proces obliczeń przedstawiony jest w załączniku A.

**Słowa kluczowe:** iloraz kolokacji, wypadki drogowe, geografia bezpieczeństwa ruchu, skupienia przestrzenne, Warszawa

## Wprowadzenie

Mapa wypadków drogowych w dużym mieście wygląda na pierwszy rzut oka tak samo jak mapa ruchu: wypadki najczęściej występują tam, gdzie jeździ najwięcej pojazdów. Ten obraz jest prawdziwy, ale mało użyteczny, ponieważ nie odróżnia miejsc, w których zdarzeń jest dużo, bo dużo jest czynników sprzyjających

zdarzeniom drogowym, od miejsc, w których zdarzeń pewnej kategorii jest więcej, niż wynikałoby to z liczby tych czynników. Dopiero te drugie lokalizacje wskazują punkty wymagające interwencji. Jest to więc aspekt geograficzny w klasycznym sensie, gdyż dotyczy lokalizacji, skali zjawiska i relacji pomiędzy zjawiskami sąsiadującymi w przestrzeni.

Iloraz kolokacji opisuje dokładnie tę sytuację. Dla wybranego rodzaju zdarzeń ustala, jaki udział mają zdarzenia badanej kategorii w otoczeniu i porównuje ten udział z udziałem w całym w obszarze referencyjnym, jakim może być np. obszar miasta. Miara wprowadzona przez Leslie i Kronenfelda (2011) ma trzy własności, które czynią ją istotną dla badań geograficznych. Jest **odniesiona do określonego obszaru referencyjnego**: mianownik zawiera całą populację zdarzeń, a więc rozkład czynników sprzyjających, przez co gęste śródmieście nie jest automatycznie obszarem problemowym. Jest **kierunkowa**: relacja rodzaju zdarzenia A z rodzajem zdarzenia B nie musi być taka sama jak relacja B z A. I ponadto dla każdego zdarzenia jest wyznaczana osobno (Cromley i in. 2014). Zamienia to jedną liczbę w warstwę mapy, którą można nakładać na inne warstwy danych. Rozszerzenie czasoprzestrzenne (GTWCLQ) wprowadza do sąsiedztwa przestrzennego okno czasowe, dzięki czemu można pytać nie tylko „gdzie”, ale i „w jakim odstępie czasu”. Miarę tę m.in. zastosowano do badania przestępczości w Manchesterze (Li i in. 2022) oraz do incydentów zdrowotnych w Nanning (Li i in. 2026).

Do analizy zdarzeń drogowych według wiedzy autorów tej metody, jak dotąd, nie stosowano, pomimo że rejestry wypadków są dla niej materiałem niemal idealnym: mają dokładną lokalizację, czas z dokładnością do minuty i rozbudowany katalog rodzajów zdarzeń. Niniejsza praca wypełnia tę lukę i jej celem jest ustalenie odpowiedzi na trzy pytania:

1. **Według jakiej skali** przestrzennej i czasowej można identyfikować skupienia się poszczególnych kategorii zdarzeń drogowych i czy skala ta jest cechą rodzaju zdarzenia, czy cechą miasta lub innej jednostki osadniczej?
2. **Jaki jest rozkład przestrzenny skupień**, których nie wyjaśnia samo natężenie ruchu?
3. Czy relacje między rodzajami zdarzeń są **jednostronne**, to jest – czy w sąsiedztwie zdarzeń jednego rodzaju drugi rodzaj pojawia się częściej niż odwrotnie? Odpowiedzi na te pytania udzielimy na podstawie studium przypadków dwóch miast o odmiennej strukturze przestrzennej położonych na różnych kontynentach.

Powtarzalność obrazu w Warszawie i Chicago traktujemy jako sprawdzenie, czy ustalamy własność zjawiska, czy właściwość jednego zbioru danych.

Praca zorganizowana jest tak, aby dała się czytać bez znajomości aparatu statystycznego. W punkcie 2 wyjaśniamy, w jaki sposób wykonywane są pomiary oraz interpretowane są ryciny. W punkcie 3 scharakteryzowane są dane wykorzystane w analizach umożliwiających osiągnięcie celów badawczych. Punkt 4 zawiera wyniki analizy czaso-przestrzennej zdarzeń drogowych, należących do różnej kategorii. W punkcie 5 przeprowadzono interpretację wyników oraz sformułowano wnioski dla praktyki. Sposób wyznaczania wartości liczbowych GTWCLQ oraz wprowadzone przez nas ulepszenia opisuje załącznik A. Czytelnik, który zechce metodę zastosować do własnych danych, znajdzie tam też listę pojęć, które należy uwzględnić we własnym opisie.

## Wykonywanie pomiarów oraz interpretacja rycin

### Przykład empiryczny

Określenie miary przedstawimy na poniższym przykładzie. Bierzemy konkretne najechanie na dziurę w Warszawie, zakreślamy wokół niego okrąg o promieniu 300 m i liczymy dwie rzeczy: ile zdarzeń drogowych w ogóle znalazło się w tym okręgu oraz ile z nich to również najechania na dziurę. Pierwsza liczba jest miarą okazji, druga miarą zjawiska, które badamy. Ich stosunek to udział dziur w jej otoczeniu. Ten udział porównujemy z udziałem dziur w całym mieście, który wynosi 1,26% (3433 najechania na dziurę na 271 706 wszystkich zdarzeń). Iloraz obu udziałów jest wartością miary CLQ.

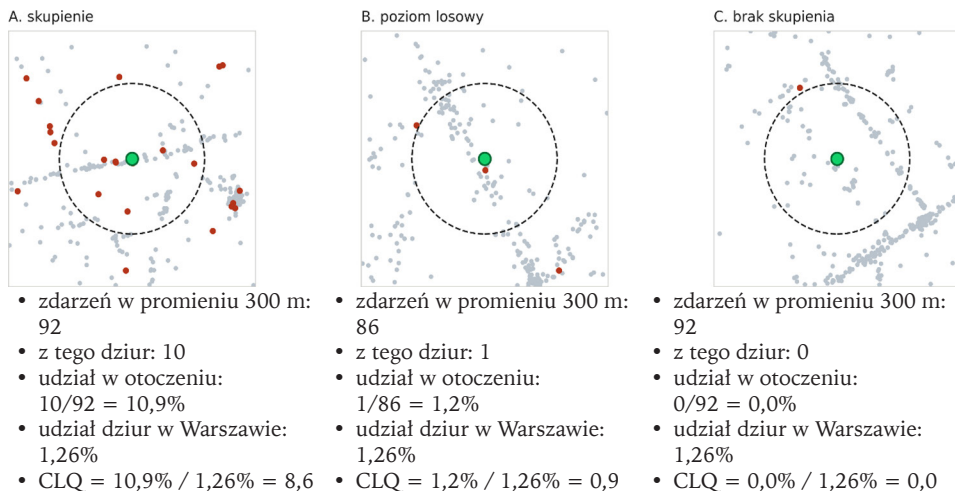
Wymaga to rozróżnienia nazwy metody od symbolu wyniku. CLQ (ang. co-location quotient, iloraz kolokacji) jest miarą wprowadzoną przez Leslie i Kronenfelda (2011) i oznacza wielkość liczbową: stosunek udziału zdarzeń danej kategorii w sąsiedztwie zdarzenia do udziału tej kategorii w całym obszarze referencyjnym. GTWCLQ jest natomiast nazwą procedury wyznaczania tej wielkości; składają się na nią wersja lokalna, czyli osobna wartość dla każdego zdarzenia (Cromley i in. 2014), oraz sąsiedztwo definiowane jednocześnie w przestrzeni i w czasie wraz z ważeniem obu wymiarów (Li i in. 2022). Wynikiem procedury GTWCLQ dla ustalonej pary skal – promienia  $r$  i okna czasowego  $\tau$  – jest wartość ilorazu kolokacji, zapisywana jako  $CLQ(r, \tau)$ . Stąd konwencja przyjęta w tej pracy: GTWCLQ występuje tam, gdzie mowa o metodzie, a CLQ tam, gdzie mowa o liczbie odczytywanej z tabeli lub z ryciny. Mapy na rycinach 5–8 są przypadkiem szczególnym tej samej procedury: promień wynosi 300 m, a sąsiedztwo nie jest ograniczane oknem czasowym, obejmuje więc cały okres objęty rejestrem, wobec czego wartość zapisujemy krócej jako CLQ.

Rycina 1 obrazuje te obliczenia na trzech prawdziwych zdarzeniach wybranych w celu ilustracji trzech możliwych rozkładów przestrzennych badanego zdarzenia na wybranym obszarze. W pierwszym otoczeniu występują 92 zdarzenia, z czego 10 to dziury w jezdni, czyli 10,9%; wobec 1,26% w mieście daje to  $CLQ = 8,6$ , a więc blisko dziewięciokrotnie większe zagęszczenie niż przy rozmieszczeniu przypadkowym. W drugim jest 86 zdarzeń i jedna dziura, czyli 1,2% – praktycznie tyle, ile wynosi udział miejski, więc  $CLQ = 0,9$  i skupienie nie występuje. W trzecim, choć zdarzeń jest tyle samo co w pierwszym (92), nie ma wśród nich ani jednej innej dziury, więc  $CLQ = 0$ .

Trzeci przypadek zasługuje na szczególną uwagę, bo bez niego łatwo o błąd w odczycie mapy, polegający na tym, że zdarzenie o  $CLQ = 0$  nie jest miejscem spokojnym: może występować w bardzo ruchliwym punkcie. Jest natomiast zdarzeniem **odosobnionym w swoim rodzaju** – takim, które w promieniu 300 m nie ma zdarzenia tego samego typu. Rozróżnienie „mało zdarzeń” i „mało zdarzeń tego rodzaju wobec obszaru referencyjnego” jest bardzo istotne.

Interpretację wyników przeprowadza się z wykorzystaniem poniższych informacji.

