

# Spatial similarity calculation - Räumliche Ähnlichkeitsberechnung



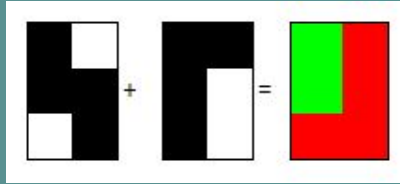
# Warum?

- entdecken der relevantesten Datensätze in den ständig wachsenden globalen SDI (Spatial Data Infrastructures)
- das Finden ähnlicher Objekte wird immer populärer in geographischen Informationssystemen
- zentraler Faktor in der intelligenten Automated Map Generalization einer Karte
  - offensichtlich lässt die Ähnlichkeit nach, je stärker generalisiert wurde
  - keine Erfolge im Beschreiben des Verhältnisses (Ähnlichkeitsgrad)
  - Software weiß nicht bis zu welchem Ausmaß eine Karte generalisiert werden soll oder wann der Algorithmus zur Generalisierung terminieren soll

# Eigene Idee zu Ähnlichkeitsberechnung von Vektordaten

- Umwandlung von Vektor in Rasterdaten
- Ähnlichkeit = gleiche Pixel/Vereinigung der Pixel beider Objekte
- Beispiel: 2 gleiche Pixel von insgesamt 6

➤ Ähnlichkeit:  $2/6 = 1/3$



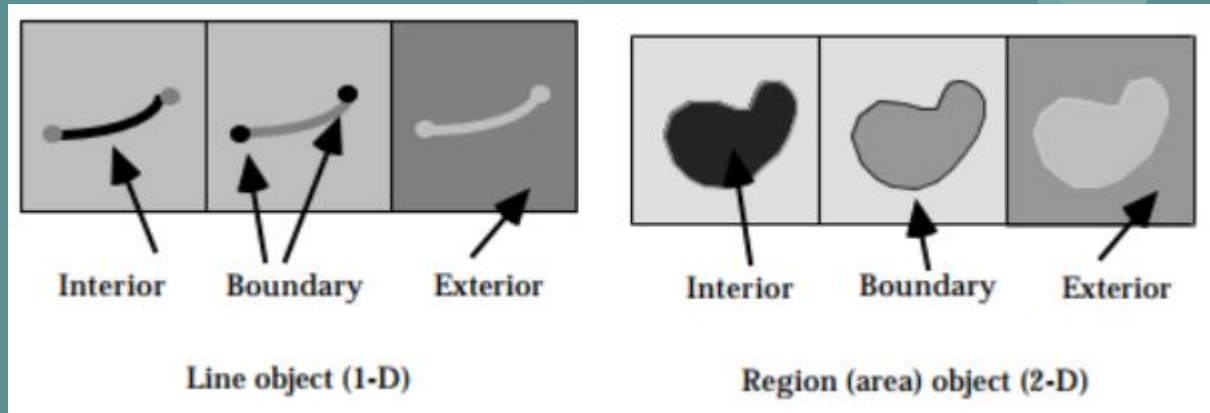
- Problem: boolesche Operationen können die Komplexität räumlicher Ähnlichkeit häufig nicht adäquat beschreiben

# Kategorien zur Messung der räumlichen Ähnlichkeit

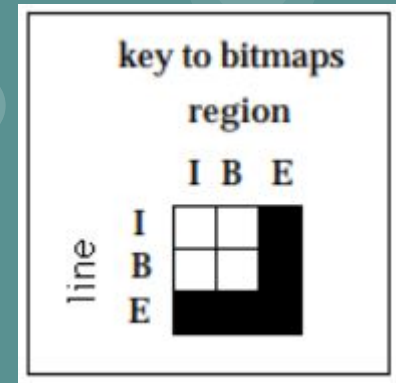
- Topologie
  - generelle Struktur und Beziehung der Objekte
  - Benutzung von topologischen Primitiven: Inneres, Äußeres, Rand
- Richtung
  - qualitative räumliche Beschreibung wo ein Objekt gelegen ist in Relation zu anderen Objekten
  - Bsp: nördlich, südöstlich, davor, hinten-rechts
- Metrische Distanz
  - messbare Entfernung von räumlichen Objekten
- Attribute
  - jede nicht räumliche Eigenschaft -> z.B. Art der Landnutzung

