



Projekt końcowy

BADANIE TOPOLOGICZNYCH
ZALEŻNOŚCI MIĘDZY
KRAKOWSKIMI SZKOŁAMI
ŚREDNIMI

Jagoda Kurosad, Katarzyna Bęben

Słowny opis problemu:

W naszym projekcie skupiliśmy się na badaniu zależności między topologią krakowskich szkół średnich a ich poziomem (miejscem w rankingu www.waszaedukacja.pl). Chcieliśmy sprawdzić, jak mierzalne cechy zależne od człowieka wpływają na dalszy sukces placówki. W tym celu posłużyliśmy się poznanymi narzędziami z zakresu topologicznej analizy danych.

Opis danych:

Źródło: www.naszeszkoly.krakow.pl

W badaniu wzięliśmy pod uwagę takie parametry jak:

- liczba uczniów w szkole
- procent dziewczyn w liczbie uczniów szkoły
- średnia liczba uczniów w klasie
- liczba nauczycieli
- procent nauczycieli dyplomowanych
- procent nauczycieli stażystów
- progres potencjału z matematyki (różnica między procentem osób kończących szkołę z wysokim potencjałem z przedmiotu a zaczynających szkołę z wysokim potencjałem – badania na podstawie wyników egzaminu gimnazjalnego i matury)
- progres potencjału z j. polskiego (jak wyżej)

Były to współrzędne punktów chmury, zaś chmurę tworzyło kolejne 10 lat (2012-2021).

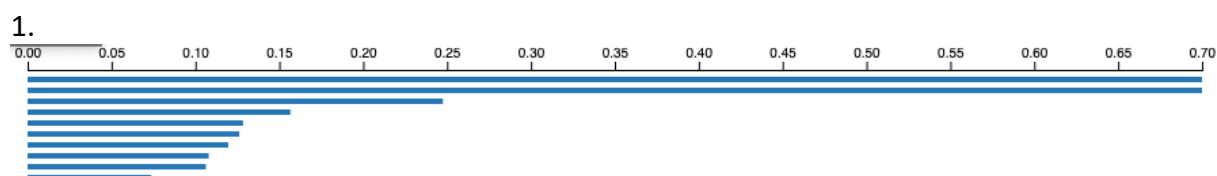
Wstępne hipotezy do analizy:

Miejsce w rankingu krakowskich szkół średnich zależy od ich topologii.

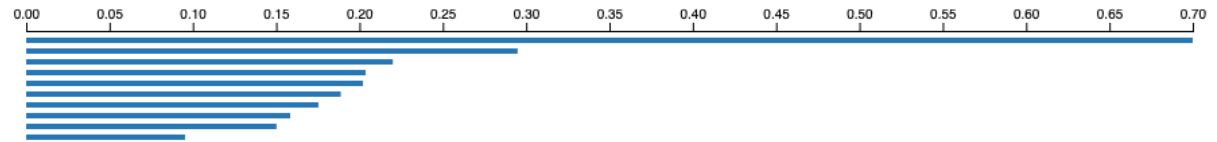
Krótki opis procedury filtracji:

Przeprowadzając filtrację stworzyliśmy na podstawie udostępnionych na stronie internetowej wykresów pliki z danymi wymienionymi powyżej, po czym przeprowadziliśmy ich normalizację globalną używając stworzonego przez nas programu (załączonego do folderu) o nazwie program.c, sprowadzając wartości do przedziału $<0,1>$. Do stworzenia barcodów wykorzystaliśmy stronę <https://live.ripser.org> wprowadzając tam nasze dane w postaci chmury punktów zapisanych w pliku .txt. Narzędzie używało filtracji kompleksu Vietoris-Ripsa.

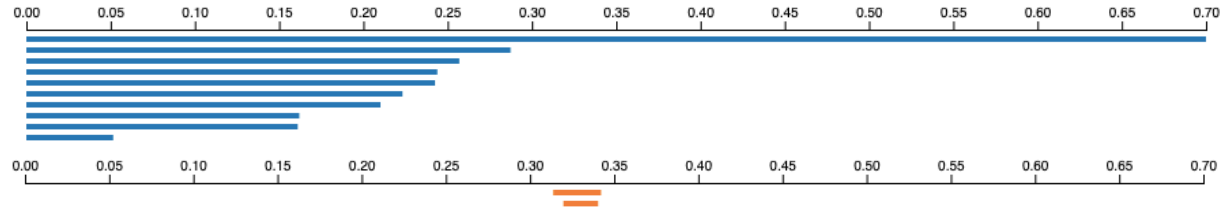
Otrzymałyśmy następujące barcodey (odpowiednio w kolejności zgodnej z pozycją w rankingu internetowym szkół), gdzie te w kolorze niebieskim odnoszą się do sympleksów 0-wymiarowych, zaś barwy pomarańczowej 1-wymiarowych (te pojawiają się tylko w sytuacji, w której takie sympleksy powstały):