**Punktem referencyjnym na skali jest jedynka, nie zero.**  $CLQ = 1$  znaczy „tyle, ile przy rozmieszczeniu przypadkowym”. Wartości powyżej jedynki to



Ryc. 1. Skąd bierze się wartość ilorazu kolokacji – trzy rzeczywiste otoczenia dotyczące Warszawy (System Ewidencji Wypadków i Kolizji 2016–2024 – SEWIK, kategoria „najeżdżanie na dziurę”). Zielony punkt to zdarzenie, dla którego liczymy wartość, przerywany okrąg to jego otoczenie o promieniu 300 m, punkty ciemnoczerwone to inne najeżdżania na dziurę, szare – pozostałe zdarzenia drogowe, czyli tło ekspozycji ruchu. W każdym panelu podano kolejno: liczbę zdarzeń w otoczeniu, liczbę dziur wśród nich, ich udział, udział dziur w całym mieście i wynikający z tego iloraz. Panel A to skupienie (CLQ 8,6), panel B poziom losowy (CLQ 0,9), panel C brak skupienia (CLQ 0,0) mimo równie ruchliwego otoczenia jak w panelu A.

Zielony punkt to zdarzenie, dla którego liczymy wartość; przerywany okrąg to jego otoczenie o promieniu 300 m. Ciemnoczerwone punkty to inne najeżdżania na dziurę, szare – pozostałe zdarzenia drogowe (to ruchu).

CLQ powyżej 1: w otoczeniu tego zdarzenia dziur jest więcej niż wynika z ich udziału w całym mieście. CLQ = 1: tyle samo, co przy rozmieszczeniu losowym. CLQ = 0: w promieniu 300 m nie ma ani jednej dziury.

Źródło: opracowanie własne.

skupienie, poniżej — rozpraszanie się zdarzeń. Na wszystkich rycinach w tej pracy poziom losowy jest oznaczony osobno: na skalach ciągłych zieloną kreską, a na mapach zielonymi punktami.

**Iloraz nie jest prawdopodobieństwem.** CLQ = 2 nie znaczy „dwa razy większa szansa wypadku”, lecz „dwa razy większy udział tego rodzaju zdarzeń w otoczeniu niż wynika z jego udziału w zdarzeniach występujących w całym mieście”.

**GTWCLQ opisuje współwystępowanie, nie przyczynę i nie następstwo.** Okno czasowe jest symetryczne, to jest traktuje odstęp „przed” i „po” jednakowo. Wysoka wartość ilorazu jest przesłanką do sprawdzenia miejsca występowania zdarzeń w terenie.

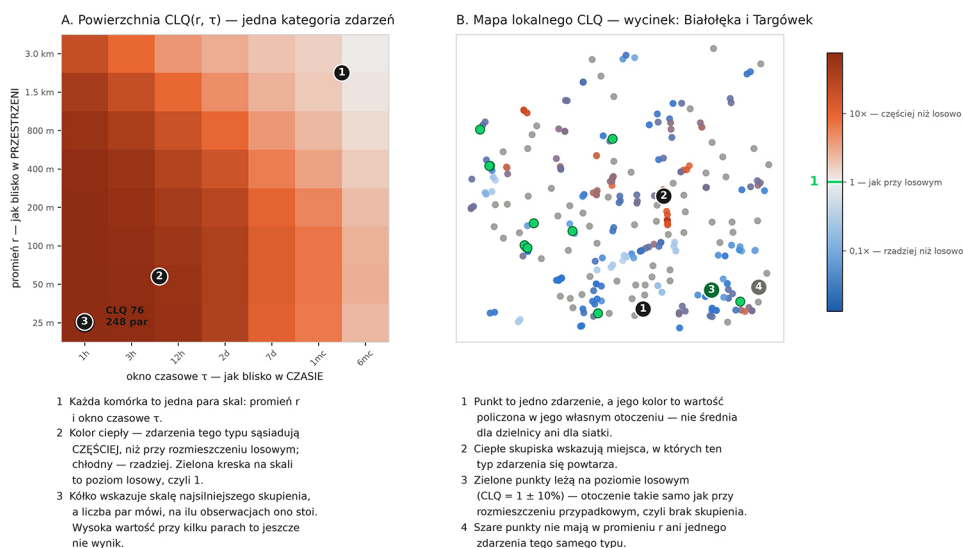
## Ilustracja graficzna i jej interpretacja

Ilustracją miary są dwa diagramy zaprezentowane na rycinie 2, zbudowane na podstawie danych z Warszawy. **Powierzchnia CLQ(r,  $\tau$ )** (panel A ryciny 2)

pozwała ustalić odpowiedź na pytanie o zasięg przestrzenny zachodzenia zdarzeń. Zamiast obliczać iloraz w jednym, z góry wybranym promieniu, obliczamy go dla całego zbioru promieni o określonych długościach (od 25 m do 3 km) i okien czasowych (od godziny do pół roku). Wynik jest tabelą, w której każda komórka to jedna para skal, oznaczająca w pionie „jak blisko w przestrzeni”, w poziomie „jak blisko w czasie”. Kolor ciepły oznacza skupienie, chłodny rozpraszanie się, a przejście między nimi występuje na poziomie losowym. Taka prezentacja zmienia sens pomiaru: skala przestaje być parametrem, który trzeba wybrać przed obliczeniami, i staje się wynikiem, bo z ryciny można wywnioskować, **w jakiej odległości i przy jakim interwale czasu** zjawisko jest najsilniejsze.

Na powierzchni umieszczamy także liczbę par zdarzeń w miejscach, gdzie występuje wartość maksymalna. Jest to informacja równie ważna jak sama wartość: iloraz policzony na kilkunastu parach bywa bardzo wysoki i zarazem bardzo niepewny, bo wystarczy jedno zdarzenie mniej, aby zmniejszył się o połowę. Czytelnik powinien analizować obie liczby razem – wysokie CLQ przy dużej liczbie par to wynik wiarygodny, a wysokie CLQ przy kilkunastu parach to silna przesłanka.

**Mapa CLQ** (panel B ryciny 2) odpowiada na pytanie o miejsce wystąpienia zdarzenia. Każdy punkt to jedno zdarzenie, a jego barwa to wartość policzona w jego własnym otoczeniu, tak jak na rycinie 1. Nie jest to zatem mapa uśredniona



Ryc. 2. Ilustracja graficzna miary dla danych warszawskich (najeżdżanie na dziurę, SEWIK 2016–2024). Panel A: powierzchnia  $CLQ(r, \tau)$ ; oś pionowa to promień sąsiedztwa, oś pozioma okno czasowe, kolor oznacza wartość ilorazu, a kółko wskazuje komórkę najsilniejszego skupienia wraz z liczbą par z nim związanych. Panel B: wycinek mapy CLQ dla północno-wschodniej części Warszawy (Białoleka i Targówek, 287 zdarzeń); barwa punktu to wartość w jego własnym otoczeniu. Jak interpretować ten układ: numerowane znaczki na obu panelach odpowiadają objaśnieniom pod nimi; zielona kreska na skali barwnej i zielone punkty na mapie oznaczają poziom losowy, czyli  $CLQ = 1$

Źródło: opracowanie własne.

według dzielnic ani według komórek siatki: rozdzielczość odpowiada rozdzielczości danych, a wartość dotyczy konkretnego punktu. Dwa rodzaje punktów wyróżniamy osobno, bo w ciągłej skali barwnej byłyby nie do odczytania. Punkty zielone to zdarzenia na poziomie losowym (CLQ w pasie  $1 \pm 10\%$ ), natomiast szare to zdarzenia o  $CLQ = 0$ , czyli bez ani jednego sąsiada tego samego rodzaju.

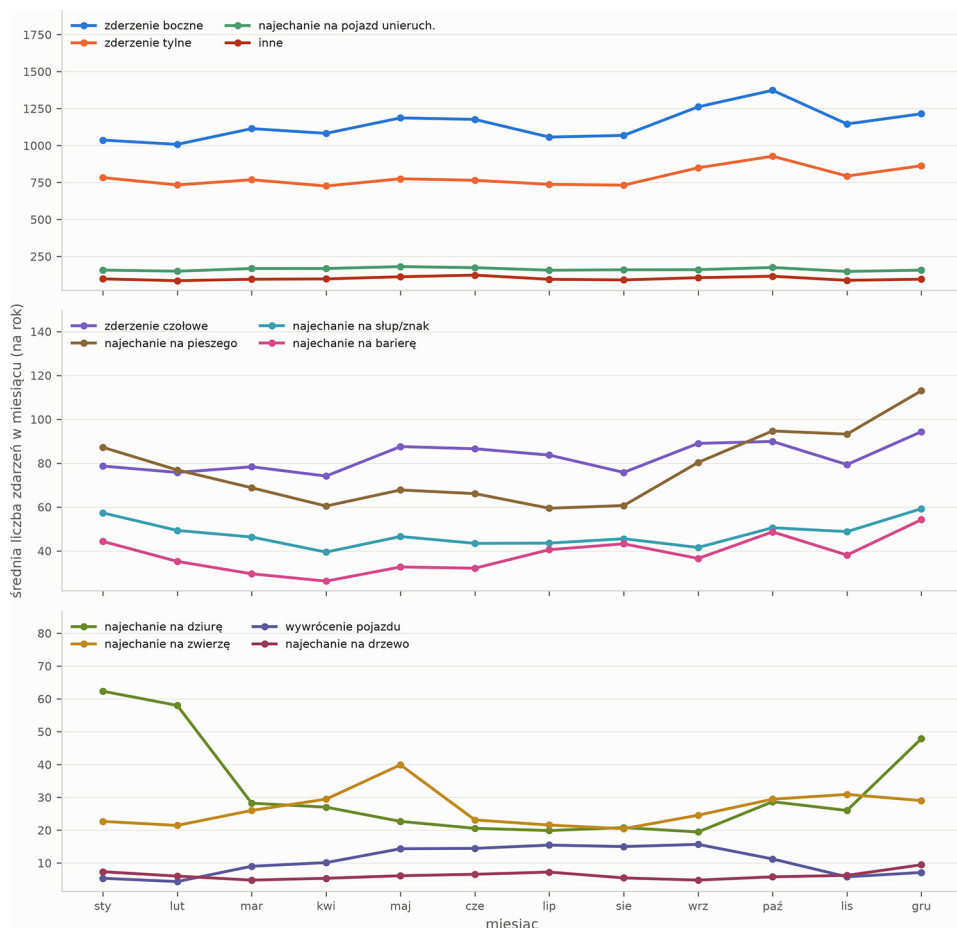
Wartość  $CLQ = 1$  na mapach praktycznie nie występuje i wynika to z algorytmu obliczenia CLQ, a nie z charakterystyki danych. Bowiem przy udziale zdarzeń rzędu procenta iloraz „liczba sąsiadów tego typu do liczby sąsiadów razem” przyjmuje wartości skokowe i jedynek zwykle pomija. Dlatego poziom losowy wyróżniamy jako pas  $1 \pm 10\%$ , a nie jako pojedynczą wartość.

