



NEIGHBORHOOD RECOMMENDATION

Balázs Dedinszky

7. November 2020

Business Problem



- Someone wants to move from one city to another
- Help them find similar neighborhood as they live now for quicker integration
- Priorize facilities that are important to families with children (e.g. schools, medical centers, playgrounds)
- The aim is to prepare a concept on four cities (Budapest, London, New York, Toronto) that can be applied on any other cities

We define „similar neighborhood” as an area where the same type and number of facilities (shops, pubs, schools, etc.) can be found nearby.

Data sources



Name of neighborhoods

- Either list of administrative areas in a city
- Or list of postal codes
- Collected from wikipedia pages
- Web scraping



Geocoding

- Assigning GPS coordinates to list of neighborhoods
- Done by remote service (Google Maps)
- API calls



Searching for facilities

- Searching venues at coordinates
- Foursquare API for most popular venues
- Google Places for additional priorities
- Within a pre-set radius (1000 m)

Methodology – 1

Collecting list of neighborhoods:

- Toronto and New York lists are available in .csv format from pervious lab sessions, these files are reuse in this porject.
- Budapest neighborhoods are collected from an open geographical database. The list had to be converted from geojson format, so it's already geocoded. (Not covered by this report.)
- London neighborhood list is collected from https://en.wikipedia.org/wiki/London_postal_district. Pandas library's built-in web-scraping feature is used.
- It was a matter of consideration whether list of negihborhood names or postal code were a better approach for geocoding.

Geocoding:

- Google Maps API calls on London list (was already available for other cities)
- Had to be careful because of some identical neighborhood names.

Methodology – 2

Feature collection:

- Foursquare API for getting top 100 popular venues within 1000 m radius of the center.
- Adding important non-profit facilities (e.g. schools) with keyword searches in Google Places
- One-hot coded list of categories aggregated for each neighborhood
- Number of facilities by categories are used as features.

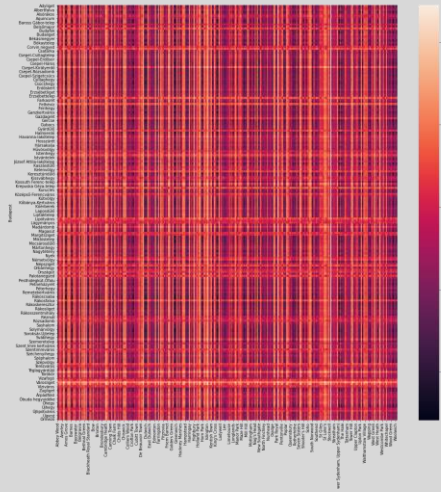
Calculating Similarity Index:

- Similarity is defined as similar kind of facilities with similar frequency.
- Euclidean distance is calculated between every pair of neighbourhoods.
- Result was normalized so each pair had a Similarity Index between 0-100 (100 being fully similar)

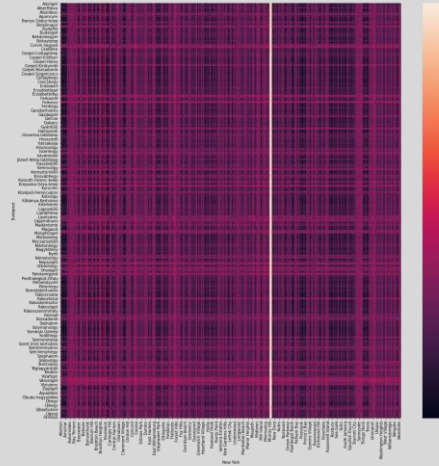
Results – overall view

Distances are calculated for each neighborhood-pairs and visualized on a heatmap:

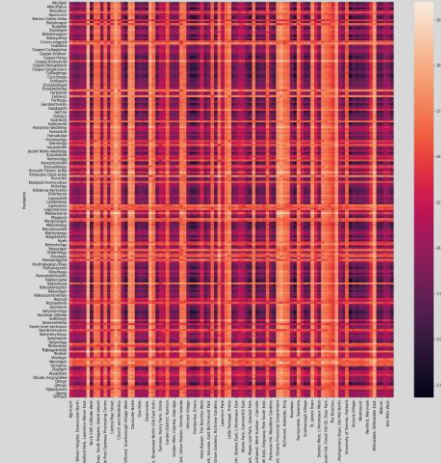
Budapest - London



Budapest – New York



Budapest – Toronto



The results are not really comprehensible this way some scenarios were prepared (see next slide)

Results – scenarios

Someone wants to move from Budapest to London, New York or Toronto.

Three neighborhoods in Budapest has been selected:



Budafok

- Suburban area
- Detached houses and parks



Terézváros

- Inner city
- Big, old condominiums
- Lot of venues



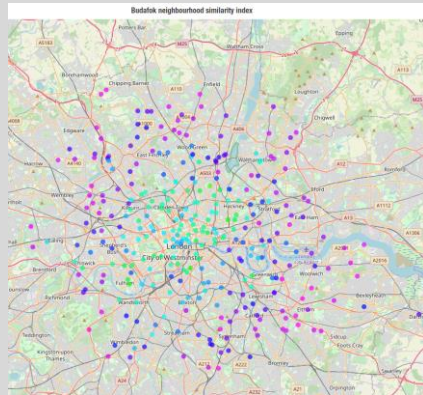
Újpest

- In-between
- Mixed residential and industrial area

3 source neighborhood and 3 target cities: the model is tested in 9 different scenarios.

