## Statistická analýza liniových prvků

## Linie mohou na mapách reprezentovat dva příbuzné objekty:

• **Vlastní linie** - reprezentují a lokalizují skutečně lineární geografické fenomény (řeky, silnice, potrubí)

• **Hrany** - rozdělují plochy a povrchy (hraniční linie, lomové linie). Hrany nemají šířku.

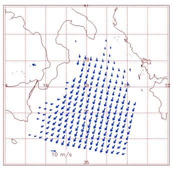
**Geometrické charakteristiky -** linie může být prezentována jako:

• **Jednoduchá** spojnice – pouze dvou bodů (koncový a počáteční – délka je euklidovská vzdálenost

• Posloupnost několika liniových segmentů - **řetězec**

## Linie mohou vystupovat na třech úrovních, které představují jistou hierarchii:

1. „**Prosté“ linie** – např. zlomy – lze určit jen délku a orientaci. Mohou existovat jako jednoduché spojnice dvou bodů či jako „řetězec“
2. „**Trajektorie**“ – vektor pole větru – lze určit velikost (délku), orientaci a směr
3. **Sítě** - dopravní sítě, říční síť – lze určit prostorové uspořádání – topologické vztahy, konektivitu, dostupnost, …



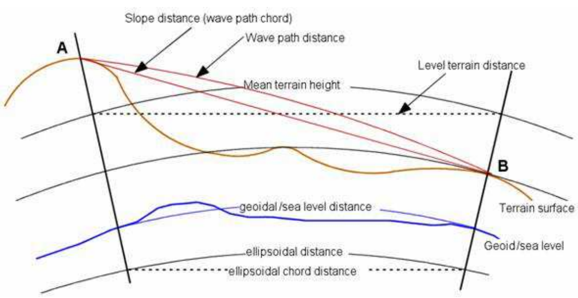

## Problémy prezentace linií

• Generalizace a zjednodušení průběhu

• Linie jako spojnice posloupnosti lomových bodů, mezi lomovými body je rovná.

• **Problém měření vzdáleností** - Někdy se místo měření vzdálenosti v délkových jednotkách používá cestovní čas a dopravní náklady.

## Problém měření vzdáleností




## Prostorové atributy liniových prvků

**Délka linie** může být definována jako:

• přímá vzdálenost (vypočtená z Pythagorovy věty)

• „skutečná“ vzdálenost (součet přímých vzdáleností jednotlivých segmentů)

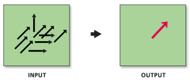
**Orientace linie** - orientace neurčuje směr (např. JV = SZ) – orientace zlomů, ulic. Nemá smysl otázka odkud - kam?

**Směr linie** - typicky např. vektor pole větru

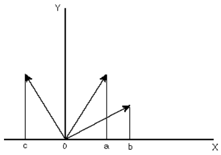
Uvedené atributy linií lze vyjádřit pro jednotlivé segmenty sítě či pro celou síť jako celek (průměrná délka segmentů sítě, převládající orientace či směr segmentů sítě, …).

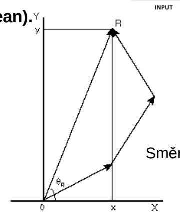
**Topologie** - atributy popisující jejich strukturu a uspořádání jako celek a dále popisují vztahy segmentů uvnitř sítě (topologii) – konektivita, segmentů, dostupnost, apod.

## Směrová statistika
### (Directional statistics)




**Směrový průměr (directional mean).**




Směr výsledného vektoru

$$\tan\theta_R = \frac{oy}{ox}$$

kde *oy* je suma délek vektorů ve směru osy *y* a *ox* suma délek vektorů ve směru osy *x*.

Protože pracuje se směrem (úhlem) a ne s délkou, je možné linii prezentovat na základě jednotkových vektorů. Vektorovým součtem – přidáním počátku druhého vektoru na konec prvního dostaneme směrový průměr.

## Směrový průměr

Délka ve směru osy *y* je v podstatě *sin* úhlu a délka na ose *x* je *cos* úhlu. Potom, jsou-li vektory označeny *a, b, c* a odpovídající úhly $\theta_a$, $\theta_b$, $\theta_c$, potom:

$$\tan\theta_R = \frac{\sin\theta_a + \sin\theta_b + \sin\theta_c}{\cos\theta_a + \cos\theta_b + \cos\theta_c}$$

Obecně, máme-li *n* vektorů *v* a úhel vektoru *v* od osy *x* je $\theta_v$, výsledný vektor *OR* má úhel $\theta_R$, měřený proti směru hodinových ručiček od osy *x:*

$$\tan\theta_R = \frac{\sum \sin\theta_v}{\sum \cos\theta_v}$$

což je tedy tangenta úhlu výsledného vektoru. Směrový průměr je potom *arctan* z výše uvedeného výrazu.