## Dane wykorzystane w analizach zdarzeń

1. Głównym zasobem danych, dotyczących Warszawy w okresie 2016–2024, jest System Ewidencji Wypadków i Kolizji (SEWIK) Komendy Głównej Policji. Po ograniczeniu do zdarzeń o poprawnej geolokalizacji w granicach miasta zostaje 271 706 zdarzeń w dwunastu kategoriach o liczebności co najmniej 200. Zdarzeniom przyporządkowana jest ich lokalizacja wyrażona poprzez parę współrzędnych geograficznych miejsca ich wystąpienia. Zdarzeniom przyporządkowana jest ich lokalizacja wyrażona poprzez parę współrzędnych geograficznych miejsca ich wystąpienia. Czas zapisany jest z dokładnością do minuty, co pozwala budować pełną oś okien czasowych, od godziny do miesięcy.
2. Natomiast dane dotyczące wypadków drogowych w Chicago w 2023 r. pochodzą ze zbioru *traffic crashes* obejmującego 109 545 zdarzeń w piętnastu kategoriach o liczebności co najmniej 200, również z czasem występowania do jednej minuty. Trzy kategorie mają jednoznaczne odpowiedniki w SEWIK i tylko one wchodziły do porównań międzymiastowych: PEDESTRIAN (najeżdżanie na pieszego), REAR END (zderzenie tylne) oraz SIDESWIPE SAME DIRECTION (zderzenie boczne).

Jednostką analizy jest w obu zbiorach **pojedyncze zdarzenie**, a nie gmina, dzielnica czy komórka siatki. Ma to konsekwencję istotną dla badań geograficznych, wynik nie zależy od podziału administracyjnego, czyli nie podlega problemowi zmiennej jednostki odniesienia, a jednocześnie daje się później zagregować do dowolnego podziału.

Zanim przejdziemy do analizy przestrzennej, warto zwrócić uwagę na zmienność czasu, bowiem rytm roczny poszczególnych kategorii ma znaczenie dla interpretacji skupień. Rycina 3 przedstawia przeciętną liczbę zdarzeń w każdym miesiącu kalendarzowym. Najeżdżanie na dziurę ma wyraźny szczyt zimowy i wczesnowiosenny, to bezpośredni skutek cykli zamarzania i rozmrażania, które niszczą nawierzchnię, oraz zimowego utrzymania dróg. Najeżdżanie na zwierzę najczęściej zdarza się w maju, co jest zgodne z sezonem wzmożonej aktywności zwierząt i porą doby, w której ruch nakłada się na ich wędrówki. Potrącenia pieszych narastają w miesiącach późnojesiennych (listopad–grudzień) gdy godziny szczytu popołudniowego wypadają po zachodzie słońca. Zatem zdarzenia mogą mieć zróżnicowane rozkłady przestrzenne i czasowe.



Ryc. 3. Rozkład sezonowy rodzajów zdarzeń drogowych w Warszawie (2016–2024). Przeciętna liczba zdarzeń w każdym miesiącu kalendarzowym (średnia z dziewięciu lat), w trzech panelach według rzędu wielkości, aby kategorie masowe nie eliminowały zdarzeń sporadycznych. Każdy panel ma własną skalę pionową, więc porównywać można kształt krzywej, a nie wysokość między panelami; interesujące są maksima – zimowe dla najechania na dziurę, majowe dla najechania na zwierzę

Źródło: opracowanie własne.

## Wyniki analizy czaso-przestrzennej

### Identyfikacja długości promienia obszaru

Rycina 4 przedstawia powierzchnie CLQ( $r$ ,  $\tau$ ) dla czterech kategorii o odmiennej naturze, a tabela 1 zestawia najsilniejsze skupienie każdej z dwunastu kategorii warszawskich wraz z **liczbą par, na których ono bazuje**. Obraz jest wyraźny: kategorie różnią się nie tylko siłą skupienia, ale i skalą, w której ono występuje.

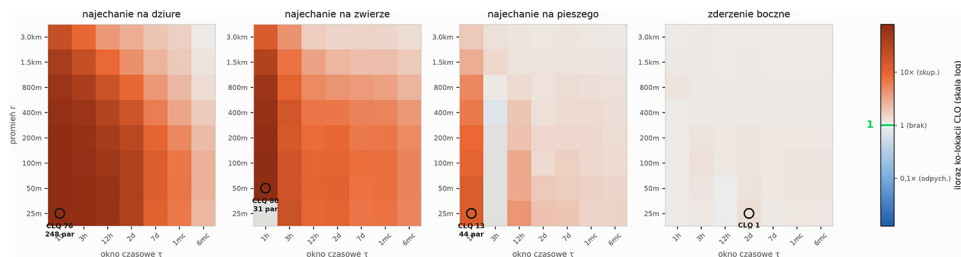
**Najeżdżanie na dziurę** skupia się punktowo i krótkoterminowo: największą wartość osiąga przy promieniu 25 m i oknie jednej godziny (CLQ 76,0), a obejmuje 248 par zdarzeń, czyli najliczniejszą podstawę w całym zestawie. Geograficznie znaczy to tyle, że mamy do czynienia ze zjawiskiem przywiązany do **jednego miejsca w jezdni**: kilkadziesiąt metrów to długość jednego uszkodzenia nawierzchni, a godzina to czas, w którym kolejne pojazdy wjeżdżają w to samo uszkodzenie. Rozmiar pozostaje duży także przy szerszych oknach, co znaczy, że defekt trwa i generuje kolejne zdarzenia przez dni i tygodnie, a nie tylko w ciągu jednego popołudnia.

**Najeżdżanie na pieszego** posiada maksimum również w skali mikro (25 m), lecz o wartości znacznie mniejszej (13,3) i w odniesieniu do 44 par. Wróćmy do tego w punkcie 4.3, w którym zaprezentowana jest mapa (ryc. 7).

**Kategorie środowiskowe** – najeżdżanie na zwierzę (CLQ 79,5 przy 50 m) i na drzewo (97,7 przy 800 m) – generują najwyższe ilorazy w całym zestawieniu, ale bazują na 31 i 33 parach. To zbyt mało, aby traktować wartość jako miarodajną. Są to jedynie poszlaki, wskazujące na powtarzanie się kolizji ze zwierzęciem lub w tym samym miejscu przy trasie wylotowej.

**Zderzenia boczne** nie koncentrują się wcale (CLQ 1,3). Jest to zjawisko rozłożone w przybliżeniu równomiernie po całej sieci ulic proporcjonalnie do ruchu. Kategoria ta jest w pracy pożyteczna w sensie referencyjnym, uświadamia bowiem, że wysokie wartości w przypadku innych kategorii nie są artefaktami sposobu liczenia.

Kolumny  $r$  i  $\tau$  podają promień i okno czasowe, przy których wartość jest największa; kolumna „par” to liczba par zdarzeń, na bazie których ta wartość jest ustalana, czyli miara jej wiarygodności. Kategorie uszeregowano według liczby par, a nie według CLQ, ponieważ wartość opartą na kilkadziesiąciu parach traktuje się inaczej niż opartą na kilkuset.



Ryc. 4. Skala samoskupienia czterech rodzajów zdarzeń – Warszawa 2016–2024. Powierzchnie CLQ( $r$ ,  $\tau$ ) na wspólnej skali barwnej; kółko oznacza komórkę najsilniejszego skupienia wraz z liczbą par, na których ta jest ustalana. Oś pionowa to promień sąsiedztwa, oś pozioma to okno czasowe, barwa ciepła to skupienie, chłodna rozchodzenie się, zielona kreska na skali to poziom losowy (CLQ = 1). Rozmiar ciepłego pola informuje, jak szeroko w przestrzeni i czasie utrzymuje się zjawisko: w przypadku dziur w jezdni pole jest duże i oparte na 248 parach, w przypadku zwierząt wąskie i oparte na 31 parach, w przypadku zderzenia bocznego nie ma go prawie wcale

Źródło: opracowanie własne.

Tabela 1. Najsilniejsze samoskupienie dwunastu rodzajów zdarzeń (Warszawa 2016–2024)

| Rodzaj zdarzenia                   | r     | $\tau$ | CLQ  | par        |
|------------------------------------|-------|--------|------|------------|
| najechnie na dziurę                | 25 m  | 1 h    | 76,0 | 248        |
| najechnie na pojazd unieruchomiony | 50 m  | 1 h    | 6,7  | 166        |
| inne                               | 25 m  | 1 h    | 9,2  | 120        |
| najechnie na barierę               | 25 m  | 1 h    | 26,2 | 120        |
| zderzenie czołowe                  | 25 m  | 1 h    | 8,4  | 80         |
| najechnie na pieszego              | 25 m  | 1 h    | 13,3 | 44         |
| wywrócenie pojazdu                 | 400 m | 1 h    | 40,5 | 35         |
| najechnie na drzewo                | 800 m | 1 h    | 97,7 | 33         |
| najechnie na słup lub znak         | 25 m  | 1 h    | 22,4 | 33         |
| najechnie na zwierzę               | 50 m  | 1 h    | 79,5 | 31         |
| zderzenie tylne                    | 25 m  | 1 h    | 2,2  | (podpróba) |
| zderzenie boczne                   | 25 m  | 48 h   | 1,3  | (podpróba) |

Źródło: opracowanie własne.