# Beispiel Topologie: 9-intersection model

- beschreibt binäre topologische Beziehungen hinsichtlich Schnittpunkte von Innerem, Äußerem und Rand von zwei räumlichen Objekten (hier Linie und Polygon)

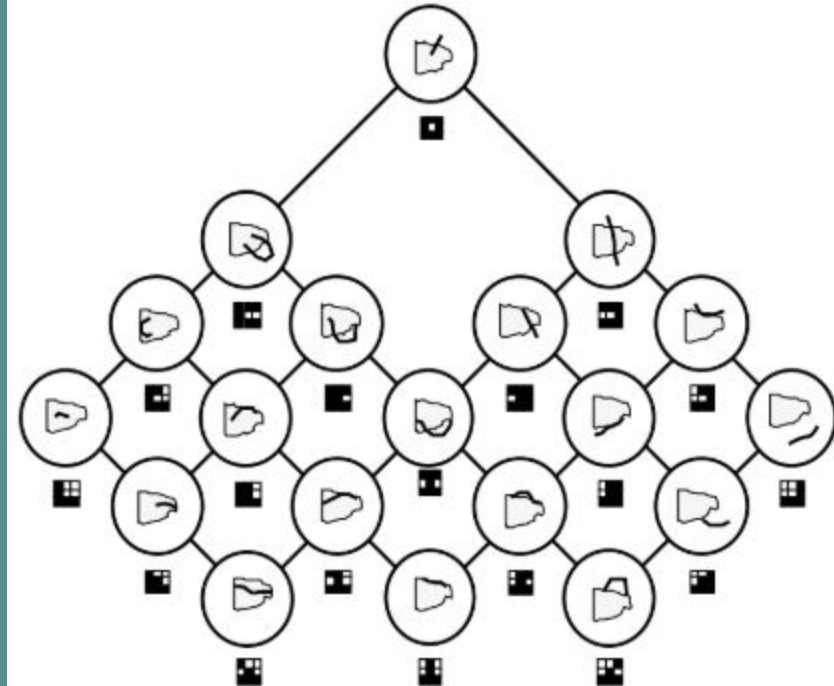


- zur Beschreibung der Schnittpunkte: 3x3 Bitmap



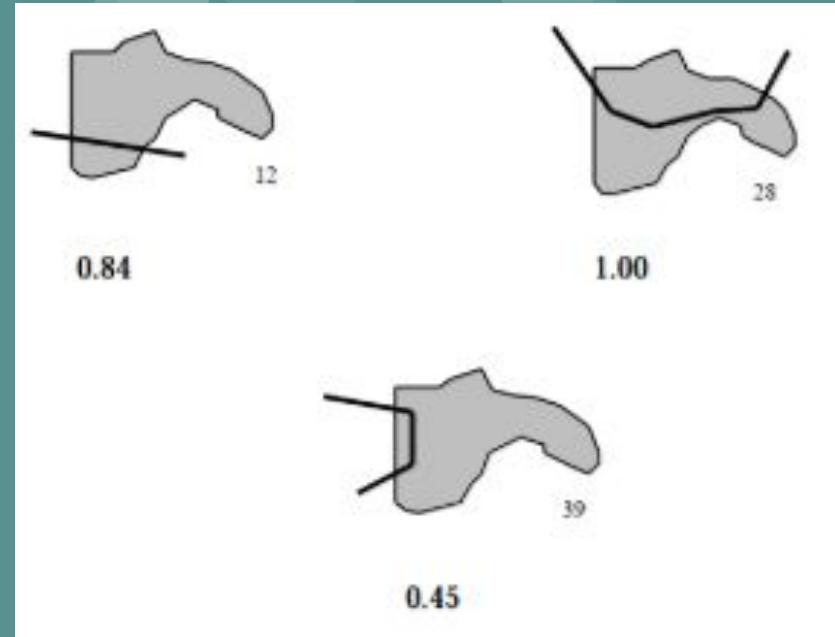
# Beispiel Topologie: 9-intersection model