Results – London

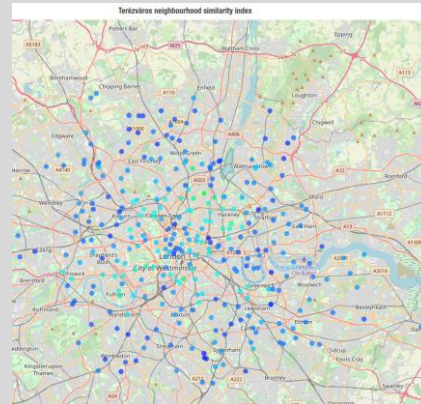
Budafok



Top 3 similar neighborhoods:

- Kingston Vale
- Sydenham Hill
- Well Hall

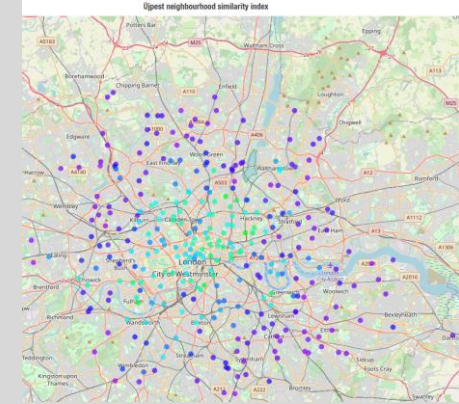
Terézváros



Top 3 similar neighborhoods:

- Gypsy Hill
- Shoreditch
- Greenwich

Újpest



Top 3 similar neighborhoods:

- Osidge
- The Hyde
- Hanwell

Click on browser icons to open interactive maps

Results – New York

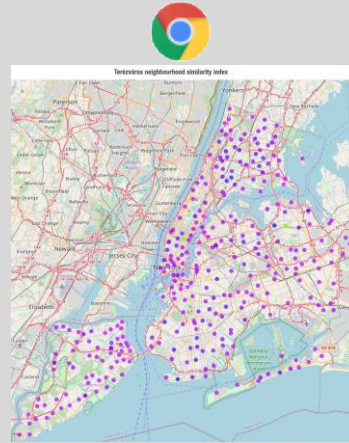
Budafok



Top 3 similar neighborhoods:

- Park Hill
- Arden Heights
- Oakwood

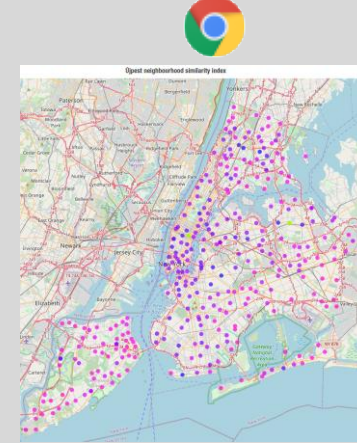
Terézváros



Top 3 similar neighborhoods:

- Yorkville
- Edgewater Park
- Thorgs Neck

Újpest



Top 3 similar neighborhoods:

- Brookville
- Hunts Point
- Arlington

Click on browser icons to open interactive maps

Results – Toronto

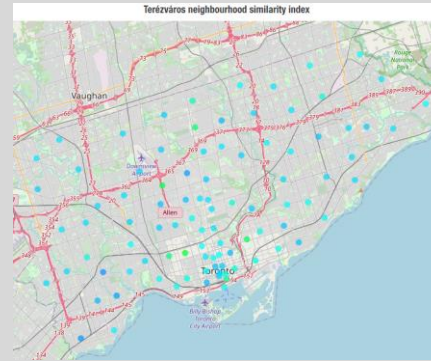
Budafok



Top 3 similar neighborhoods:

- Parkwoods
- Rouge Hill
- Upper Rouge

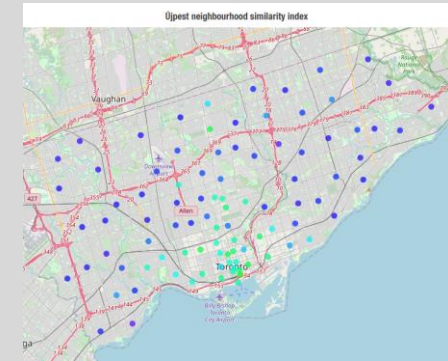
Terézváros



Top 3 similar neighborhoods:

- The Kingsway
- Alderwood
- Old Mill South

Újpest



Top 3 similar neighborhoods:

- New Toronto
- Parkwoods
- Islington Avenue

Click on browser icons to open interactive maps

Discussion

Overall findings:

- results in line with expectations: a suburban area of Budapest is more similar to outer areas of other cities, inner city area is more similar to inner ones.
- It seems that Toronto in overall is more similar to Budapest than the other two cities. It is not surprising as Toronto is much closer to Budapest in terms of area and population than the other two megapolises.
- Even though the overall standard deviation is quite similar in each matrices, it is more difficult in New York maps to differentiate more similar and less similar areas in the three scenarios. See considerations below.

Model seems more useful in Toronto and London than in New York. Possible reasons:

- Overall size and population density of a city has an impact on individual neighborhoods
- Limitation of foursquare search (max. 100/neighborhood)
- Diversity of facility categories may be bigger in bigger cities

Possible improvements



- Raising search limits (100 in foursquare, 20 in Google)
- Aggregating similar features (e.g. „restaurant“, „italian restaurant“ and „pizzeria“ can be grouped together)
- Additional features: adding other attributes (e.g. crime rate, pollution, school results)