Statistická analýza liniových prvků

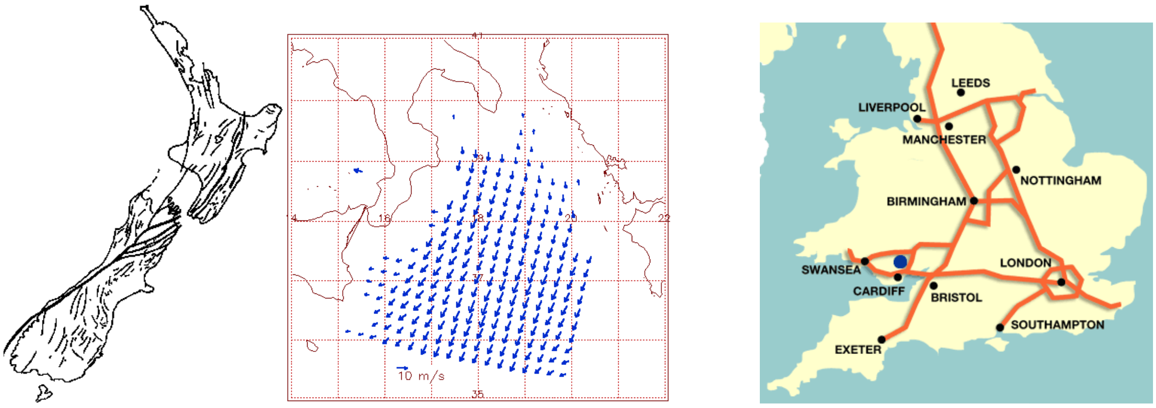
Linie mohou na mapách reprezentovat dva příbuzné objekty:

- **Vlastní linie** - reprezentují a lokalizují skutečně lineární geografické fenomény (řeky, silnice, potrubí)
- **Hrany** - rozdělují plochy a povrchy (hraniční linie, lomové linie). Hrany nemají šířku.

Problémy prezentace „přirozených linií“ v prostředí GIS jsou spojeny především s procesy generalizace a zjednodušení průběhu. Linie je prezentována 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.

Pro analýzu linií jsou vedle délky významné také atributy jako orientace, směr či spojení. Existence spojení mezi soustavou bodů, které tvoří linii, znamená, že lokace (body) na sobě nejsou nezávislé, ale jsou spojené v určitém směru. Body spojené v určitém pořadí musí zachovávat tuto posloupnost.



Obr. 3.1 Liniové prvky na digitální mapě - prosté linie, trajektorie, síť

Linie mohou v GIS vystupovat **na třech úrovních**, které představují jistou hierarchii (obr. 3.1):

1. „**Prosté“ linie** – např. zlomy – lze určit jen délku a orientaci. Může existovat jako jednoduchá 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, …

Geometrické charakteristiky – linie může být prezentována jako:

- **Jednoduchá** spojnice – pouze dvou bodů (koncový a počáteční – délka je Euklidovská vzdálenost
- Posloupnost několika liniových segmentů - **řetězec**

Příklady analýzy prostorových vazeb liniových prvků:

- analýza převládající orientace, průměrné délky spoje,
- charakterizování liniových vzorků – „uspořádání sítí“
- dopravní dostupnost
- gravitační modely
- hledání optimální trasy

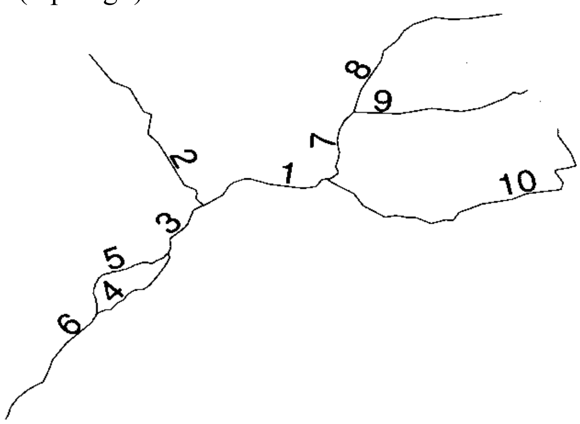
Prostorové atributy liniových prvků

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 – vektor pole větru

Topologie (sítí)
Výše uvedené atributy linií lze vyjádřit i pro jednotlivé segmenty sítě či pro celou síť jako celek (průměrná délka sítě, převládající orientace či směr segmentů sítě). Vedle toho jsou pro charakterizování sítí důležité atributy popisující jejich strukturu a uspořádání jako celek a dále popisují vztahy segmentů uvnitř sítě (topologii).