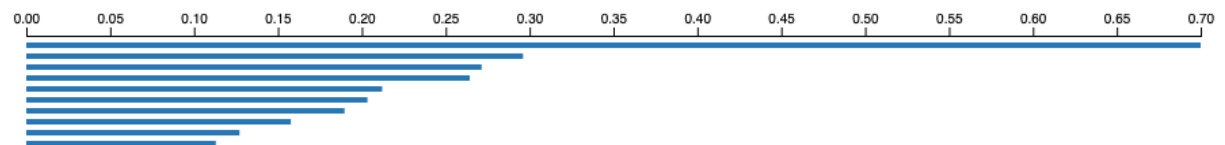
2.



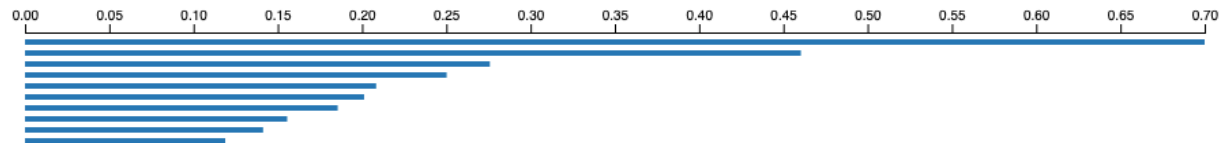
3.



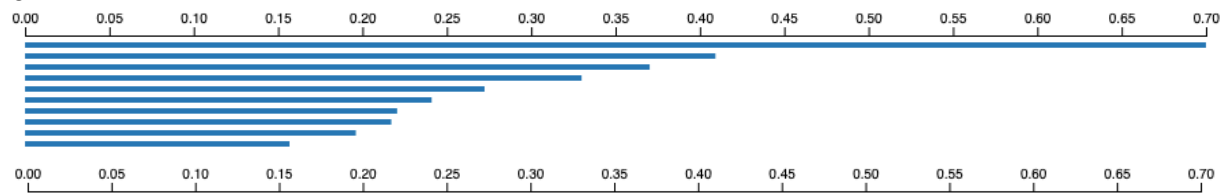
4.



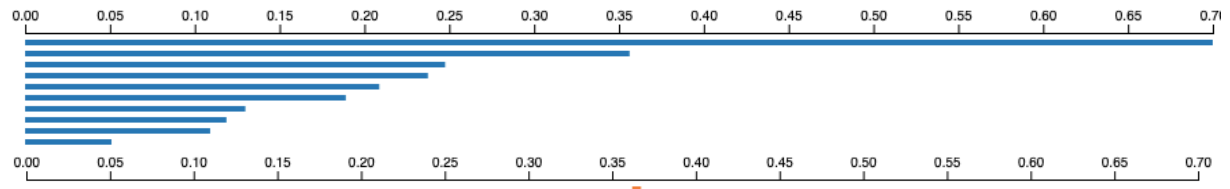
5.



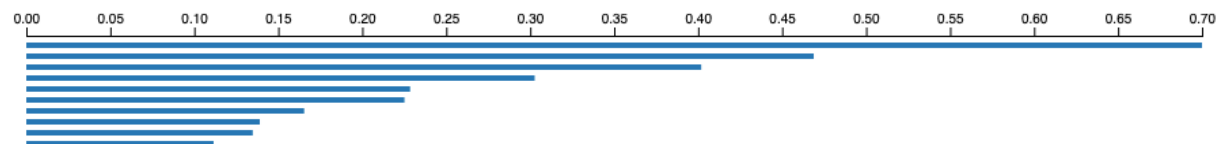
6.