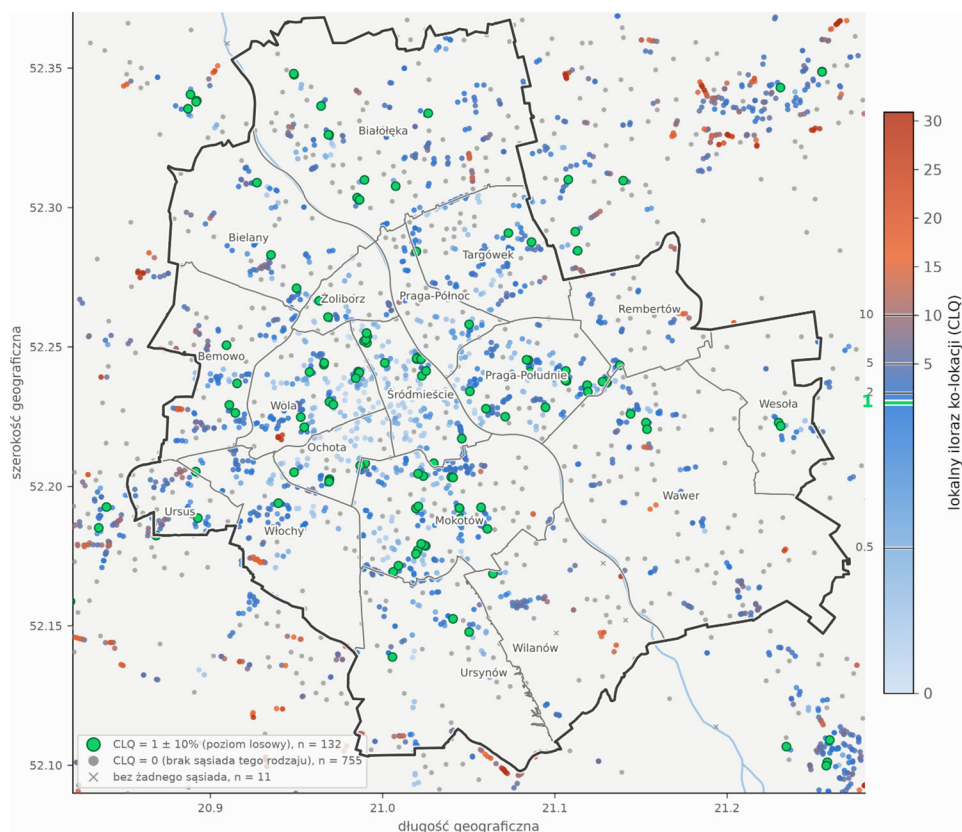
Dodatkowo należy wyjaśnić, dlaczego maksimum każdej kategorii wypadka przy najkrótszym oknie czasowym. Jest to własność miary, nie własność zjawiska. Pary zdarzeń bliskie w czasie są najrzadsze, więc ich nadwyżka nad wartością oczekiwaną daje najbardziej skrajny iloraz. Kategorie różnicuje zatem nie położenie wartości maksimum, lecz **wysokość, szerokość ciepłego pola** (jak daleko i długo w przestrzeni i czasie sygnał się utrzymuje) oraz **liczba par**. Tak właśnie należy interpretować rycinę 4.

### Lokalizacja skupień najechnań na dziurę na obszarze miasta

Rycina 5 przedstawia mapę lokalnego ilorazu kolokacji dla wszystkich 3433 najechnań na dziurę w Warszawie. Mediana wartości wynosi 1,37, dziewiąty decyl 9,94, a 41% zdarzeń leży w strefach o co najmniej dwukrotnym zagęszczeniu. Innymi słowy: dla dwóch piątych tych zdarzeń miejsce, w którym się wydarzyły, jest miejscem powtarzalnym.

**Rozkład tych stref nie pokrywa się z rozkładem ruchu.** Najwyraźniejsze skupienie występuje na północno-wschodnich obrzeżach Warszawy, tj. w Białołęce i dalej na wschód, a nie w śródmieściu, gdzie ruch jest największy. Jest to obszar intensywnej zabudowy mieszkaniowej w okresie ostatnich dwóch dekad. Sieć ulic powstawała etapami, a znaczna część jezdnii zajęta jest przez transport budowlany. Drugie skupienie tworzy pasmo wzdłuż tras wylotowych na wschód od Wisły. Śródmieście, przeciwnie, daje wartości niskie: nawierzchnie są tam remontowane częściej, a duża liczba wszystkich zdarzeń podnosi mianownik miary.

Do podobnego efektu przestrzennego prowadzi rytm roczny zobrazowany na rycinie 3: kategoria ta rośnie po zimie. Skupienia przestrzenne i sezonowość składają się w spójne wyjaśnienie geograficzne – w miejscach o niższym standardzie utrzymania nawierzchni intensyfikują zdarzenia wtedy, gdy nawierzchnia jest najbardziej narażona. Miara nie dowodzi tego mechanizmu, ale wskazuje miejsca, w których należy sprawdzić stan dróg; taka jest jej rola w praktyce.



Ryc. 5. Skupienia najechań na dziurę w Warszawie (2016–2024). Iloraz kolokacji obliczany jest w promieniu 300 m dla każdego z 3433 zdarzeń tej kategorii; tło stanowią granice 18 dzielnic i koryto Wisły (warstwy wektorowe OpenStreetMap, licencja ODbL). Barwa punktu to wartość w jego własnym otoczeniu, skala jest rozbieżna i wyśrodkowana na poziomie losowym  $CLQ = 1$ , oznaczonym na skali zieloną kreską. Zielone punkty to zdarzenia na poziomie losowym (pas  $1 \pm 10\%$ ,  $n = 132$ ), szare – zdarzenia o  $CLQ = 0$ , czyli bez sąsiadów tego samego rodzaju w promieniu 300 m ( $n = 755$ ). Mediana wartości 1,37, dziewiąty decyl 9,94, 41% zdarzeń w strefach o zagęszczeniu co najmniej dwukrotnym

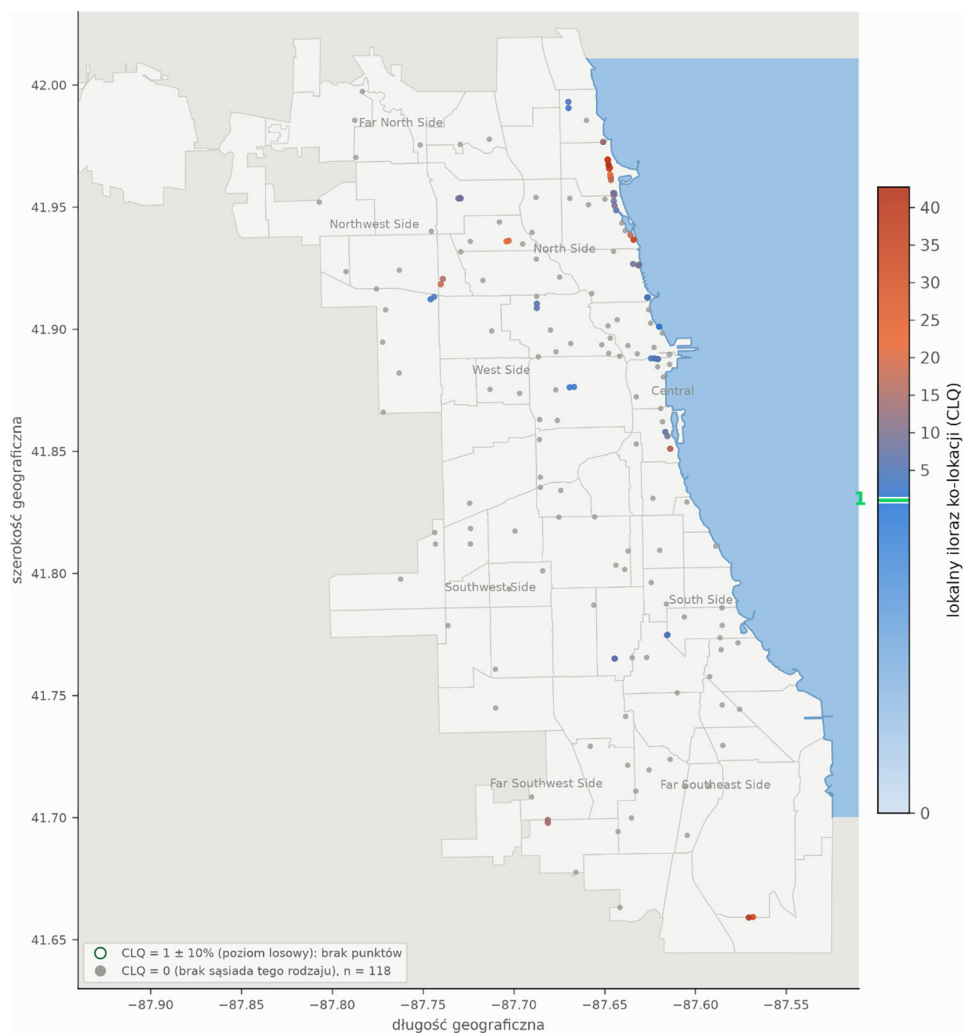
Źródło: opracowanie własne.

Ważne jest przy tym, to że skupienie **nie znika**, gdy weźmiemy pod uwagę samo rozmieszczenie czynników sprzyjających zdarzeniom. Sprawdzamy to, porównując zaobserwowaną wartość nie z rozmieszczeniem **całkowicie przypadkowym**, lecz z przypadkowym **wewnątrz obszarów o podobnym charakterze** (zabudowa i powiat; szczegóły w załączniku A). Wartości miary CLQ spadają wtedy z 22,1 do 13,1, ale pozostają wysokie. Informacja ta jest zatem czymś więcej niż tylko odbiciem tego, że część miasta jest gęściej zabudowana.

Chicago nie prowadzi rejestracji „najechania na dziurę” w katalogu pierwszego zderzenia, więc odpowiednika szukamy wśród przyczyn głównych: jest nim

kategoria „defekty nawierzchni i oznakowania” (ROAD ENGINEERING/SURFACE/MARKING DEFECTS). Kategoria ta jest znacznie rzadsza niż występowanie dziur w Warszawie: 182 zdarzenia, czyli 0,17% wobec 1,26%. Rycina 6 pokazuje ją na tle 77 jednostek podziału miasta.

Mapę tę czytamy inaczej niż mapę Warszawy, bowiem w Warszawie najechanie na dziurę interpretuje się jako kategorię zdarzenia, a w Chicago jako przyczynę.