- theoretisch  $2^9$  (=512) Möglichkeiten, aber im zweidimensionalen faktisch **19 Möglichkeiten** (siehe Diagramm)
- Kante verbindet zwei Situationen, wenn nur ein “Schnittpunkt” verschieden ist
  - je mehr Kanten, desto verschiedener ist die Topologie zweier Objekte



# Einfluss der Geometrie

- 40 Probanden wurden gebeten ihre Zustimmung zu dem Satz: “Straße kreuzt den Park” zu bewerten für topologisch identische Situationen
  - sehr verschiedene Bewertungen schließen darauf, dass auch die Geometrie stark die Ähnlichkeit beeinflusst



# Automated Map Generalization



Maßstab: 1:10.000

Automatisch?



Maßstab: 1:20.000

- räumliche Ähnlichkeitsberechnung von enormer Bedeutung für Automated Map Generalization
- kann eine Karte mit mehreren Skalen erstellen
- Grundlage: nur einer gespeicherten Karte (die des größtmöglichen Maßstabs)



# Automated Map Generalization

- der Algorithmus enthält üblicherweise einen oder mehrere Schwellenwerte
- sobald Ähnlichkeitsberechnung zwischen den beiden Karten präzise möglich sind, sind Verfahren der Berechnung dieser Schwellenwerte möglich
- wenn diese bekannt sind, kann der Algorithmus ohne Parameter funktionieren und wird somit **vollautomatisch**



räumliche Ähnlichkeitsberechnungen ist von enormer Bedeutung für automated map generalization

# Räumliche Ähnlichkeitsberechnung

- ist zu unterscheiden von textueller Ähnlichkeitsberechnung (“matching keywords”), da:
  - involvieren verschiedenste Elemente (wie räumliche Beziehung, räumliche Verteilung, geometrische und thematische Attribute, ...)
- die Definition von “Ähnlichkeit” (bzw. “Similarity”) ist stark **anwendungsorientiert**
- das Spektrum jeder räumlichen Ähnlichkeitsberechnung reicht von 0 (sich gar nicht ähneln) bis 1 (gleich)

# Räumliche Ähnlichkeitsberechnung (Vektor-Daten)

**Beispiel:** Für Objekte A1 und A2 soll die Ähnlichkeit quantifiziert werden

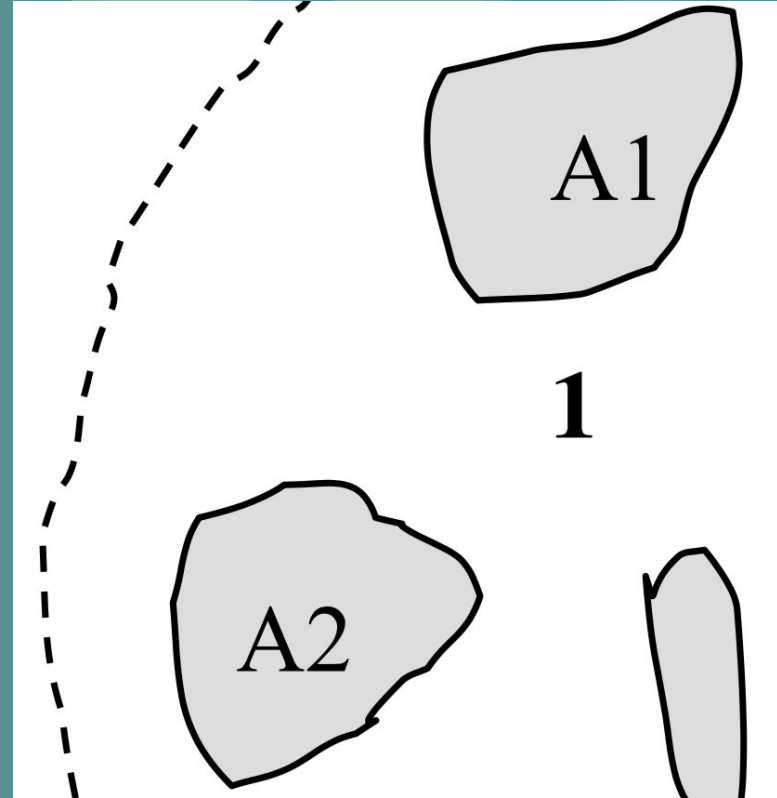
**Methode:** Den Objekten werden Eigenschaften ("properties") zugewiesen, welche verglichen werden. Die hier ausgewählten properties sind:  $P = \{\text{area, shape, arability}\}$