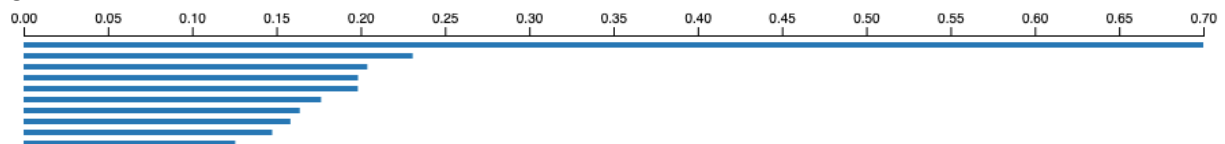
7.



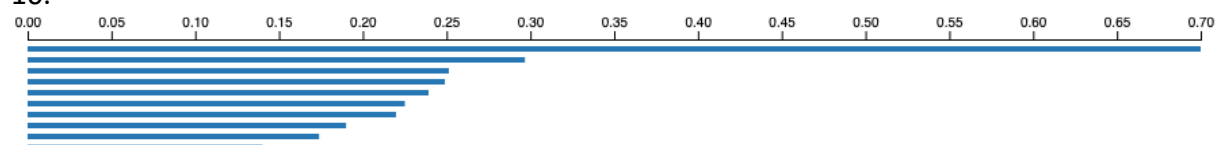
8.

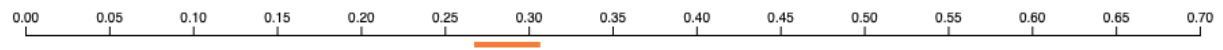


9.

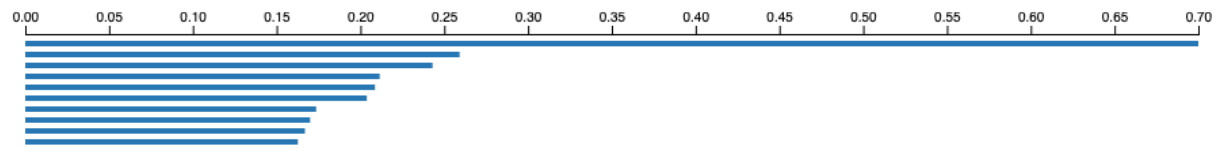


10.

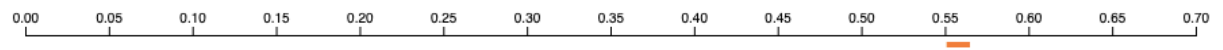
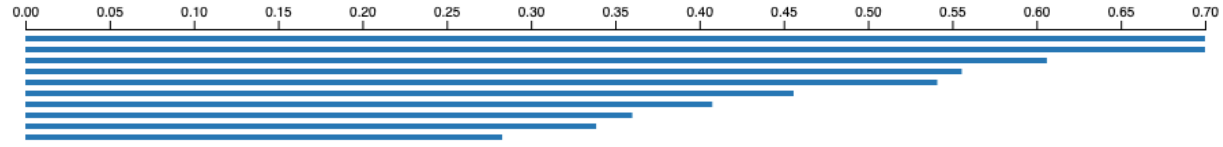




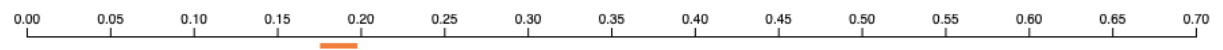
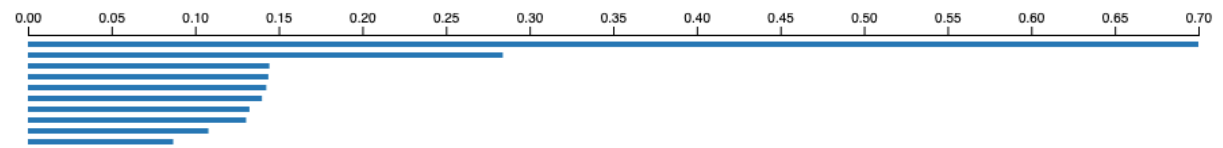
11.



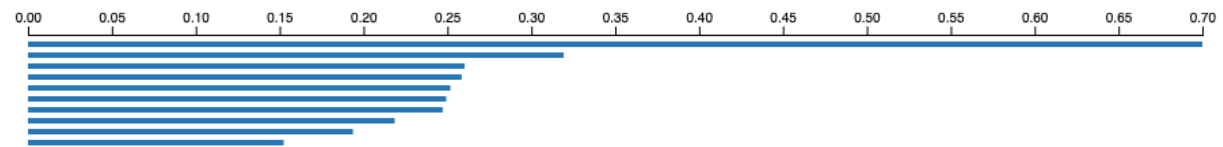
12.



