Rozwiązując korki

Analiza przestrzenna
z teorią grafów

Agenda

Wstęp

Dane przestrzenne i GIS

Modelowanie ruchu samochodowego

Podstawy teorii grafów i metryk centralności

Notebooki

O mnie

- 5 rok Informatyki - Data Science na WI AGH
- Data Science Engineer @ Sabre
- Obszar badań – GNN

Hobbistycznie

- Amator urbanistyki
- Fan gier strategicznych (4X)

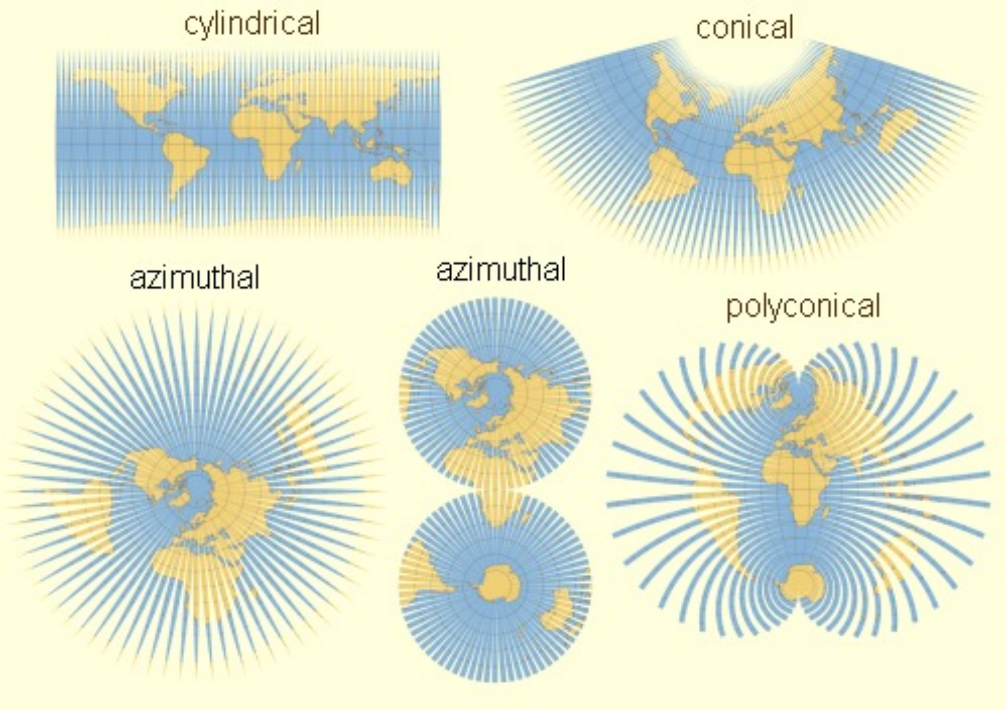
Dane przestrzenne

GIS

Projekcja Ziemi do 2D

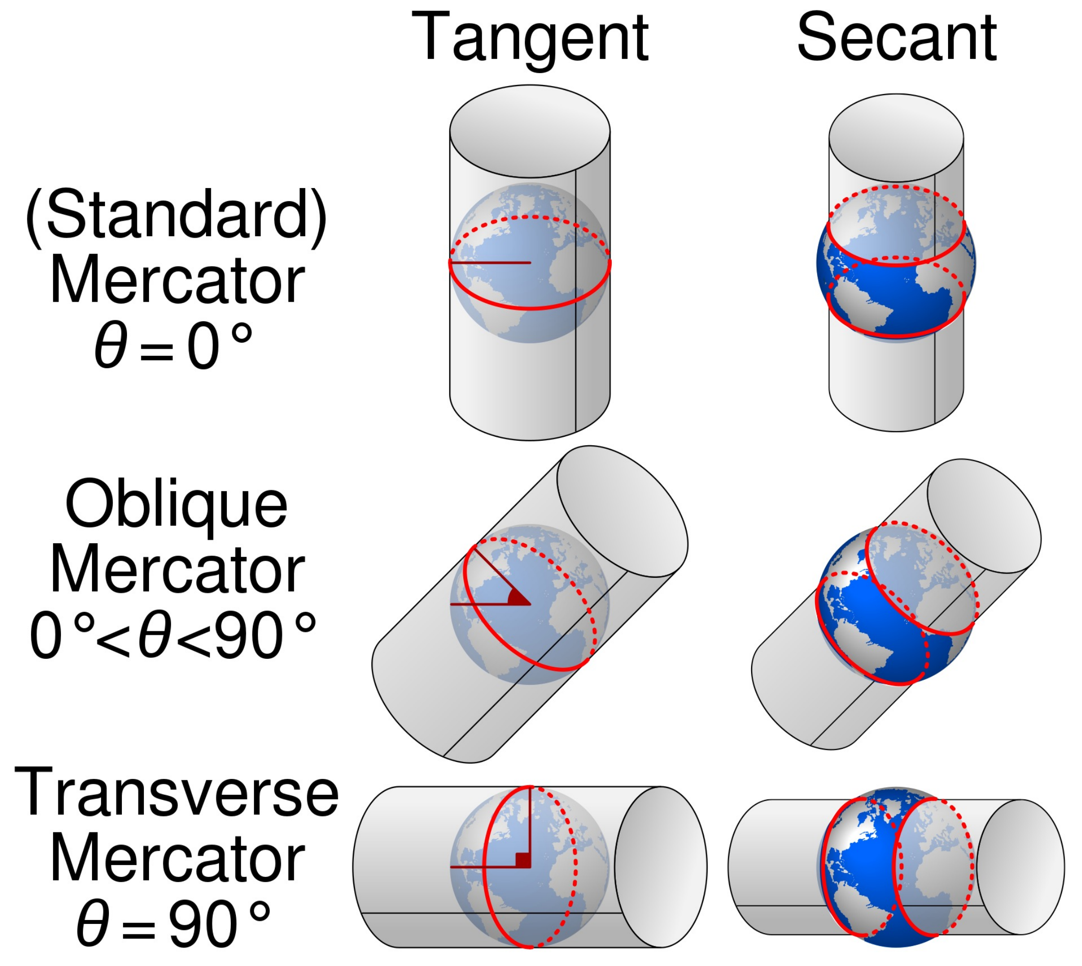