Obr.3.2 Příklad sítě

Tabulka 3.1 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

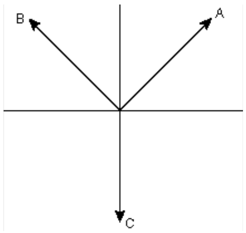
Základním topologickým aspektem sítě je způsob propojení jednotlivých segmentů – tedy její **konektivita**. Tradičním nástrojem používaným k charakterizování konektivity je **matice konektivity**. Je to matice čtvercová, binární, symetrická o *n* řádcích (sloupcích), kde *n* je počet segmentů sítě. Jednička v matici značí, že dva příslušné segmenty jsou bezprostředně spojeny. Na hlavní diagonále matice jsou nuly.

Směrová statistika (Directional statistics)

Topologii sítě lze charakterizovat jednoduchými mírami. Takovou je např. poměr mezi skutečnou délkou linie a spojnicí počátečního a koncového bodu. Tato charakteristika se určuje křivost linie (**sinusoity**). Čím větší číslo, tím větší křivost.
Směr linie – vizuální hodnocení směru linií lze provést přidáním šipek. Např. u pole větru je možné odhalit strukturu proudění v celé oblasti.

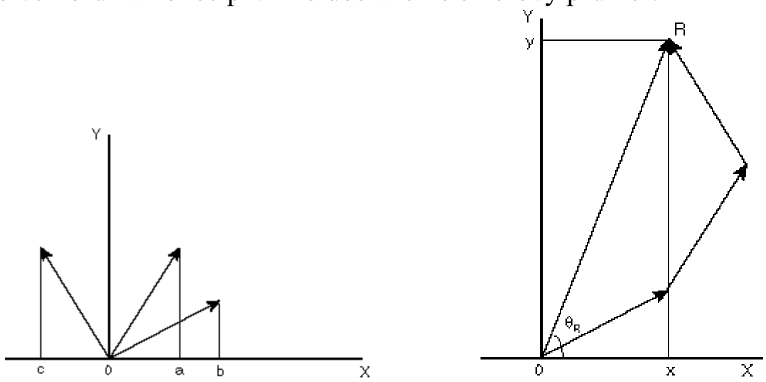
Směrový průměr (directional mean).

Využití klasických měr popisné statistiky pro charakterizování směru a orientace linií je nevhodné (viz. obr. 3.3).



Obr. 3.3 Problém popisné statistiky při určování charakteristik směru linie

Jak je patrné z obrázku, aritmetický průměr dvou vektorů s úhly 45 a 315 stupňů dává 180 (jižní směr), avšak měl by být 0 stupňů (severní směr). Průměrný směr je však nutné určit vektorovým součtem či tzv. **směrovým průměrem** (directional mean). Protože pracuje se směrem (úhlem) a ne s délkou, je možné ho prezentovat na základě jednotkových vektorů. Vektorovým součtem – přidáním počátku druhého vektoru na konec prvního dostaneme směrový průměr.



Obr. 3.4 Koncept směrového průměru

Směr výsledného vektoru lze získat také z následujícího vztahu:

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

kde *oy* je suma délek vektorů ve směru osy *y* a *ox* suma délek vektorů ve směru osy *x*. Protože všechny vektory jsou jednotkové, délka ve směru osy *y* je v podstatě sin úhlu a délka na ose *x*. je cosinus úhlu. Potom, jsou-li vektory označeny *a, b, c* a odpovídající úhly θ_a, θ_b, θ_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 θ_v, výsledný vektor OR má úhel θ_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.

