

**Adam Mickiewicz University**  
**Faculty of Mathematics and Computer Science**

GEOMETRY AND TOPOLOGY SEMINAR

1:45 PM, Wednesday, March 6, 2019  
B1-37, Collegium Mathematicum

**Speaker:** Karolina Paradysz (Uniwersytet Ekonomiczny w Poznaniu)

**Title:** Zastosowanie Topologicznej Analizy Danych w ocenie jakości estymatorów klasy SMO na przykładzie danych z Narodowego Spisu Powszechnego

**Abstract:**

Narodowy Spis Powszechny w 2011 po raz pierwszy w Polsce przeprowadzony przy pomocy metody reprezentacyjnej co wiąże się z wyborem odpowiednio skonstruowanego estymatora, który na podstawie próby najtrafniej będzie reprezentował badaną populację. W prezentacji skupiono się jedynie na jednej spośród wielu kategorii występujących w Spisie. Badaną kategorią jest niepełnosprawność, a odpowiedzi uzyskiwane od respondentów były deklaratywne, w związku z tym powstają zniekształcenia i braki odpowiedzi. Spośród wielu estymatorów pośrednich klasy SMO należy więc wybrać te, które będą spełniały kryteria formalne i merytoryczne (określają one pewien wzorzec, benchmark). W tym celu można wykorzystać także metody klastrowania, które pozwolą stworzyć wzorcowy układ odniesienia. Do stworzenia takiego benchmarku należałoby pójść o krok dalej niż zwykle klastrowanie.

Rozwiązaniem zaproponowanym w prezentacji jest analiza kształtu chmury punktów danych (Topology Data Analysis) pozwalająca pogrupować je i znaleźć zależności między nimi. Metoda ta nie wymaga wcześniejszych założeń co do szukanych zależności oraz pozwala odkryć relacje pomiędzy punktami danych, które mogłyby przejść niewykryte w ramach standardowych wspólnych metod statystycznych i taksonomicznych.

Zdaniem G. Carlssona et al. (2005) metoda wykorzystuje homologię persystentną, której główną zaletą jest mała wrażliwość na zakłócenia, szum informacyjny, niedokładność danych. Niewielkie przesunięcia punktów danych nie są w stanie znacząco zmienić diagramu persystencji i płynących z niego wniosków.