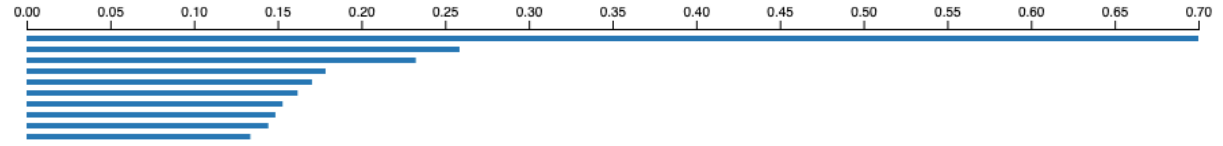
13.



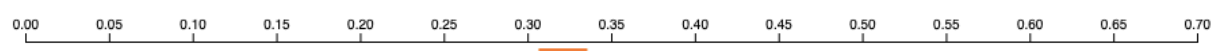
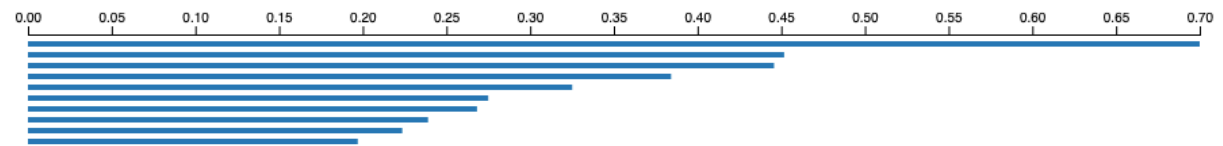
14.



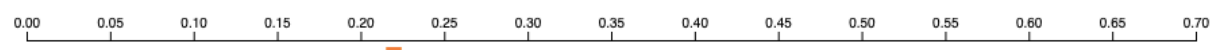
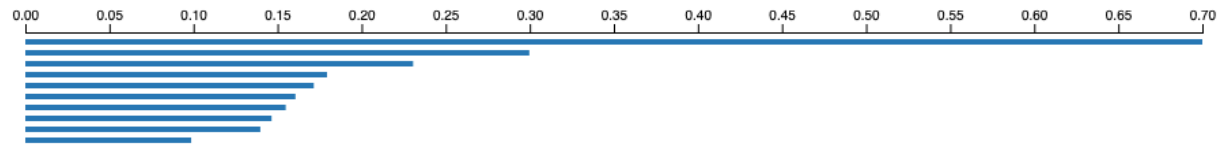
15.



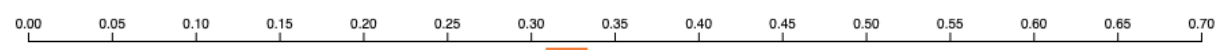
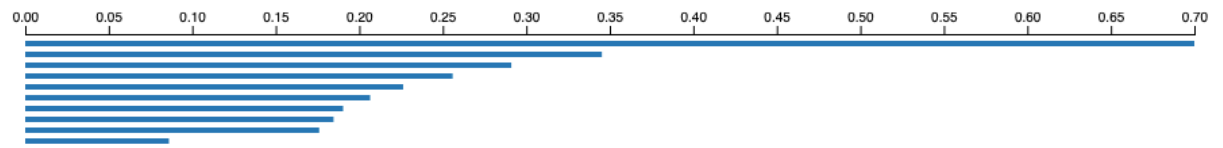
16.



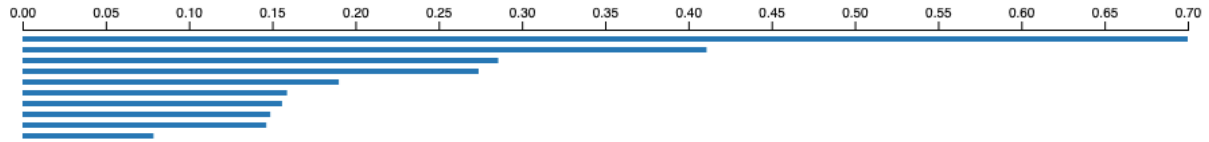
17.