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

1. čitatel i jmenovatel v tan θ_R jsou oba kladné – není nutná žádná úprava (vektor leží v 1. kvadrantu)
2. čitatel je kladný jmenovatel záporný – směrový průměr bude 180 - θ_R, (vektor leží v 2. kvadrantu)

3. čitatel i jmenovatel v *tan* θ_R jsou oba záporné – směrový průměr bude 180 + θ_R, (vektor leží v 3. kvadrantu)
4. čitatel je záporný, jmenovatel kladný – směrový průměr bude 360 - θ_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 je potom *arctan*.

Směrový rozptyl (Circular variance)

Stejně jako v případě klasické popisné statistiky je charakterizování souboru prvků pouze měrou úrovně, kterou je výše uvedený směrový průměr, je 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*. Tedy 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 (circular variance) S_v se potom vypočte:

$$S_v = 1 - OR/n$$

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

Nebude probírána síťová analýza – ta vyžaduje speciální prostředí a nástroje (maticový počet) i speciálně upravená vstupní data.
Základní pojmy používané v síťové analýze: **nódy** a **hrany** (spoje), jejich počet také charakterizuje síť. Ke křížení dvou a více hran dochází pouze ve vrcholu (planar graph topology)

Deskriptory sítě lze rozdělit do dvou skupin:
1. Deskriptory sítě jako celku
2. Deskriptory relací jednotlivých segmentů sítě.

Konektivita a matice konektivity

Matice konektivity (tab. 3.1) 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.

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 **Gamma index (γ)** – je definován jako poměr aktuálního a maximálního počtu vrcholů sítě.

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

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)

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

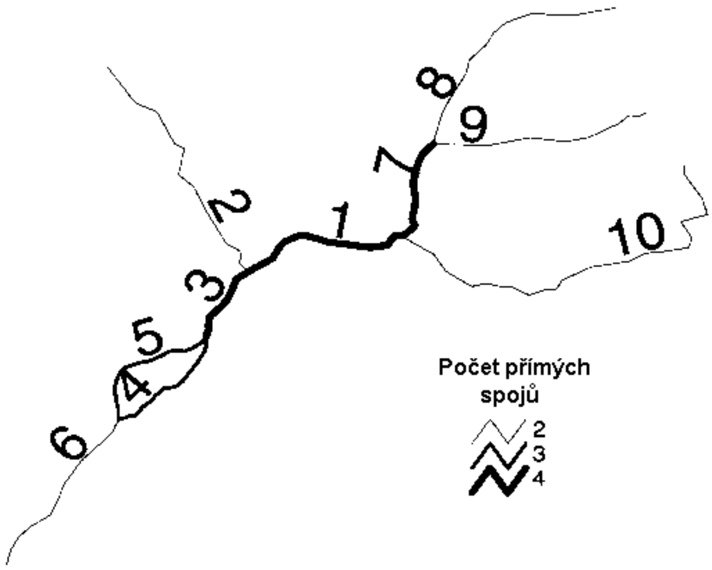
Jednoduchým ukazatelem dostupnosti hrany v rámci sítě je, s kolika jinými hranami daná linie přímo souvisí. Tuto informaci lze vyčíst z binární matice konektivity, pokud tuto doplníme např. řádkovým součtem.

Tabulka 3.3 Matice konektivity a dostupnost hran v rámci sítě

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

Tabulka 3.3 Charakteristiky dostupnosti sítě (viz. obr 3.5, 3.6, 3.7)