Wertebereich area: 1 = small, 2 = big, 3 = large

shape: Wert entspricht Anzahl an Kanten

arability: Ja = 2, Nein = 1      => arability = bebaubar?

$P1 = \{2, 6, 1\}$  und  $P2 = \{2, 9, 1\}$



## Räumliche Ähnlichkeitsberechnung zwischen $A_1$ und $A_2$ bezogen auf property $P_i$ :

$$Sim_{A_1, A_2}^{P_i} = f_i(p_{1i}, p_{2i})$$

- jede property hat ihre eigene Ähnlichkeitsfunktion (im Beispiel:  $f_1$ ,  $f_2$  und  $f_3$ )
- dessen Wert liegt zwischen 0 und 1
- ⇒ Beispiel mit property “area”:

$A_1$  und  $A_2$  sind bezüglich  $P_1$  gleich  
(sind 100% ähnlich).

$$Sim_{A_1, A_2}^{P_1} = f_1(2, 2) = 1$$

Beispiel mit property “area”

Beispiel mit  
property  
“shape”

$$Sim_{A_1, A_2}^{P_2} = f_2(6, 9) = \frac{\vee(p_{12}, p_{22})}{(p_{12}, p_{22}) / 2} = \frac{\vee(6, 9)}{(6 + 9) / 2} = 0.8$$

# Räumliche Ähnlichkeitsberechnung zwischen $A_1$ und $A_2$ :

- nicht auf nur eine property bezogen, sondern auf alle
- jede property bekommt eine Gewichtung, dessen Summe 1 ergeben muss
- im Beispiel:  $w = \{0.3, 0.6, 0.1\}$

$P = \{\text{area, shape, arability}\}$

$P1 = \{2, 6, 1\}$  und  $P2 = \{2, 9, 1\}$

$$Sim(A_1, A_2) = \sum_{i=1}^n w_i Sim_{A_1, A_2}^{P_i}$$

Berechnungen aus  
letzter Folien

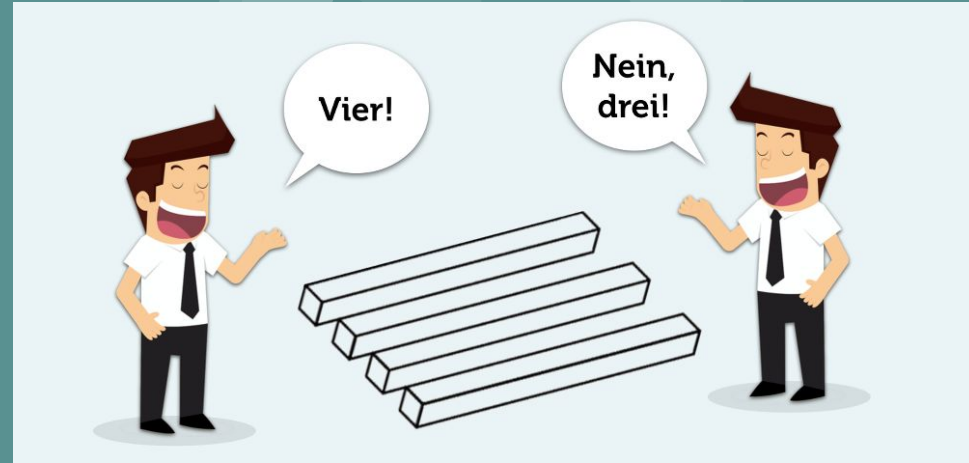
Beispiel mit Objekt A1 und A2:

$$Sim(A_1, A_2) = \sum_{i=1}^3 w_i Sim_{A_1, A_2}^{P_i} = 1 \times 0.3 + 0.8 \times 0.6 + 1 \times 0.1 = 0.88$$

Objekte A1 und A2 sind sich zu 88% ähnlich.

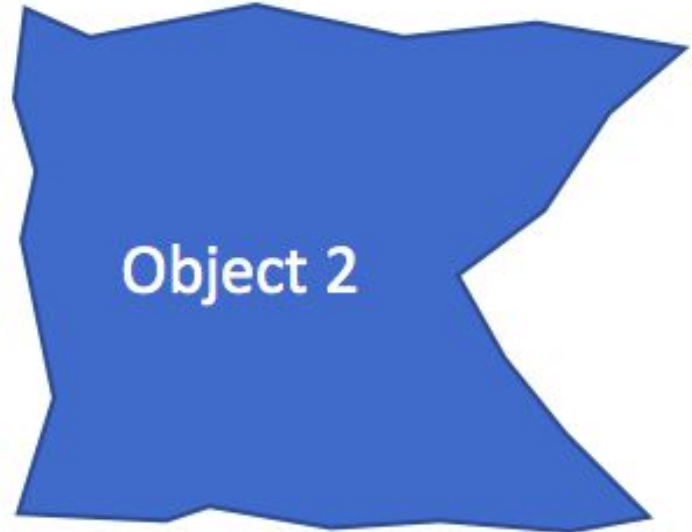
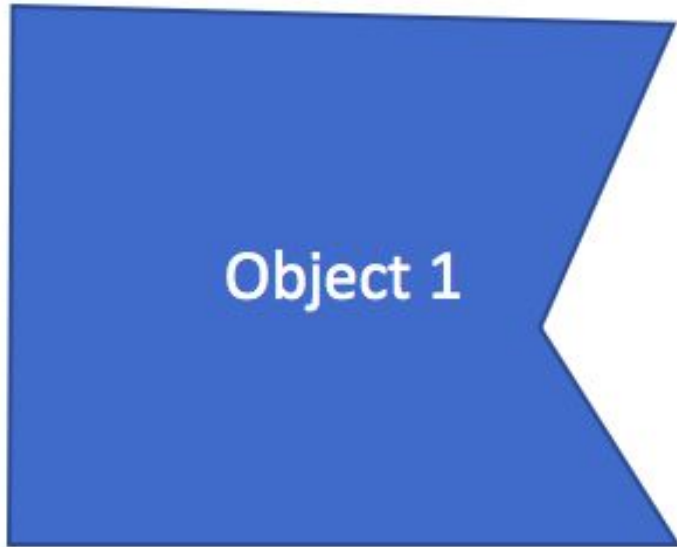
# Die Wahl der properties

- ein subjektiver Prozess
- können räumliche oder auch thematische properties enthalten
- die Gültigkeit der Definition hängt vom Ermessen des Nutzers ab



# Eigene Idee: Warum bestimmte properties täuschen können...

- die meisten Menschen würden räumliche Ähnlichkeit von Object 1 und Object 2 wohl auf mindestens 90 % schätzen, **ABER...**



# Eigene Idee: Warum bestimmte properties täuschen können...

- bzgl. der property 1 und 3 sind Object 1 und Object 2 gleich.