Aby pozwolić na sprawne analizy danych geoprzestrzennych musimy najpierw je przedstawić w formie 2D.

Jako, że Ziemia nie jest ani sferą ani elipsoidą, dopuszcza to wiele różnych metod projekcji I jeszcze więcej problemów



https://www.semanticscholar.org/paper/Unfolding-the-Earth%3A-Myriahedral-Projections-Wijk/3228862dc4d2743e55d5c077817cca0623492bcc

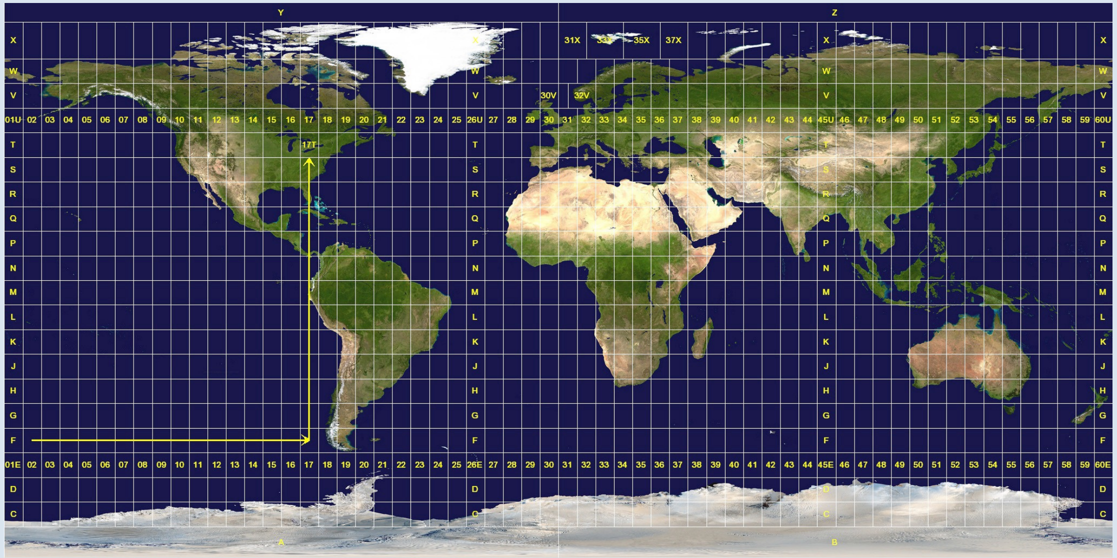
Transverse Mercartor Projection



https://commons.wikimedia.org/wiki/File:Comparison_of_Mercator_projection
https://www.thetruesize.com/