## Směrový průměr

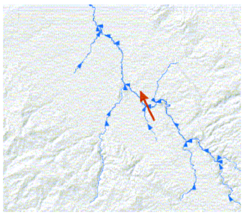
Výsledná hodnota směrového průměru musí zohledňovat specifika jednotlivých kvadrantů, jak uvádí následující pravidla:

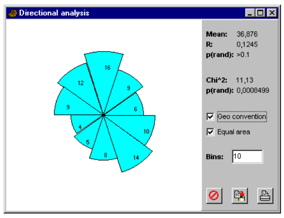
1. čitatel i jmenovatel v tan $\theta_R$ jsou oba kladné – není nutná žádná úprava (vektor leží v 1. kvadrantu)
2. čitatel je kladný jmenovatel záporný – směrový průměr bude 180 - $\theta_R$, (vektor leží v 2. kvadrantu)
3. čitatel i jmenovatel v tan $\theta_R$ jsou oba záporné – směrový průměr bude 180 + $\theta_R$, (vektor leží v 3. kvadrantu)
4. čitatel je záporný, jmenovatel kladný – směrový průměr bude 360 - $\theta_R$, (vektor leží v 4. kvadrantu)

Praktický výpočet spočívá v určení sin a cos úhlů všech vektorů. Určí se jejich sumy a vytvoří poměr, který je tangentou výsledného úhlu.

## Směrový průměr

- Průměrný (= převládající) směr údolí
- Průměr nemusí být typický
- V řadě případů je vhodnější prezentace směrovou růžicí, která je analogií rozdělení četností resp. histogramu





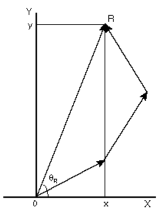

http://folk.uio.no/ohammer/past/study11.html

## Směrový rozptyl (Circular variance)

Stejně jako v případě klasické popisné statistiky je charakterizování souboru prvků pouze měrou úrovně často nedostatečné a může být i zavádějící (např. pokud dva vektory budou svírat úhel 180 stupňů). Proto je nutné použít i měr variability (rozptylu).

Pokud dáme dohromady vektory podobného směru, výsledný vektor bude relativně dlouhý. Jeho délka se bude blížit *n*, pokud bude *n* jednotkových vektorů. Naproti tomu, pokud dáme dohromady vektory opačného či značně rozdílného směru, výsledný vektor bude významně menší než *n*.

Délku výsledného vektoru můžeme použít jako statistiku, která reflektuje variabilitu ve směru jednotlivých vektorů. Na základě výše uvedeného tedy platí:




$$OR = \sqrt{(\sum \sin\theta_v)^2 + (\sum \cos\theta_v)^2}$$

## Směrový rozptyl

**Směrový rozptyl** $S_v$ se vypočte ze vztahu:

$$S_v = 1 - OR/n$$

kde *n* je počet vektorů. $S_v$ může nabývat hodnot 0 až 1. Je-li $S_v=0$, potom *OR=n* a všechny vektory mají stejný směr. Je-li $S_v=1$, potom *OR=0*, všechny vektory mají opačný směr a výsledný vektor je bod.

## Úvod do statistického popisu sítí

Základní pojmy používané v síťové analýze:

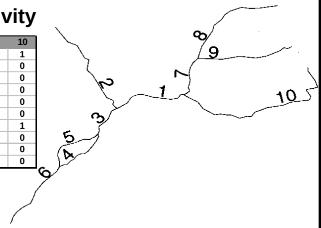
- **nódy**
- **hrany** (spoje)
- **planární graf**

**Deskriptory sítě**:

- deskriptory sítě jako celku
- deskriptory relací jednotlivých segmentů sítě.

## Konektivita a matice konektivity

| ID | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 2 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 7 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| 10 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |




Základním topologickým aspektem sítě je způsob propojení jednotlivých segmentů – **konektivita**.

Vhodným nástrojem používaným k charakterizování konektivity je matice konektivity.

Je to matice čtvercová, binární, symetrická o *n* řádcích (sloupcích), kde *n* je počet segmentů sítě. Jednička v matici značí, že dva příslušné segmenty jsou bezprostředně spojeny. Na hlavní diagonále matice jsou nuly.

## Konektivita a matice konektivity

Matice konektivity shrnuje informaci o tom, které segmenty sítě spolu souvisí (jsou bezprostředně spojeny). Lze však charakterizovat i úroveň konektivity sítě jako celku.

Pro fixní počet vrcholů má síť s větším počtem spojů lepší konektivitu. Dále existuje minimální počet spojů, který zajišťuje spojení všech vrcholů.

Bude-li *v* – počet vrcholů sítě, *e* - počet hran sítě potom:

$$e_{\min} = v - 1$$

**Minimálně propojená síť** (Minimally conneted network - MCN) – odstraníme–li jakoukoliv jednu hranu, síť se rozpadne na dva subsystémy.