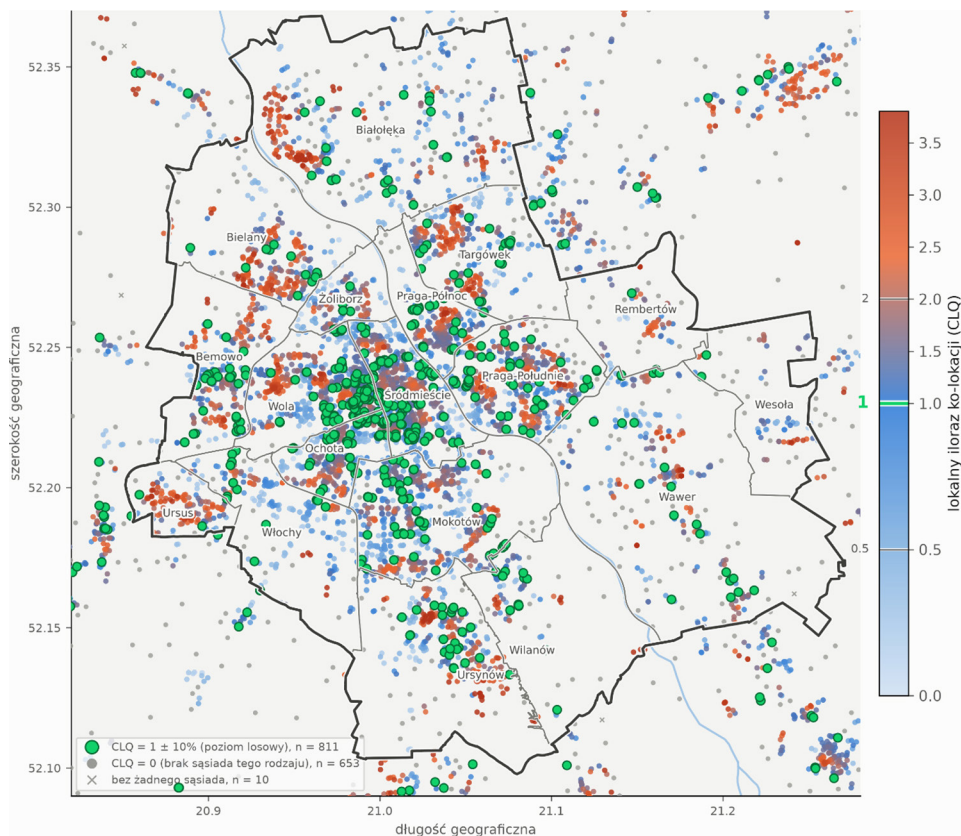
Ryc. 6. Zdarzenia z powodu defektu nawierzchni – Chicago 2023 (odpowiednik kategorii „najechanie na dziurę” w Warszawie. Przyczyna główna ROAD ENGINEERING/SURFACE/MARKING DEFECTS). CLQ w promieniu 300 m, oznaczenia jak na rycinie 5. Tło stanowią granice 77 obszarów społecznościowych (Chicago Community Areas – CHCA), granice dziewięciu dzielnic miasta oraz linia brzegowa jeziora Michigan (warstwy wektorowe OpenStreetMap, licencja ODbL)

Źródło: opracowanie własne.

## Potrącenia pieszych w Warszawie i Chicago

Najbardziej precyzyjny wynik geograficzny tej pracy dotyczy pieszych i powtarza się w obu miastach. Ryciny 7 i 8 przedstawiają lokalny CLQ dla 8357 potrąceń w Warszawie i 2726 w Chicago.

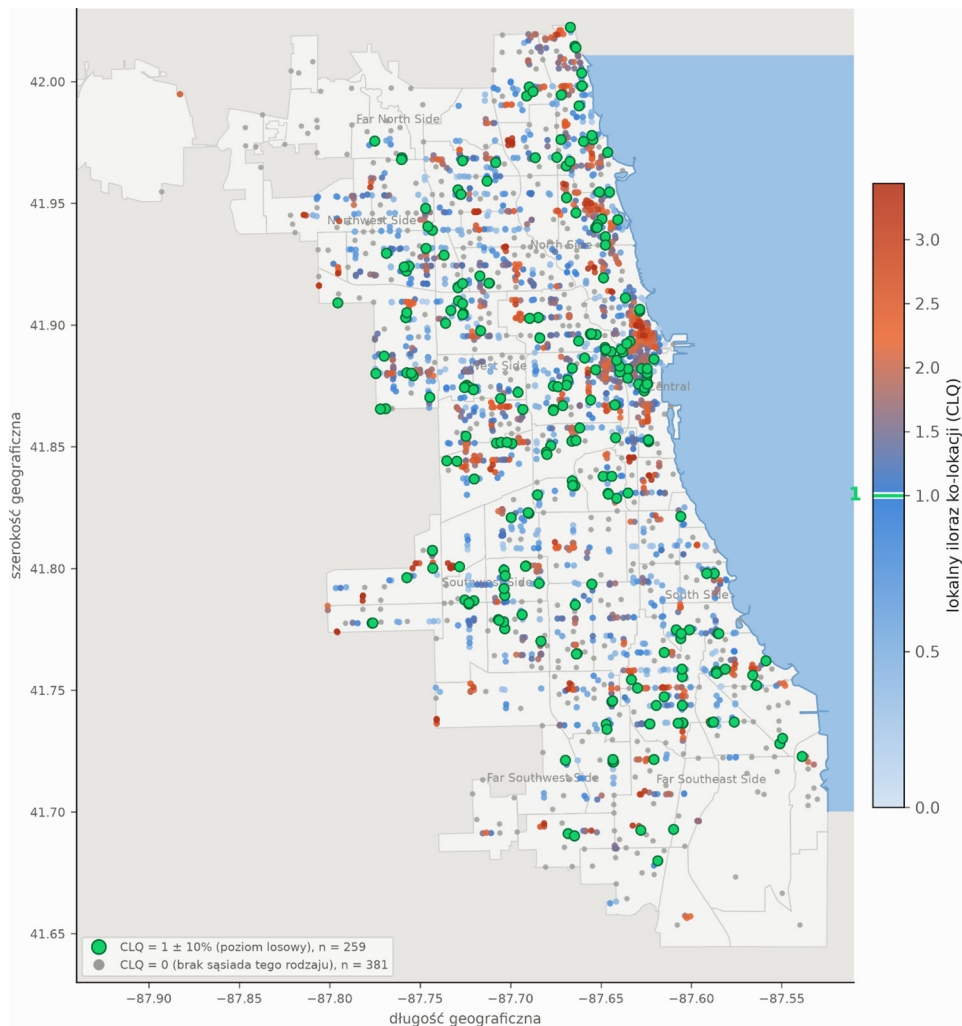
W Warszawie (rycina 7) rdzeń śródmieścia daje wartości bliskie poziomowi losowego lub niższe, mimo że bezwzględna liczba potrąceń jest tam największa. Nie jest to paradoks, lecz właśnie efekt odniesienia do kontekstu przestrzennego. W śródmieściu zabudowa i sieć ulic są gęste, więc udział potrąceń wśród zdarzeń nie odstaje od poziomu miejskiego. Skupienia przenoszą się na **arterie obrzebne** – wielopasmowe ciągi o wysokich prędkościach, przecinane przejściami



Ryc. 7. Skupienia potrąceń pieszych w Warszawie (2016–2024). CLQ w promieniu 300 m dla 8357 zdarzeń, oznaczenia jak na rycinie 5. Zielone punkty (poziom losowy, pas  $1 \pm 10\%$ ):  $n = 811$ ; szare (CLQ = 0):  $n = 653$ . Niskie wartości w rdzeniu śródmieścia nie znaczą, że jest tam bezpiecznie, lecz że udział potrąceń wśród wszystkich zdarzeń nie odstaje od miejskiego; skupienia leżą na arteriach obrzebnych. Mediana 1,27, dziewiąty decyl 2,55, 21% zdarzeń w strefach o zagęszczeniu co najmniej dwukrotnym (warstwy wektorowe OpenStreetMap, licencja ODbL)

Źródło: opracowanie własne.

obsługującymi dzielnice mieszkaniowe. Mediana wartości wynosi 1,27, dziewiąty decyl 2,55, a 21% zdarzeń leży w strefach o co najmniej dwukrotnym zagęszczeniu; rozkład jest więc znacznie bardziej wyrównany niż w przypadku dziury, dla której dziewiąty decyl większy jest od wartości 9.



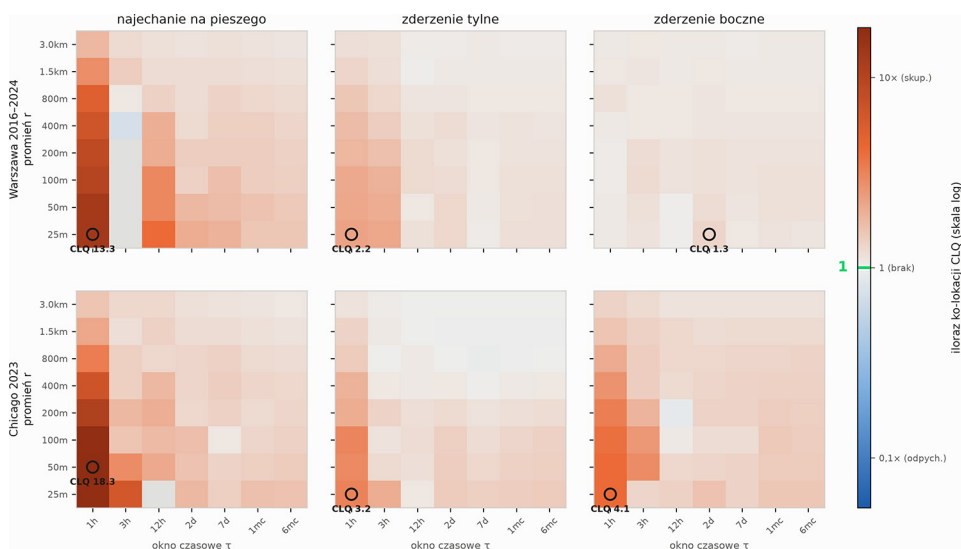
Ryc. 8. Skupienia potrażeń pieszych w Chicago 2023. CLQ w promieniu 300 m dla 2726 zdarzeń, oznaczenia jak na rycinie 5; tło stanowią granice 77 obszarów społeczno-sociowych (CHCA) i linia brzegowa jeziora Michigan. Zielone punkty:  $n = 259$ ; szare:  $n = 381$ . Mediana 1,10, dziewiąty decyl 2,32, 17% zdarzeń w strefach o zagęszczeniu co najmniej dwukrotnym. Jak interpretować ten układ – jest podobny do tego, który występuje w Warszawie. Czyli wyróżnia rdzeń (Central, Loop) na poziomie losowym. Natomiast skupienia występują na arteriach North Side i South Side (warstwy wektorowe OpenStreetMap, licencja ODbL)

Źródło: opracowanie własne.

W Chicago (ryc. 8) obraz jest taki sam: Central i Loop, czyli rdzeń o najwyższym natężeniu ruchu pieszego, dają wartości na poziomie losowym lub niższe, a najsilniejsze skupienia leżą na arteriach North Side i South Side. Statystyki opisowe są bliskie warszawskim (mediana 1,10, dziewiąty decyl 2,32, 17% zdarzeń w strefach o zagęszczeniu co najmniej dwukrotnym).