Zniekształcenia projekcji Mercartor oraz jej segmentacja (UTM)

Projekcja Mercartor wprowadza zaburzenie między różnymi obszarami na różnych długościach/szerokościach geograficznych. W związku z tym stosuje się segmentację UTM.



https://commons.wikimedia.org/wiki/File:Utm-zones.jpg

Czy sam Mercartor wystarczy?

Choosing UTM zone to use for large country?

Asked 8 years, 11 months ago Modified 6 years, 4 months ago Viewed 19k times

9

I want to reproject OpenStreetMap roads data from the current projection (WGS 84) to UTM, since I read in another GIS SE Q&A(Getting $length in meters in QGIS?) that WGS 84 cannot be used to measure length in m or km. My end goal is to calculate the total roadway kilometres within each district, so that I can enter that variable into my regression model. The country in question is Indonesia, and as you can see from the map below, it covers UTM zones 46-54.

6

You use a UTM zone when your area of interest fits completely within it or very nearly so. A UTM zone is not appropriate when your area of interest spans several zones such as in your case. A little overlap into a neighboring zone might be ok, but the further away from the zone you pick, the more distortion there will be and the more it matters. I found this page with a graphic example.

You actually want a projection designed to cover that area and minimize the appropriate distortions (shape, area, distance, direction - can't have them all minimized in one projection), as mdsumner suggests. Note that you can always take a standard projection and modify it to best suite your particular area of interest by changing the detailed settings. That does take a level of understanding to know the impact your choices have on distortions and whether or not they are acceptable. And of course the larger the area you look at, the greater the compromise you make in distortions for some areas. This Esri pdf is a good introduction to projections.

You could also get the length calculations you need on a per zone basis if you want to stick with UTM and its level of distortion/error.

1. First you'd find a dataset that is the UTM grid in polygon form with an attribute identifying the zone.
2. Intersect that with your roads so that each road gets an attribute defining what zone it is in. This also splits the road at the zone boundary since it's likely that you'll have roads crossing them. With a spatial join you could end up counting the same road twice.
3. Since you also want totals per district, you'd want to next Intersect your updated roads with the district boundary layer to again split them at boundaries and get an attribute defining their district. Be cautious with intersect - if you have roads that aren't within a district, they will get dropped because intersect only looks at areas of overlap between layers and not just one or the other.
4. Then you can add a field to hold your length calculation (you do *not* want to rely on the default shape_length field, since that is system tracked and tied to the projection of the data). Also note you need to do both intersects before calculating length, because *your* length field *won't* automatically update with changes. If you only did zones, then intersected with districts, each district piece of the road would have a length value for the whole zone and not just the district unless you recalculated it.
5. Now start reprojecting (**not** *defining*, note the difference) the data into each UTM zone, select only the records with that zone attribute, and field calculate their length into your field. You don't need to split it to separate layers/shapefiles to do that, just make sure you're only working with the correct records for the current projection. Once you've calculated all the lengths you can go back to WGS84, and your length attribute will remain as whatever units it was originally calculated to.
6. Finally, in ArcGIS you'd use Summary Statistics to total road length per district, but in QGIS you need the GroupStats plugin or something similar. Or you pull the table into another software that lets you do the same district case-based sum.