## Gama index

Podobně lze pro daný počet vrcholů vytvořit **maximální počet hran**, které spojují všechny vrcholy. Tedy maximální počet hran v síti o v vrcholech:

$$e_{\max} = 3(v - 2)$$

Jednoduchou charakteristikou konektivity sítě je **Gama index (γ)** – je definován jako poměr aktuálního a maximálního počtu hran sítě.

$$\gamma = \frac{e}{e_{\max}}$$

## Alfa index

Další jednoduchou charakteristikou konektivity sítě je **počet okruhů.** Výskyt okruhů v síti značí možnost dostat se z jednoho místa do jiného alternativními cestami. Síť s minimální konektivitou nemá žádný okruh.

Počet okruhů lze zjistit tak, že od aktuálního počtu hran v síti odečteme počet hran potřebný pro minimálně propojenou síť (MCN), tedy *e-(v-1)* nebo *e-v+1*.

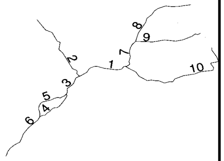
Obdobně pro daný počet vrcholů je **maximální počet okruhů** roven 2v-5.
S oběma uvedenými počty okruhů lze vytvořit poměr aktuálního počtu k počtu maximálnímu – tedy tzv. **alfa index**

$$\alpha = \frac{e - v + 1}{2v - 5}$$

Pomocí alfa indexu můžeme snadno porovnat dvě sítě.

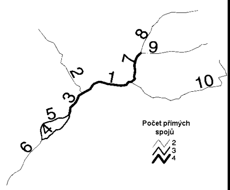
## Dostupnost sítě (Acccessibility)

| ID | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | SUMA |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 4 |
| 2 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| 3 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 4 |
| 4 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 3 |
| 5 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 3 |
| 6 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 2 |
| 7 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 4 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 2 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 2 |
| 10 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 2 |




Jedná se o charakteristiku jednotlivých **vrcholů** či **hran** sítě. Popisuje jejich dostupnost v rámci sítě. Další výklad se týká dostupnosti hran sítě, obdobné vztahy lze definovat i pro vrcholy.

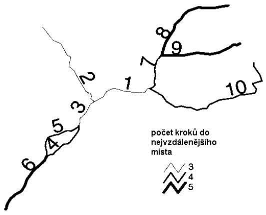
Jednoduchým ukazatelem dostupnosti hrany v rámci sítě je, s kolika jinými hranami daná linie přímo souvisí. Tuto informaci lze vyčíst z binární matice konektivity, pokud tuto doplníme řádkovým součtem.



## Dostupnost sítě

Výše uvedená charakteristika však může být zavádějící, protože nebere v úvahu relativní (topologickou) polohu hrany v rámci sítě. Hrana může mít i pouze jeden či dva spoje, přesto může být snadno dostupná, protože se nachází uprostřed sítě (a naopak).

Relativní pozici každé hrany v rámci sítě lze zjistit např. pomocí počtu hran, kterými se lze z daného spoje dostat do nejvzdálenějšího místa sítě.

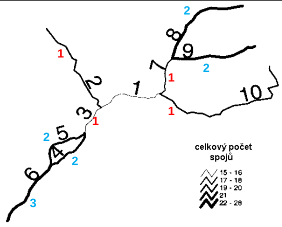



| ID | počet přímých spojů | počet kroků k dosažení nejvzálenějšího místa |
|---|---|---|
| 1 | 4 | 3 |
| 2 | 2 | 3 |
| 3 | 4 | 3 |
| 4 | 3 | 4 |
| 5 | 3 | 4 |
| 6 | 2 | 5 |
| 7 | 4 | 4 |
| 8 | 2 | 5 |
| 9 | 2 | 5 |
| 10 | 2 | 4 |

**Diametr (poloměr) sítě** – je to jedna (1) plus největší počet hran nutných k dosažení nejvzdálenějšího místa v síti.

## Dostupnost sítě

| ID | počet přímých spojů | celkový počet přímých a nepřímých spojů |
|---|---|---|
| 1 | 4 | 15 |
| 2 | 2 | 19 |
| 3 | 4 | 16 |
| 4 | 3 | 21 |
| 5 | 3 | 21 |
| 6 | 2 | 28 |
| 7 | 4 | 18 |
| 8 | 2 | 25 |
| 9 | 2 | 25 |
| 10 | 2 | 20 |




Přímé spoje pro segment 1: 4
Nepřímé spoje pro segment 1: 11

Kvalitu spojení dvou hran (vrcholů) definuje počet hran mezi nimi. Spojení mohou být přímá a nepřímá.

Každou hranu lze popsat **počtem přímých a nepřímých spojů**, které jsou třeba, aby tato byla spojena se všemi hranami ostatními.

Nepřímé spoje lze vážit počtem kroků. Zřejmě platí, že čím větší je celkový počet potřebný počet spojů, tím hůře dostupná je daná hrana. Celkový počet spojů (přímých i nepřímých) je mírou dostupnosti.