Powtarzalność tego układu w dwóch miastach o odmiennym planie zabudowy – promienistej ukształtowanej w Warszawie i prostokątnej siatce Chicago – jest wynikiem istotnym, ponieważ wskazuje ona, że w grę wchodzi nie plan ulic, lecz **relacja między prędkością ruchu a obecnością pieszego**. Skupienia powstają tam, gdzie piesi pojawiają się w otoczeniu przystosowanym dla ruchu szybkiego, a nie tam, gdzie pieszych jest najwięcej. Dla planowania jest to **wskazówka odwrotna do intuicyjnej: przeglądu wymagają nie najbardziej zatłoczone ulice śródmieścia, lecz przejścia na obwodowych arteriach**.

Zgodność zdarzeń w obu miastach uwidoczniła się także w postaci skali, w której zjawisko występuje. Rycina 9 zestawia powierzchnie CLQ( $r$ ,  $\tau$ ) dla trzech rodzajów zdarzeń obecnych w obu zbiorach, na wspólnej skali barwnej. Sygnatura potrącenia pieszego powtarza się: skupienie na skali mikro, od 25 do 100 m. Liczone w tym samym polu (25 m, okno godzinne z wykorzystaniem kompletu zdarzeń) CLQ pieszego wynosi 13,3 w Warszawie i 17,3 w Chicago. Nieco wyższa wartość w Chicago odpowiada jego ściślejszej, prostokątnej siatce ulic, w której przejścia są bardziej regularnie rozłożone. Zderzenie tylne rozkłada



Ryc. 9. Ten sam pomiar w dwóch miastach w Warszawie i Chicago. Powierzchnie CLQ( $r$ ,  $\tau$ ) dla trzech rodzajów zdarzeń obecnych w obu zbiorach (górny wiersz Warszawa 2016–2024, dolny Chicago 2023), na wspólnej skali barwnej. Barwy są porównywalne między wszystkimi sześcioma panelami, bo skala jest wspólna; zielona kreska na pasku oznacza losowy charakter współwystępowania zdarzeń. Zgodny kształt „ciepłego” pola w kolumnie oznacza, że dany rodzaj zdarzenia zachowuje się tak samo w obu miastach

Źródło: opracowanie własne.

się przestrzennie zgodnie w obu miastach. Rozbieżność dotyczy zderzenia boczno – kolorystycznie bladego w Warszawie i wyraźniejszego w Chicago. Może to wynikać z odmiennej geometrii skrzyżowań – prostokątna siatka daje więcej skrzyżowań równorzędnych na tej samej długości ulic.

## Problem symetrii pomiędzy kategoriami zdarzeń

Dotychczasowa analiza dotyczyła ustalenia odpowiedzi na pytanie: czy zdarzenia drogowe, należące do tej samej kategorii wykazują tendencję do współwystępowania w bliskim sąsiedztwie przestrzennym?

Jednak GTWCLQ pozwala również ustalać relacje przestrzenne pomiędzy zdarzeniami należącymi do różnej kategorii – łącznie z identyfikacją kierunków tych oddziaływań.

Tabela 2 zawiera zestawienie takich relacji pomiędzy zdarzeniami drogowymi dla Warszawy w okresie 2016–2024. Szczególnie interesujące są wartości liczbowe w kolumnie 5 (iloraz). Występują w niej bowiem wartości liczbowe, które informują, ile razy CLQ  $A \rightarrow B$  przewyższają wartości CLQ  $B \rightarrow A$ . Czyli pozwalają ustalić stopień asymetrii wzajemnego położenia przestrzeni różnych zdarzeń drogowych.

Tabela 2. Jednostronność relacji między rodzajami zdarzeń (Warszawa, promień 400 m, okno 48 h)

| Rodzaj A                    | Rodzaj B                            | CLQ $A \rightarrow B$ | CLQ $B \rightarrow A$ | iloraz | p-war. |
|-----------------------------|-------------------------------------|-----------------------|-----------------------|--------|--------|
| 1                           | 2                                   | 3                     | 4                     | 5      | 6      |
| najechanie na drzewo        | zderzenie tylne                     | 1,00                  | 0,24                  | 4,2×   | 0,001  |
| najechanie na drzewo        | zderzenie boczne                    | 0,84                  | 0,32                  | 2,6×   | 0,005  |
| najechanie na zwierzę       | zderzenie boczne                    | 0,77                  | 0,38                  | 2,0×   | 0,001  |
| najechanie na słup lub znak | najechanie na pojazd unieruchomiony | 1,08                  | 0,73                  | 1,5×   | 0,048  |
| wywrócenie pojazdu          | zderzenie tylne                     | 0,89                  | 0,89                  | 1,0×   | 0,982  |

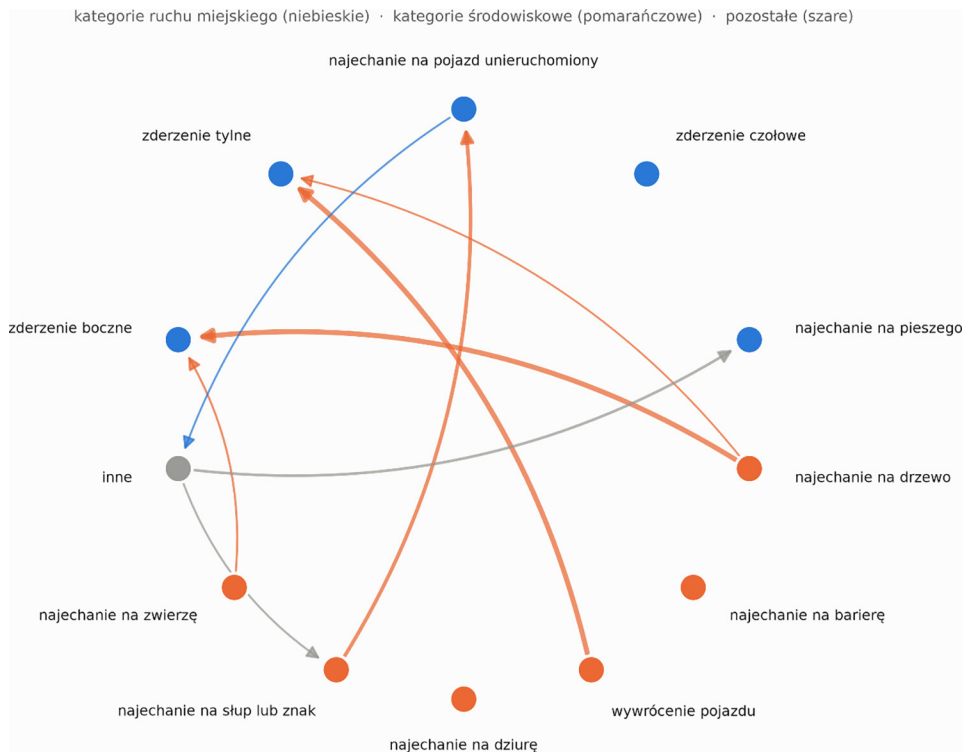
Zwraca uwagę jednostronność relacji pomiędzy różnymi kategoriami analizowanych zdarzeń, gdyż we wszystkich przypadkach zdarzenia typu A występują częściej w otoczeniu zdarzenia B, aniżeli B w otoczeniu A.

W tej samej tabeli kolumna 6 zawiera p-wartości. Statystycznie potwierdzają one występujące asymetrie przestrzenne, dotyczące kolokacji zdarzeń drogowych A i B.

Dokładniejsza analiza pozwala stwierdzić, że trzy pierwsze pary zdarzeń A i B w tabeli 2 mają wyraźnie jednostronną relację i dotyczą zdarzeń o charakterze środowiskowym (drzewo, zwierzę) względem zdarzeń typowo miejskich (zderzenie tylne, zderzenie boczne).

Interpretacja geograficzna jest następująca: najechania na drzewo i na zwierzę zdarzają się w przestrzeni, gdzie zdarzenia miejskie występują w normalnej dla miasta proporcji.

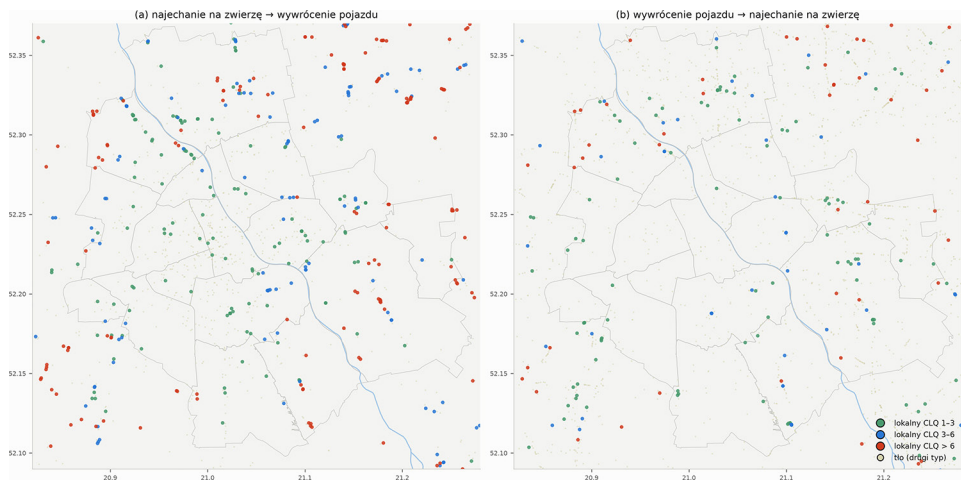
Natomiast w przypadku zderzeń tylnych i bocznych obiekty: drzewo i zwierzę pojawiają się rzadziej niż przeciętnie. Zatem miara CLQ nie identyfikuje mechanizmu wypadków, lecz asymetrię zasięgów dwóch kategorii zjawisk.



Ryc. 10. Jednostronność relacji między rodzajami zdarzeń – Warszawa 2016–2024, promień 400 m, okno 48 h. Strzałka biegnie w kierunku silniejszej relacji, a jej grubość odpowiada ilorazowi obu kierunków. Pokazano pary o ilorazie co najmniej  $1,8\times$ . Strzałka od A do B znaczy, że w otoczeniu zdarzeń rodzaju A zdarzenia rodzaju B występują względnie częściej niż odwrotnie, a nie że A powoduje B; brak strzałki między dwoma kategoriami znaczy, że relacja jest obustronnie podobna

Źródło: opracowanie własne.