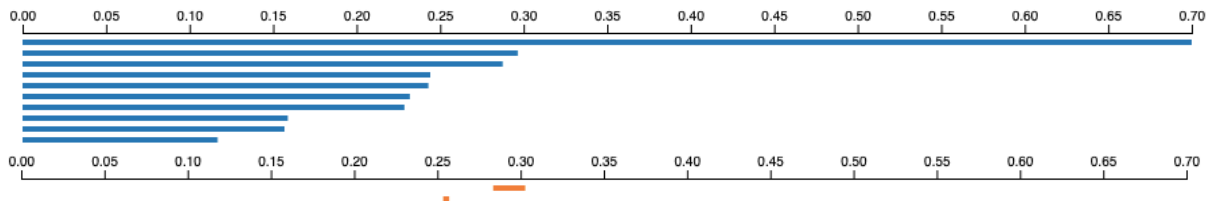
18.



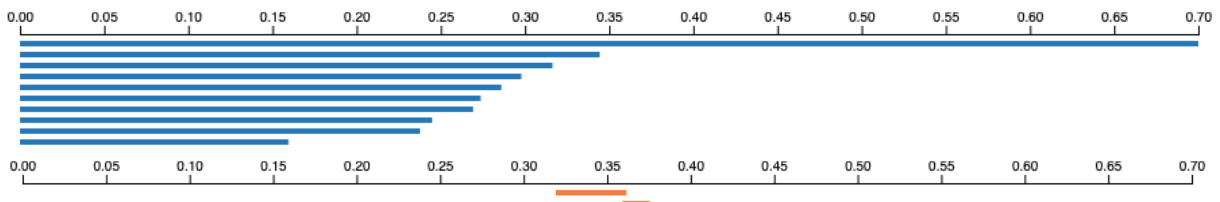
19.



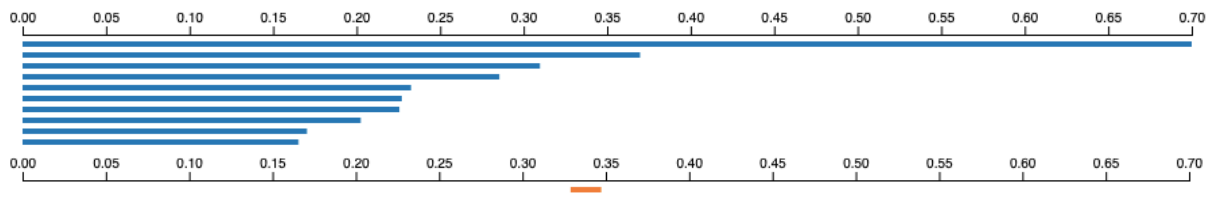
20.



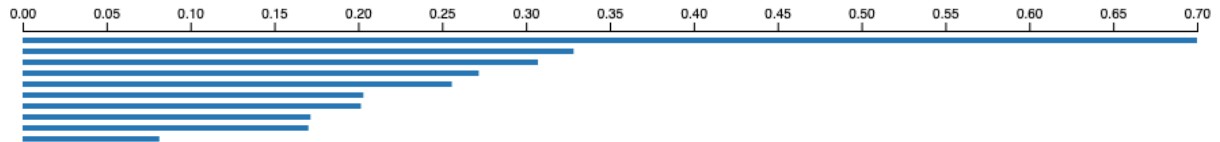
21.



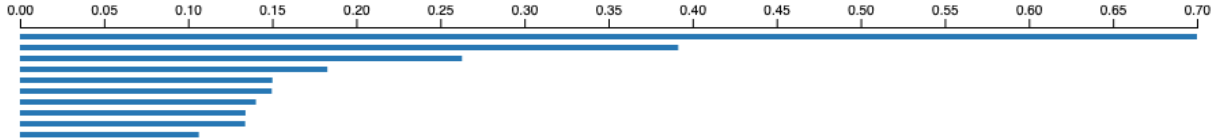
22.



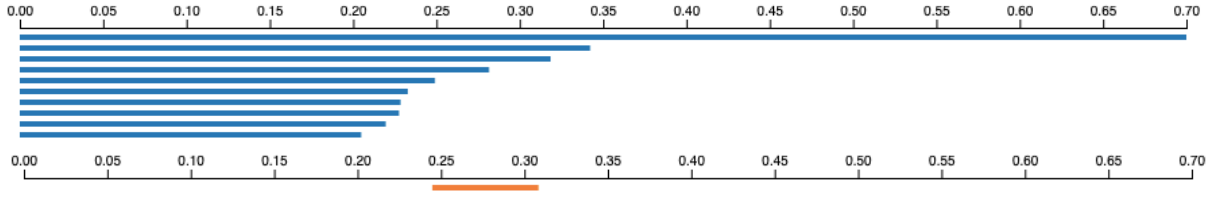
23.