As an experiment, you can always add another length field, set a different projection that covers more than just a UTM zone, such as EPSG:3001 I mentioned earlier in a comment, field calculate that length, and compare it to your UTM lengths to see what kind of difference it makes.

https://gis.stackexchange.com/questions/141496/choosing-utm-zone-to-use-for-large-country

Jak przechowujemy mapy – GIS

Dane przestrzenne są przechowywane w formie bazy danych, w której trzymamy geometrię (np. drogi, rzeki, kraju) wraz z dodatkowymi atrybutami.

```
gdf['railway'].unique()
Executed at 2024.03.30 20:29:25 in 30ms

1-10   18 rows × 1 columns np.ndarray
   0
0  None
1  rail
2  funicular
3  tram
4  abandoned
5  preserved
6  disused
7  construction
8  platform
9  razed
```

```
<class
'geopandas.geodataframe.GeoDataFra
me'>
RangeIndex: 3333144 entries, 0 to
3333143
Data columns (total 37 columns):
 #  Column        Dtype
---  ------       -----
 0  ogc_fid       int64
 1  WKT_GEOMETRY  object
 2  osm_id        object
 3  name          object
 4  highway       object
 5  waterway      object
 6  aerialway     object
 7  barrier       object
 8  man_made      object
 9  railway       object
 10  z_order      float64
 11  other_tags   object
 12  table_name   object
 13  ref          object
 14  address      object
 15  is_in        object
 16  place        object
 17  type         object
 18  osm_way_id   object
 19  aeroway      object
 20  amenity      object
 21  admin_level  object
 22  boundary     object
 23  building     object
 24  craft        object
 25  geological   object
 26  historic     object
 27  land_area    object
 28  landuse      object
 29  leisure      object
 30  military     object
 31  natural      object
 32  office       object
 33  shop         object
 34  sport        object
 35  tourism      object
 36  geom         geometry
dtypes: float64(1), geometry(1),
int64(1), object(34)
memory usage: 940.9+ MB
```

Modelowanie ruchu samochodowego

Powody i metody

https://en.wikipedia.org/wiki/File:Gospers_glider_gun.gif

Po co modelować samochody?

- Projektowanie nowych dróg
- Odciążanie infrastruktury miasta
- Analiza krytycznych punktów infrastruktury

Metody modelowania ruchu

Modelowanie mikroskopowe

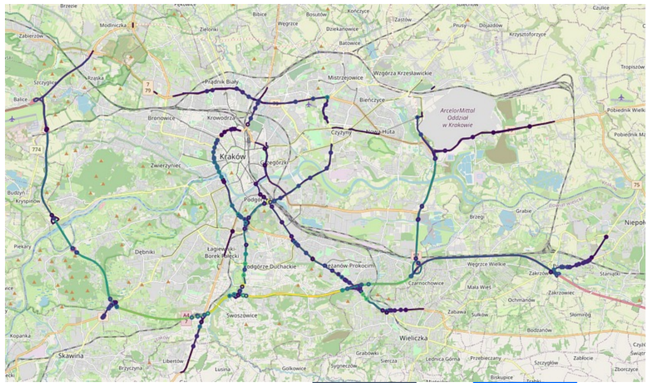
- Symulujemy każdego użytkownika ruchu osobno

Np. Automat komórkowy

Modelowanie makroskopowe

- Modelujemy tylko generalne trendy ruchu

Np. centralności grafowe

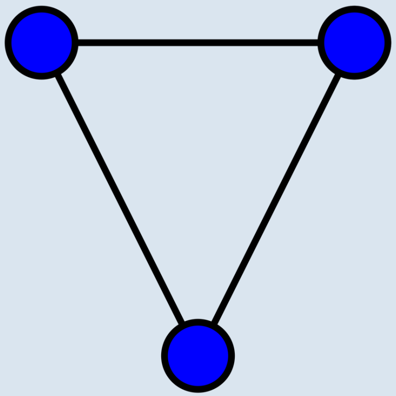


https://en.wikipedia.org/wiki/File:Gospers_glider_gun.gif
https://towardsdatascience.com/mapping-the-jams-traffic-analysis-using-graph-theory-a387135ea748