Rycina 11 zestawia dwa kierunki relacji dla pary najeżdżanie na zwierzę–wywrócenie pojazdu, czyli pary najsilniej kolokowanej przy małych promieniach. Oba panele obejmują ten sam obszar i te same dwie kategorie zdarzeń, różnią się tylko tym, z jakiej perspektywy rozpatrujemy te zdarzenia. Skupienia w obu przypadkach wypadają na obrzeżach i trasach wylotowych, zgodnie z niemiejskim charakterem tych zdarzeń, ale ich gęstość i natężenie nie są takie same i obrazują różnicę kolokacji, stwierdzone w tabeli 2.



Ryc. 11. Dwa kierunki tej samej relacji na mapie: najechanie na zwierzę i wywrócenie pojazdu – Warszawa 2016–2024 (warstwy wektorowe OpenStreetMap, licencja ODbL). Barwa punktów odpowiada wartościom ilorazu kolokacji (promień 300 m) liczonego na zdarzeniach kategorii nadawcy, tło stanowią zdarzenia kategorii odbiorcy. Jak interpretować te ryciny: panel (a) odpowiada na pytanie „ile wywróceń występuje w otoczeniu najechań na zwierzę”, panel (b) na pytanie odwrotne. Różnica gęstości i natężenia między panelami obrazuje jednostronność relacji przedstawionej w tabeli 2

Źródło: opracowanie własne.

## Interpretacja i wnioski

Skala jest cechą zjawiska, nie obrazowaniem analizy, ponieważ najechanie na dziurę jest zjawiskiem w przestrzeni kilkudziesięciu metrów, a potrącenie pieszego obejmuje ciągi komunikacyjne, kategorie środowiskowe cechuje rozciągłość kilometrowa. Wynika stąd praktyczny wniosek: analiza prowadzona przy uwzględnieniu jednego, z góry wybranego promienia lub w jednej jednostce podziału administracyjnego ujawni tylko te zjawiska, których skala jest zbliżona do wybranej, a pozostałe uśredni. Zastosowana w tej pracy powierzchnia  $CLQ(r, \tau)$  pozwala tej sytuacji uniknąć, bo skalę odczytuje się z danych.

**Odniesienie do kontekstu przestrzennego zmienia mapę problemów.** Największa liczba zdarzeń i największe zagęszczenie zdarzeń danej kategorii to dwa różne zjawiska. To drugie wskazuje miejsca wymagające interwencji. W obu badanych miastach **śródmieście**, mimo największej liczby potrąceń, nie jest obszarem podwyższonego zagęszczenia potrąceń, a są nimi **arterie obrzeżne**. Informacja taka byłaby niedostępna w przypadku analizy operującej na liczbach bezwzględnych lub na gęstości zdarzeń na kilometr.

**Powtarzalność między miastami wzmacnia siłę wnioskowania.** Ten sam układ potrąceń pieszych w Warszawie i Chicago przy odmiennym planie ulic sugeruje, że ustalamy relację „prędkość ruchu wobec obecności pieszego”, a nie właściwości jednego zbioru danych. Zestawienie pakietów danych dla dwóch

miast jest przy tym niskokosztowe, bo wymaga tylko rejestru punktowego z etykietą kategorii i znacznika czasu. Takie rejestry są coraz szerzej dostępne.

**Zjawiska o różnym zasięgu wykazują relacje jednostronne.** Kategorie śródowniskowe i miejskie nie są względem siebie symetryczne, ponieważ zajmują różne podzbiory przestrzeni miejskiej. Jednostronność jest mierzalna, a jej pomiar chroni od nadinterpretacji, bo identyfikuje również te pary zdarzeń, w których różnica kierunków jest pozorna.

Do zastosowania analizowanej miary potrzebny jest zbiór punktów z trzema charakterystykami: położeniem, czasem i etykietą kategorii zdarzenia. Wskazane jest, aby dane spełniały cztery warunki.

1. położenie musi być rzeczywiste, a nie zastąpione centroidem jednostki administracyjnej;
2. katalog rodzajów powinien być rozłączny i stabilny w czasie; zmiana definicji kategorii w połowie okresu daje sztuczne skupienie czasowe;
3. kategoria musi być dostatecznie liczna; przy udziale poniżej kilku dziesiątych procenta większość zdarzeń nie ma w otoczeniu żadnego zdarzenia tego samego rodzaju i mapa staje się wskazaniem miejsc, a nie polem zagęszczenia, co występowało w przypadku Chicago;
4. warto zapoznać się z procedurą powstawania rejestru, bo różnice w procedurze zgłoszeniowej między miastami są zwykle większe niż różnice geograficzne.

W interpretacji wyników pomocne są cztery pytania, które powinno się zadać każdej wartości ilorazu: na jakiej skali została policzona, na ilu parach zdarzeń występuje, jak wygląda wobec rozmieszczenia przypadkowego oraz czy nie jest odbiciem samej struktury okazji, bowiem część obszaru jest gęściej zabudowana. Sposób odpowiadania na dwa ostatnie pytania opisuje załącznik A.

Należy wyraźnie podkreślić, że miara GTWCLQ opisuje **współwystępowanie** zdarzeń, a **nie zależność przyczynową**. Okno czasowe jest symetryczne, czyli nie porządkuje zdarzeń w czasie. Wartości dla kategorii rzadkich, bazujących na kilkudziesięciu parach, traktujemy jako wskazania wymagające potwierdzenia na większym materiale obserwacyjnym, na przykład na danych krajowych. Wynik zależy też od definicji obszarów uznanych za jednorodne. Wpływ tej definicji jest niewielki dla kategorii licznych i większy dla rzadkich. Wreszcie porównania międzymiastowe ograniczają się do rodzajów zdarzeń o jednoznacznym odpowiedniku w obu katalogach, co dla Warszawy i Chicago daje trzy z kilkunastu kategorii.

## Konflikt interesów / wkład autorów

Autorzy deklarują brak konfliktu interesów. Oświadczają, że tekst artykułu jest w całości ich dziełem. Podział pracy przy artykule wyglądał następująco:

- konceptualizacja: MK, WR, M (Marcin) K,
- metodologia: MK, WR, M (Marcin) K,
- organizacja badań: MK, WR, M (Marcin) K, AK,
- analiza formalna: MK, WR,
- pisanie: MK, WR, M (Marcin) K, AK.

## Dostępność danych i kodu

Dane Warszawy pochodzą z systemu SEWIK, dane Chicago z miejskiego portalu danych otwartych (zbiory podane w sekcji „Źródła danych”). Kody obliczeń są dostępne u autorów.

## Literatura / References

- Benjamini Y., Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, B*, 57(1): 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Cliff A.D., Ord J.K. 1981. *Spatial processes: models and applications*. Pion, London.
- Cromley R.G., Hanink D.M., Bentley G.C. 2014. Geographically weighted colocation quotients: specification and application. *The Professional Geographer*, 66(1): 138–148. <https://doi.org/10.1080/00330124.2013.768130>
- Leslie T.F., Kronenfeld B.J. 2011. The colocation quotient: a new measure of spatial association between categorical subsets of points. *Geographical Analysis*, 43(3): 306–326. <https://doi.org/10.1111/j.1538-4632.2011.00821.x>
- Li L., Cheng J., Bannister J., Mai X. 2022. Geographically and temporally weighted co-location quotient: an analysis of spatiotemporal crime patterns in greater Manchester. *International Journal of Geographical Information Science*. <https://doi.org/10.1080/13658816.2022.2029454>
- Li L., Cheng J., Mai X., Duan L., Mo L., Tan J. 2026. Spatiotemporal association analysis of incident data: a multi-temporal framework based on Geographically and Temporally Weighted Co-Location Quotient. *Geo-spatial Information Science*. <https://doi.org/10.1080/10095020.2026.2612768>

## Źródła danych

- SEWIK – System Ewidencji Wypadków i Kolizji, Komenda Główna Policji. Zdarzenia drogowe 2016–2024.
- City of Chicago Open Data. 2024. Traffic crashes – crashes (85ca-t3if) oraz Boundaries – community areas (igwz-8jzy). <https://data.cityofchicago.org/>

## The geography of road-crash co-location: space-time association of events in Warsaw and Chicago

**Abstract:** Road crashes of different kinds are not distributed independently of one another: some concentrate in the same places and within a short time of each other, while others clearly avoid one another. The tool for measuring such association is the co-location quotient (CLQ) in its space-time form (GTWCLQ), which compares the neighbourhood of one kind of event with what would be expected under a random arrangement. This paper is applied and didactic: it shows what the measure says about urban space, how to read its results and how not to over-interpret them. The material consists of all road events recorded in Warsaw between 2016 and 2024 (271,706 events, SEWIK register) and in Chicago in 2023 (109,545 events, municipal open-data portal). Three groups of results emerge. First, the scale of association is a property of the kind of event rather than of the city: potholes cluster at the scale of tens of metres and hours, pedestrian crashes along arterial corridors, and sideswipe collisions hardly at all. Second, the spatial pattern of pedestrian crashes is the same in both cities: the dense core yields values at or below the random level, while pronounced clusters lie on peripheral arterial roads. Third, the relation between two kinds of events can be one-sided: environmental categories such as animal and tree collisions occupy a subset of urban space, so urban collision types appear in their neighbourhood in ordinary proportion while the converse does not

hold. This one-sidedness can be measured and tested, and the test also rejects one pair that looked one-sided on the raw values alone. The computation, including the improvements introduced here – the full scale surface instead of a single bandwidth, a reference distribution obtained by permuting type labels within blocks of similar terrain, and an explicit test of directional difference – is set out in plain terms in Appendix A, together with a twelve-item list of what to report when applying the measure to one's own data.

**Keywords:** co-location quotient, road crashes, road-safety geography, spatial clusters, Warsaw, Chicago

**Data przekazania tekstu:** 7.09.2026; **data zaakceptowania tekstu:** 21.09.2026.

**Article submitted:** 7.09.2026; **article accepted:** 21.09.2026.

## Załącznik A. Procedura obliczania ilorazu kolokacji oraz jej modyfikacja

Załącznik ten jest zwięzłym wykładem, na tyle dokładnym, aby obliczenia dały się powtórzyć, i pisany z myślą o czytelniku, który nie zajmuje się statystyką.