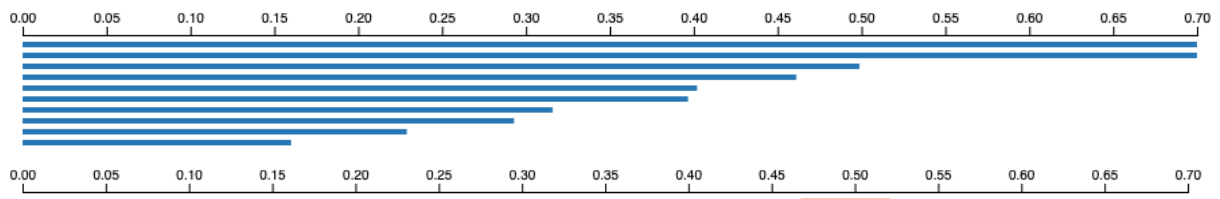
24.



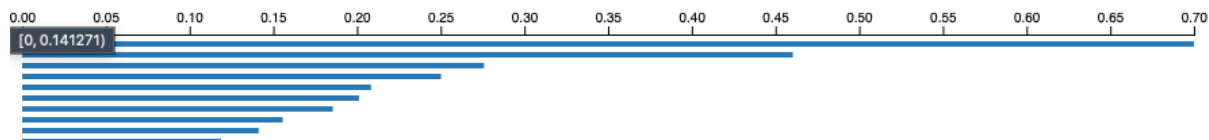
25.



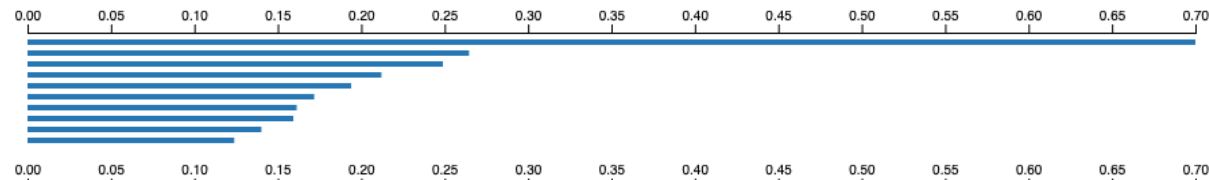
26.



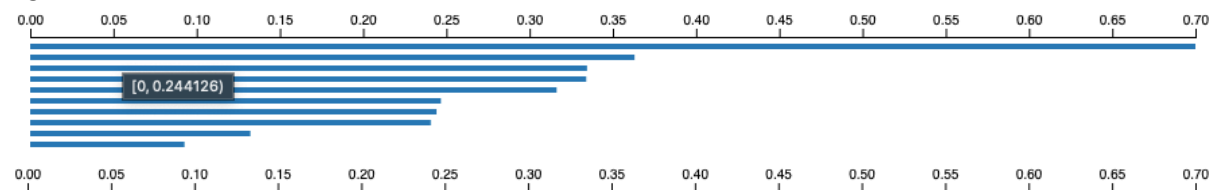
27.



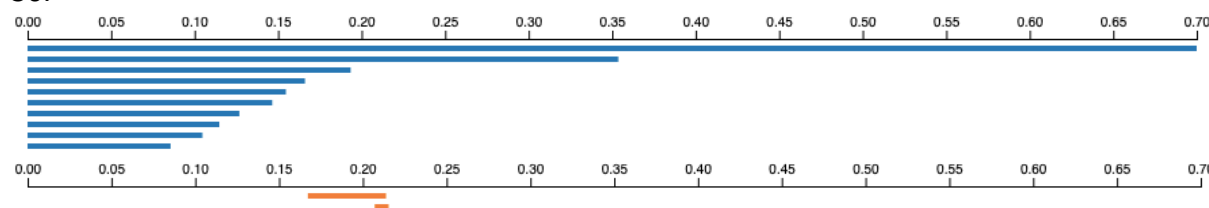
28.