ABER: Objekt 2 hat deutlich mehr Ecken:

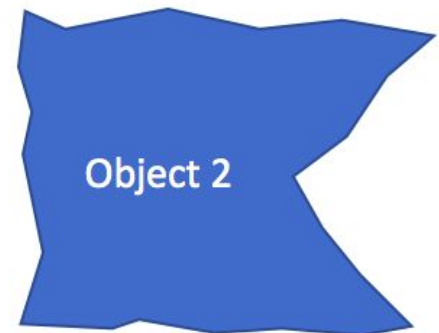
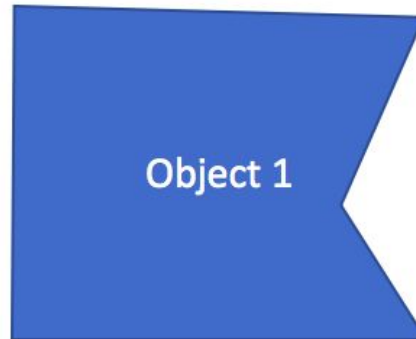
P1 = {2,5,1} and P2 = {2,21,1}

P = {area, shape, arability}

$$Sim_{A_1, A_2}^{P_2} = f_2(5, 21) = \frac{v(p_{12}, p_{22})}{(p_{12}, p_{22}) / 2} = \frac{v(5, 21)}{(5 + 21) / 2} = 0.38$$

=> 
$$Sim(A_1, A_2) = \sum_{i=1}^3 w_i Sim_{A_1, A_2}^{P_i} = 1 \times 0.3 + 0.38 \times 0.6 + 1 \times 0.1 = 0.628$$

- die räumliche Ähnlichkeit beträgt nach mit unseren drei properties 62,8 %
- deutlich weniger als erwartet (> 90%)

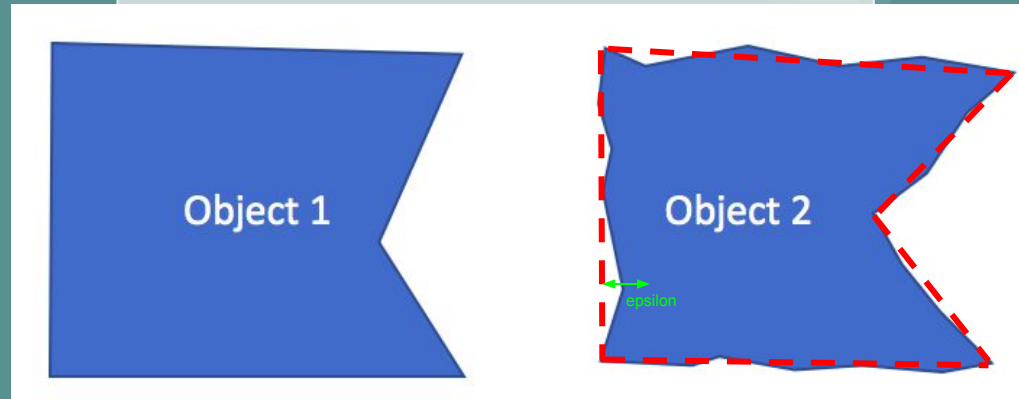
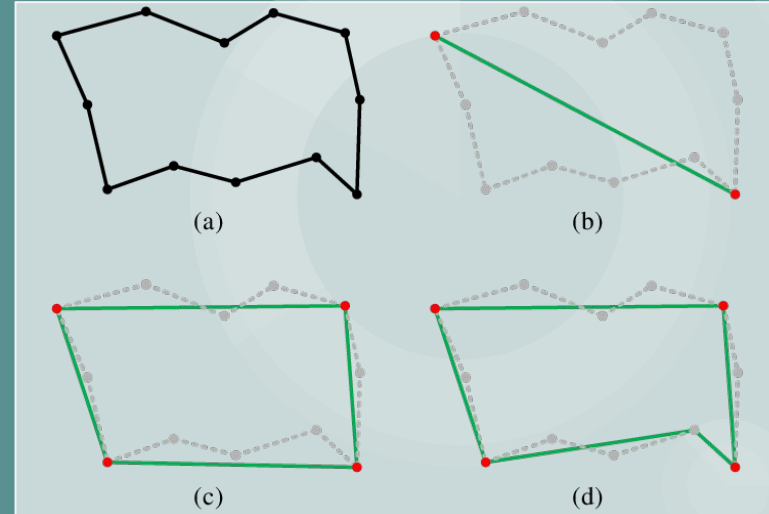