### A.1. Miara podstawowa

Punktem wyjścia jest iloraz kolokacji dla dwóch rodzajów zdarzeń, A i B (Leslie, Kronenfeld 2011). Dla każdego zdarzenia rodzaju A ustalamy jego sąsiedztwo i wyznaczamy, jaki udział mają w nim zdarzenia rodzaju B. Udziały te uśredniamy po wszystkich zdarzeniach rodzaju A i dzielimy przez udział rodzaju B w całej zbiorowości zdarzeń:

Mianownik jest tym, co czyni miarę użyteczną w mieście: zawiera rozkład okazji, więc wynik odpowiada na pytanie o nadwyżkę zdarzeń w zbiorowości ogółem, a nie o liczbę zdarzeń. Wartość równa 1 oznacza brak nadwyżki zdarzeń, powyżej 1 – skupienie, poniżej – rozpraszanie się zdarzeń. Dla  $A = B$  miara opisuje samo-skupienie jednego rodzaju i tak używamy jej w punktach 4.1–4.3.

Miara jest kierunkowa, ponieważ liczba zdarzeń każdej kategorii jest włączana do obliczeń na innym etapie:  $CLQ(A \rightarrow B)$  i  $CLQ(B \rightarrow A)$  to dwie różne liczby i porównanie ich jest osobnym pytaniem badawczym (punkt 4.4).

W wersji lokalnej (Cromley i in. 2014) nie uśredniamy według zdarzeń kategorii A, lecz ustalamy wartość dla każdego zdarzenia osobno. Otrzymujemy wówczas tyle liczb, ile jest zdarzeń danego kategorii i taką postać mają mapy na rycinach 5–8. Obliczenia dla jednego zdarzenia pokazuje rycina 1.

W wersji czasoprzestrzennej sąsiedztwo definiuje się jednocześnie w przestrzeni i w czasie, sąsiadem zdarzenia jest zdarzenie występujące w promieniu  $r$  i różniące się czasem nie więcej niż o wartość  $\tau$ . To rozszerzenie, wraz z wersją lokalną, nosi nazwę geograficznie i czasowo ważonego ilorazu kolokacji (GTWCLQ).

### A.2. Identyfikacja parametrów obszaru

Etap dotyczy wyznaczania sąsiedztwa. Zamiast wybierać jeden promień i jedno okno czasowe, liczymy iloraz dla całej siatki (zbioru) promieni  $m$  oraz okien  $h$ .

Sąsiedztwo jest przy tym proste do opisania: para zdarzeń jest włączona do obliczeń z wagą 1, jeśli mieści się w promieniu i w ustalonym oknie, a poza nimi z wagą 0.

Korzyść jest podwójna. Po pierwsze, znika parametr, którego wartość trzeba przyjąć przed obliczeniami, a od którego zależy wynik końcowy. Skala staje się zmienną niezależną wyniku, tak jak promień w funkcji K Ripleya. Po drugie, można postawić pytanie, którego przy jednym promieniu postawić nie można: w jakiej skali zjawisko jest najsilniejsze. Odpowiedź jest wynikiem merytorycznym i różni się między rodzajami zdarzeń – według pomiarów empirycznych w Chicago zderzenie boczne osiąga maksimum przy 167 m, co odpowiada odstępowi między skrzyżowaniami, a potrącenie pieszego przy 1833 m – i jest to związane z długością ciągu komunikacyjnego.

Oś przestrzenna jest zawsze skumulowana (wszystkie pary do danego promienia), a oś czasowa liczona jest w dwóch trybach: skumulowanym, gdy sprawdzamy istotność, oraz przyrostowym, gdy ustalamy skalę. Rozróżnienie to jest konieczne, bo pewność wyniku rośnie wraz z liczbą par, więc przy coraz szerszych oknach rośnie zawsze; skalę czasową odczytuje się natomiast z samej wartości ilorazu.

### A.3. Porównanie wyników analizy

Aby stwierdzić, czy dana wartość jest większa, niż mogłaby występować losowo, potrzebny jest punkt odniesienia. Budujemy go przez permutację etykiet kategorii zdarzenia przy zachowanych lokalizacjach i czasach, zdarzenia zostają tam, gdzie wystąpiły, a losowo zmieniamy tylko to, jakiej są kategorii. Powtarzamy to wielokrotnie (299 przy przeglądzie całej siatki, 999 przy testach kluczowych) i otrzymujemy rozkład wartości, których należałoby oczekiwać przy braku związku między kategorią zdarzenia a jego położeniem. Zaobserwowaną wartość odnosimy do średniej i rozrzutu tego rozkładu; dzięki temu wynik nie zależy od liczby powtórzeń symulacji, a jedynie od rangi informacji pochodzącej z danych.

Drugie ulepszenie polega na tym, że permutację prowadzimy w dwóch wariantach. W wariacie globalnym etykiety mogą trafić w dowolne miejsce miasta. W wariacie warunkowym permutujemy je tylko wewnątrz bloków wyznaczonych przez charakter terenu (zabudowany lub nie) oraz podział na jednostki przestrzenne. Bloki o liczebności poniżej 50 zdarzeń łączymy w jeden blok reszty, co dla Warszawy daje 26 bloków. Różnica między obu wariantami ma sens geograficzny: jest to ta część współwystępowania, którą wyjaśnia samo rozmieszczenie zdarzeń w strukturze osadniczej i administracyjnej. Jeśli skupienie należy do wariantu warunkowego, jest czymś więcej niż odbiciem tej struktury – tak jest w wypadku zdarzenia najeżdżania na dziurę (punkt 4.2).

### A.4. Sprawdzanie hipotezy o jednostronności relacji

Aby orzec, że relacja między kategoriami zdarzeń jest jednostronna, nie wystarczy porównanie dwóch wartości. Różnicę  $D = CLQ(A \rightarrow B) - CLQ(B \rightarrow A)$  odnosimy do jej rozkładu przy zastosowaniu permutacji etykiet (999 powtórzeń, oba warianty tasowania) i tylko istotną różnicę zapisujemy jako strzałkę w grafie

(ryc. 10). Dla kategorii rzadkich liczymy przy tym wszystkie zdarzenia danego rodzaju, a nie ich próbę, aby wynik nie zależał od losowania. Test ten odrzucił jedną z par, która przy oglądzie samych wartości wyglądała na jednostronną (tab. 2, wiersz ostatni).

#### **A.5. Wielokrotne sprawdzanie i inne szczegóły procesu obliczeń**

Przegląd całej siatki skal i wszystkich par kategorii zdarzeń oznacza wiele jednoczesnych weryfikacji, a przy wielu sprawdzeniach część wyników ma charakter skrajny. Stosujemy więc kontrolę współczynnika fałszywych ustaleń metodą Benjaminiego–Hochberga (1995) na całej siatce, a w odczycie krzywej wzdłuż promienia posługujemy się obwiednią obowiązującą dla całej krzywej naraz, a nie dla pojedynczego punktu. Kontrola jest wymagająca – np. w Chicago przechodzi ją 89 z 225 komórek tablicy kategorii, wobec 154 przy obwiedni punktowej – i właśnie dlatego orzeczenia, które ją przechodzą, są wiarygodne.

Dwie decyzje obliczeniowe należy uwypuklić, gdyż wpływają na wynik. Zdarzenie nie jest własnym sąsiadem, a usuwamy je z sąsiedztwa po identyfikatory rekordu, nie po współrzędnej, ponieważ różne zdarzenia dzielą tę samą lokalizację; w kategoriach rzadkich zaniechanie tego zawyżałoby samoskupienie wielokrotnie. Powierzchnie  $CLQ(r, \tau)$  i ich rozkłady odniesienia liczymy z próby 6000 zdarzeń docelowych (losowanie jednorodne, ustalone ziarno), ze względu na ogólnie rozumiany koszt zestawiania par w promieniu 3 km; wartości w tabeli 1 oraz wszystkie mapy lokalne liczone są jednak na komplecie zdarzeń danej kategorii.

Dla dużych zbiorów przygotowuje się również postać zamkniętą wartości oczekiwanej i rozrzutu ilorazu, pozwalającą pominąć symulację; w niniejszej pracy wszystkie wyniki podajemy jako efekt permutacji.

#### **A.6. Pojęcia i definicje stosowane w procedurze obliczania GTWCLQ**

Aby wynik mógł być powtórzony i porównany z innymi, w opisie własnego zastosowania należy uwzględnić dwanaście elementów, zestawionych w tabeli A1. Lista wynika z konstrukcji miary.

Tabela A1. Co należy podać, opisując zastosowanie GTWCLQ, wraz z rozwiązaniem przyjętym w tej pracy

| Nr | Element opisu                    | Zastosowanie w niniejszej pracy                                                                                               |
|----|----------------------------------|-------------------------------------------------------------------------------------------------------------------------------|
| 1  | definicja sąsiedztwa             | promień i okno czasowe, wynik podany dla całej siatki skal (A.2)                                                              |
| 2  | postać wagi w przestrzeni        | waga 1 wewnątrz promienia, 0 poza nim                                                                                         |
| 3  | postać wagi w czasie             | waga 1 wewnątrz okna, 0 poza nim                                                                                              |
| 4  | zasięg przestrzenny              | osiem promieni od 25 m do 3 km, oś skumulowana                                                                                |
| 5  | zasięg czasowy                   | siedem okien od 1 h do pół roku, oś skumulowana i przyrostowa                                                                 |
| 6  | sposób normalizacji              | udział rodzaju w całej zbiorowości zdarzeń, z korektą na wykluczenie zdarzenia                                                |
| 7  | punkt odniesienia                | rozkład z permutacji etykiet rodzaju, w wariancie globalnym i warunkowym (A.3)                                                |
| 8  | model permutacji                 | etykiety rodzaju przy zachowanych położeniach i momentach czasu; warunkowo w blokach zabudowa $\times$ jednostki przestrzenne |
| 9  | liczba powtórzeń                 | 299 przy przeglądzie siatki, 999 przy testach kluczowych; wynik niezależny od tej liczby                                      |
| 10 | struktura okazji i brzeg obszaru | bloki zabudowa $\times$ powiat jako jawny model struktury zdarzeń                                                             |
| 11 | zdarzenia współlokalizowane      | zdarzenie wykluczone z własnego sąsiedztwa według identyfikatora identyfikatorze rekordu                                      |
| 12 | kierunek relacji                 | różnica kierunków sprawdzona wobec rozkładu odniesienia, nie odczytana z dwóch liczb (A.4)                                    |