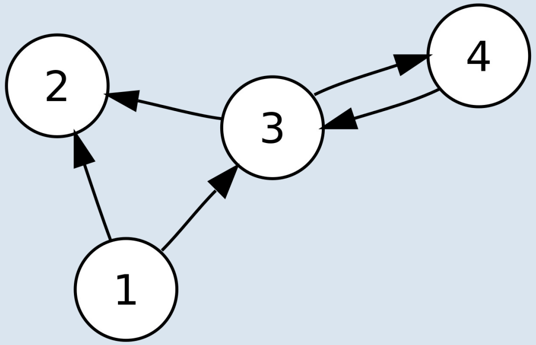
Teoria Grafów

Miary centralności

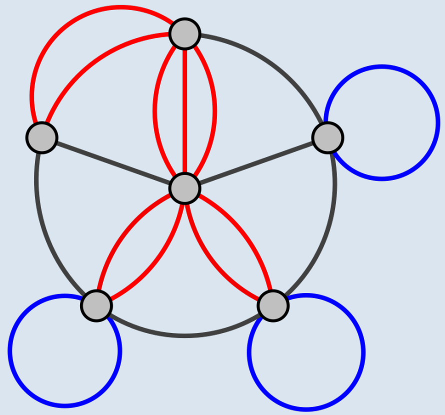
Rodzaje grafów



Grafy proste
(nieskierowane)

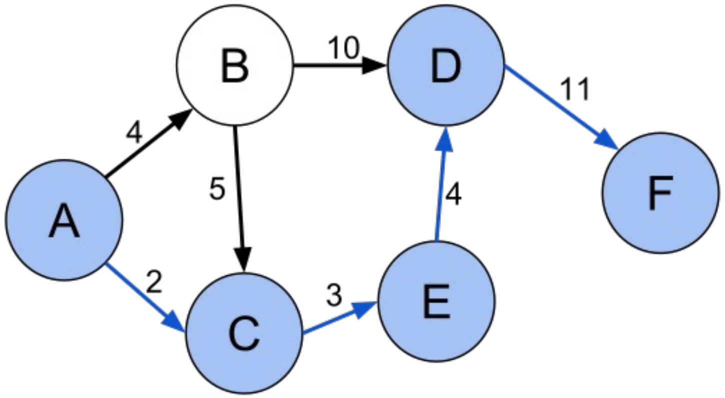


Grafy
skierowane



Mutligrafy
[skierowane]

https://commons.wikimedia.org/wiki/File:Undirected.svg
https://commons.wikimedia.org/wiki/File:Directed_graph_no_backgrour
https://commons.wikimedia.org/wiki/File:Multi-pseudograph.svg



Grafy ważone

Krawędzie grafu mogą mieć przypisane do siebie wartości rzeczywiste, nazywane wagami.

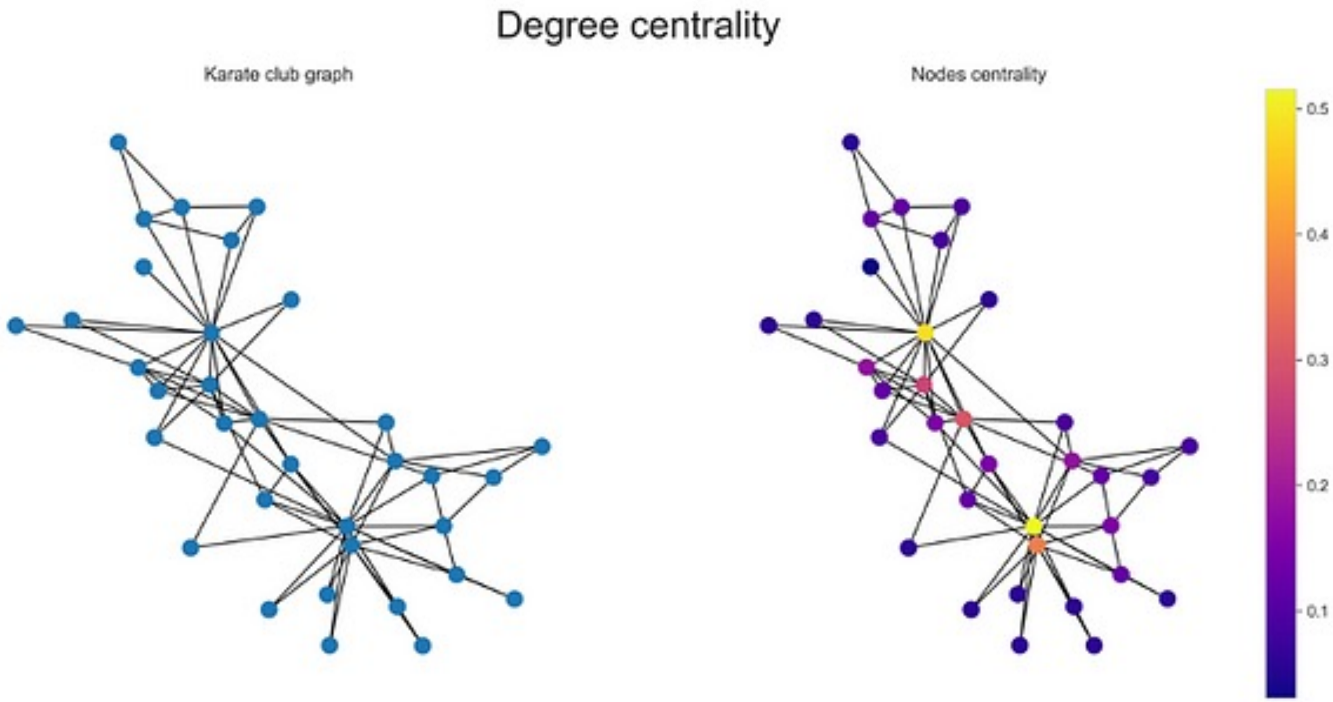
Są one wykorzystywane w różnych algorytmach gdzie koszt przejścia przez krawędź może być różny, np:

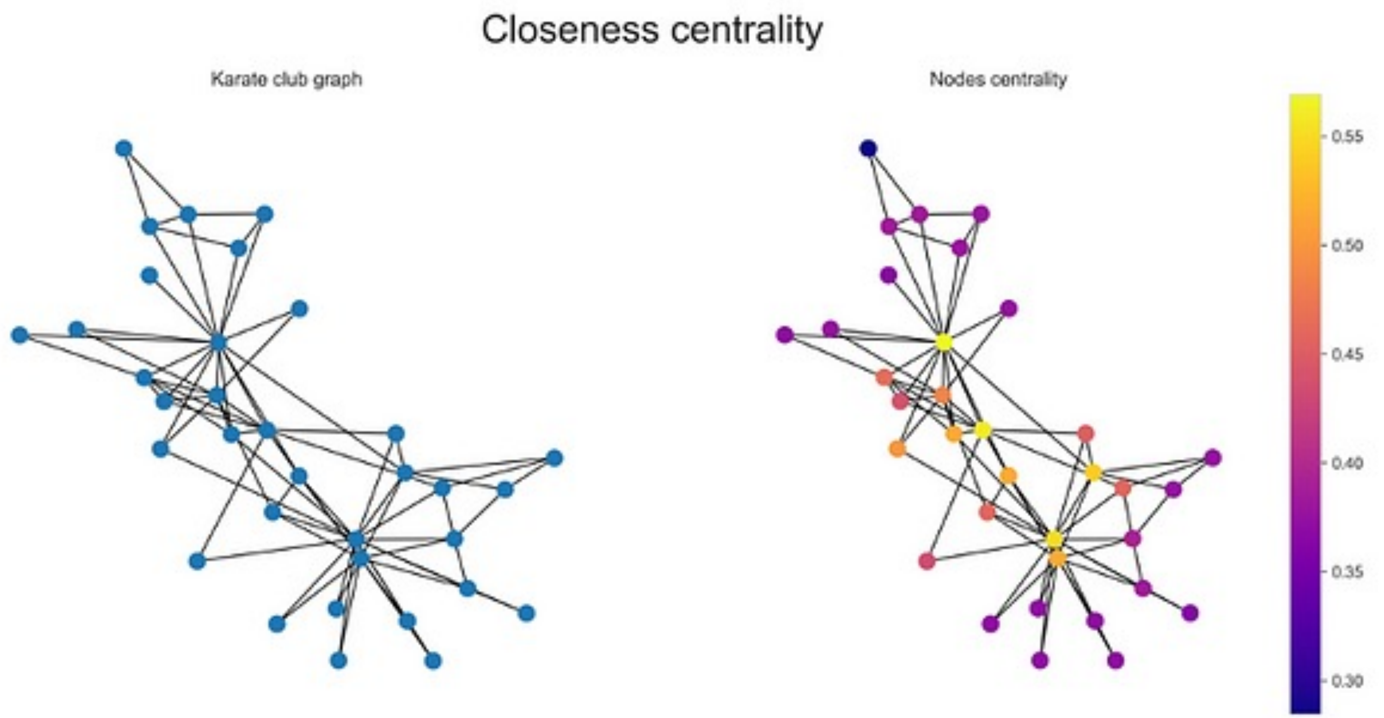
- Najkrótsze ścieżki
- Problemy przepływu