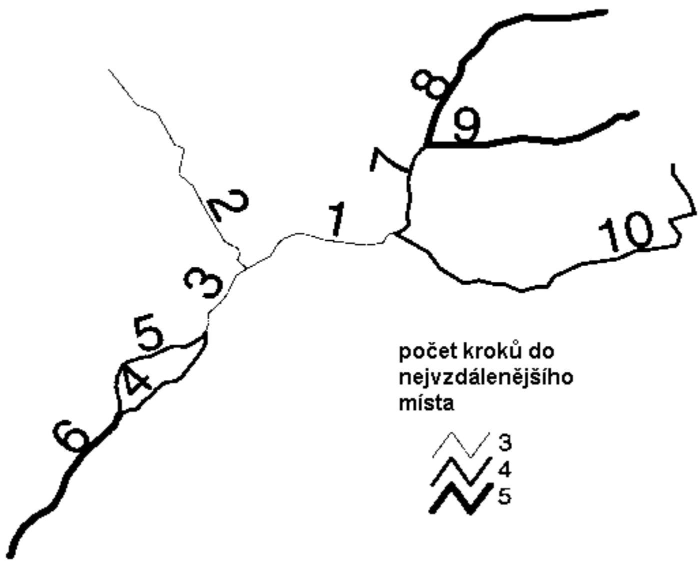
ID	počet přímých spojů	počet kroků k dosažení nejvzdálenějšího místa	celkový počet přímých a nepřímých spojů
1	4	3	15
2	2	3	19
3	4	3	16
4	3	4	21
5	3	4	21
6	2	5	28
7	4	4	18
8	2	5	25
9	2	5	25
10	2	4	20



Obr. 3.5 Dostupnost jednotlivých segmentů sítě charakterizovaná počtem přímých spojů

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).

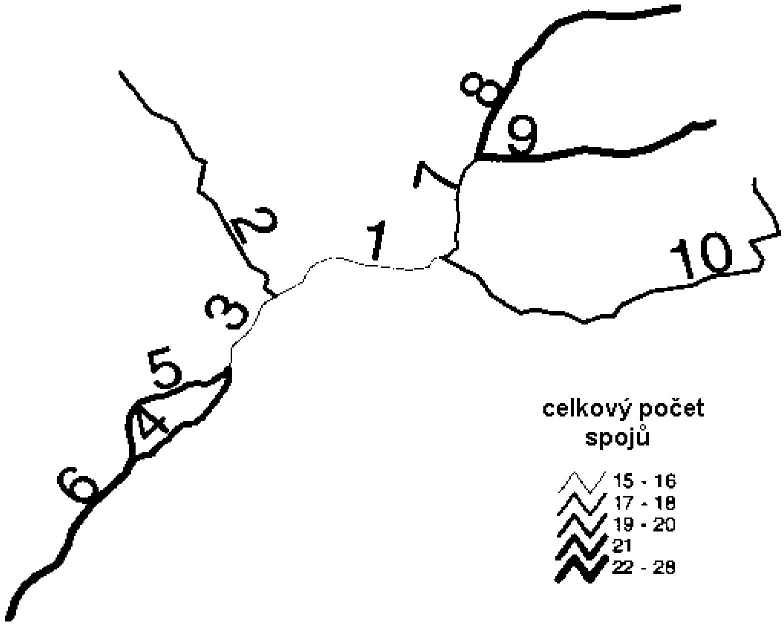
Relativní pozici každé hrany v rámci sítě lze zjistit např. pomocí počtu hran, kterými se lze z daného spoje dostat do nejvzdálenějšího místa sítě.



Obr. 3.6 Dostupnost jednotlivých segmentů sítě charakterizovaná počtem kroků nutných k dosažení nejvzdálenějšího místa sítě.

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

Kvalitu spojení dvou hran (vrcholů) definuje počet hran mezi nimi. Spojení mohou být přímá a nepřímá. Tedy počet přímých a nepřímých spojů, které jsou třeba, aby byla daná hrana spojena se všemi hranami ostatními. Nepřímé spoje lze vážit počtem kroků. Zřejmě platí, že čím větší je celkový potřebný počet spojů, tím hůře dostupná je daná hrana. Celkový počet spojů (přímých i nepřímých) je mírou dostupnosti.



Obr. 3.7 Dostupnost jednotlivých segmentů sítě charakterizovaná počtem přímých a nepřímých spojů nutných k dosažení jakéhokoliv místa v síti