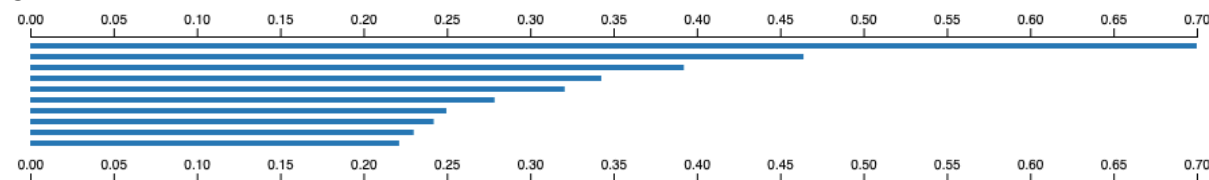
29.



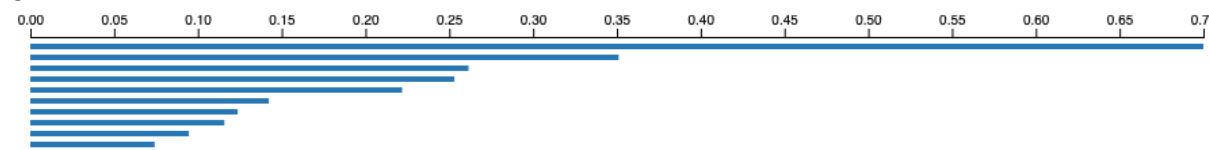
30.



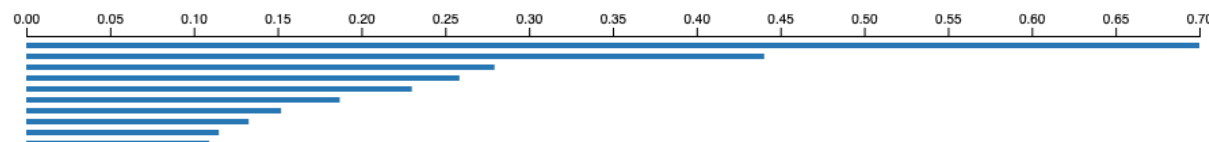
31.



32.



33.



Wnioski:

Na podstawie analizy powstałych barcodów zauważyliśmy nieznaczną większą dysproporcję barcodu najlepszego z liceów (wąska podstawa przy równocześnie żyjących przez całą filtrację dwóch dziurach). Pozostałe barcody nie wykazują szczególnych cech korelujących z miejscem w rankingu. Nie widać również podobieństw barcodów szkół o podobnym poziomie (miejscu w rankingu).

POSZERZENIE PROJEKTU:

OPIS PROBLEMU:

W kolejnej analizie sprawdzaliśmy czy na podstawie takich samych danych można rozróżnić topologicznie najlepsze krakowskie licea i technika.

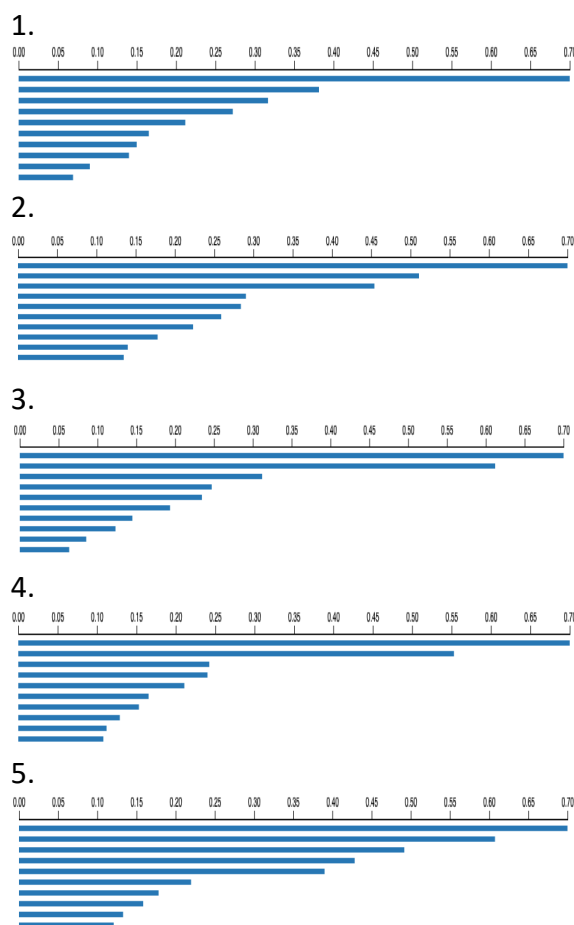
Wstępne hipotezy do analizy:

Za pomocą topologicznej analizy danych można rozróżnić krakowskie licea i technika

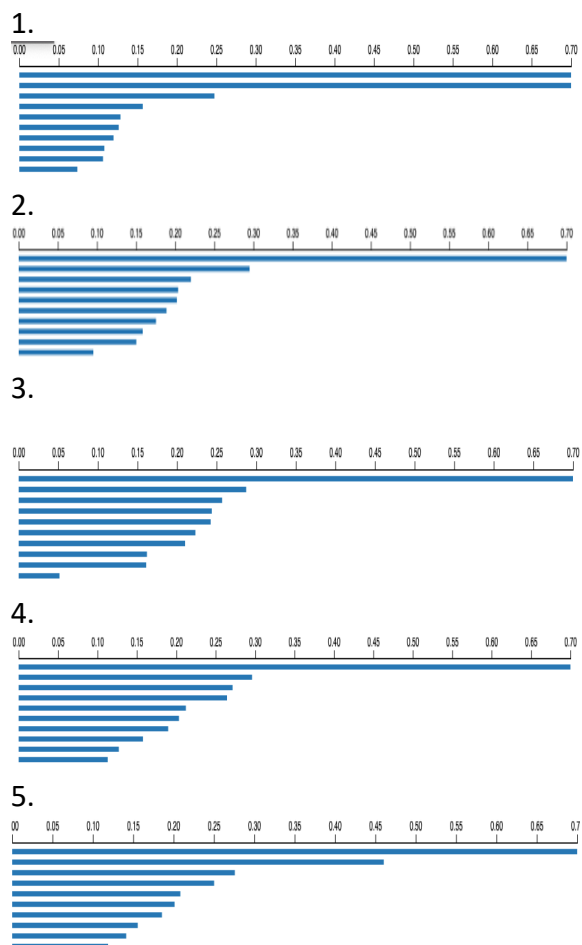
Dane, narzędzia i procedura: Przy poniższym badaniu skorzystaliśmy z tego samego źródła i użyliśmy takich samych narzędzi do TDA, jak w poprzednim etapie projektu. Pliki zostały stworzone przez chmury punktów zawierające takie same kategorie danych, a jako metodę filtracji ponownie wykorzystaliśmy kompleksy Vietoris-Rips. Do wyliczenia Bottleneck Distance skorzystaliśmy z używanego podczas zajęć programu BD.exe, do którego wczytaliśmy pliki utworzone za pomocą perseusWin.exe.

Otrzymane barcodesy (kolejność według internetowego rankingu):

TECHNIKA:



LICEA:



BD dla liceum:

| | 1 | 2 | 3 | 4 | 5 |
|---|---|--------|--------|--------|--------|
| 1 | X | 500000 | 500000 | 500000 | 117 |
| 2 | | X | 55 | 76 | 500000 |
| 3 | | | X | 70.5 | 500000 |
| 4 | | | | X | 500000 |
| 5 | | | | | X |

BD dla technikum:

| | 1t | 2t | 3t | 4t | 5t |
|----|----|--------|--------|--------|--------|
| 1t | X | 500000 | 500000 | 500000 | 500000 |
| 2t | | X | 500000 | 500000 | 500000 |
| 3t | | | X | 86 | 500000 |
| 4t | | | | X | 500000 |
| 5t | | | | | X |

Liceum x Technikum

| | 1t | 2t | 3t | 4t | 5t |
|---|--------|--------|--------|--------|--------|
| 1 | 500000 | 500000 | 133 | 105 | 500000 |
| 2 | 121 | 500000 | 500000 | 500000 | 500000 |
| 3 | 117 | 500000 | 500000 | 500000 | 500000 |
| 4 | 106 | 500000 | 500000 | 500000 | 500000 |
| 5 | 500000 | 500000 | 70 | 45 | 500000 |

WNIOSKI:

Po analizie wyglądu wygenerowanych barcodów dla 5 najlepszych krakowskich liceów i techników zauważamy nieznaczne różnice dotyczące ich kształtów (bacody dla liceów mają bardziej zbliżony kształt do prostokątów, natomiast barcody techników swoim kształtem bardziej przypominają trójkąty ostrokątne/schodki).