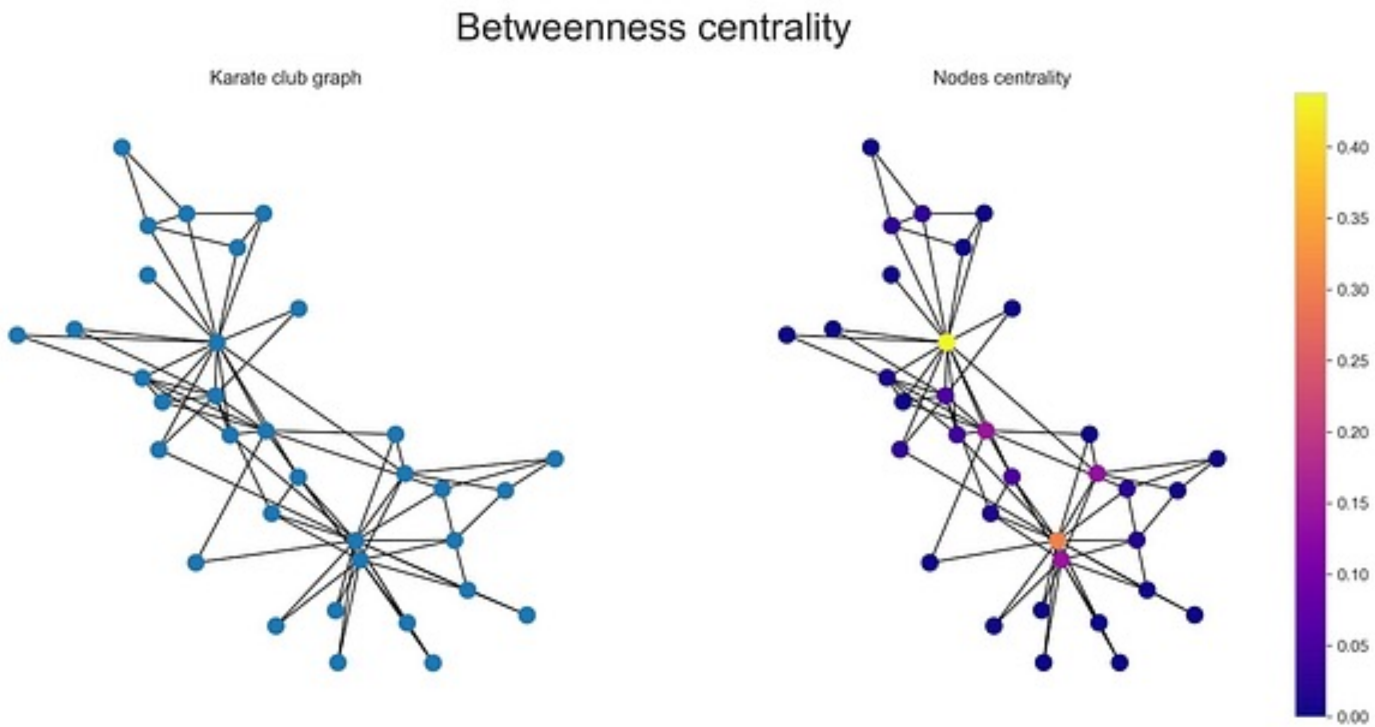
https://commons.wikimedia.org/wiki/File:Shortest_path_with_direct_weights.svg

Miary centralności

Miary centralności służą do opisania jak bardzo “centralne” (kluczowe / znajdujące się w środku) są elementu grafu. Opisują one liczbą rzeczywistą krawędzie lub wierzchołki.


Degree centrality
Karate club graph
Nodes centrality
0.5
0.4
0.3
0.2
0.1


Closeness centrality
Karate club graph
Nodes centrality
0.55
0.50
0.45
0.40
0.35
0.30


Betweenness centrality
Karate club graph
Nodes centrality
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00

Dziękuję za uwagę

Mateusz Praski

mateusz.praski@gmail.com

in/mateusz-praski @ LinkedIn