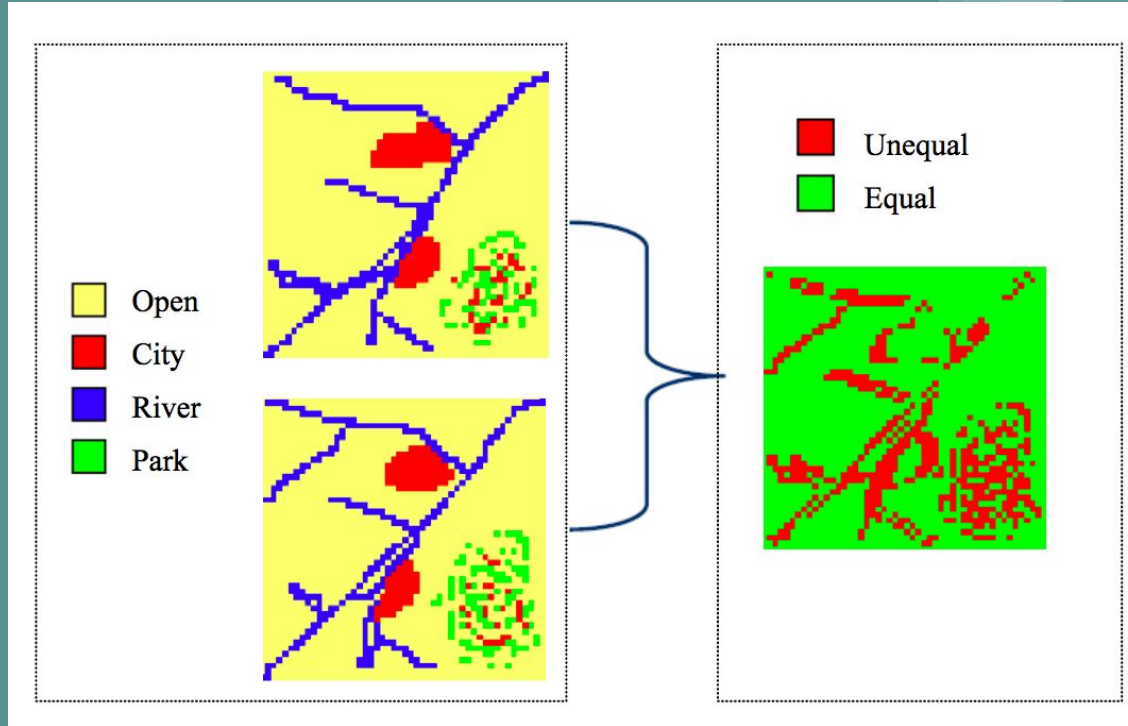
# Eigene Idee: Warum bestimmte properties täuschen können...

## Unsere Idee der Anwendung:

1. wenn Douglas-Peucker-Algorithmus bei Objekt mindestens eine Ecke bei ausreichend geringem Epsilon (Interpretationssache!) entfernt, wende ihn an
2. zähle erneut die Ecken und berechne die Ähnlichkeit der Objekte => im Beispiel 100%



# Einblick in Ähnlichkeit von Rasterdaten



# Methoden

Annahmen:

- beide Karten beschreiben die gleiche räumliche Fläche
- wir wollen die Ähnlichkeit einer diskreten Variable vergleichen (Bsp.: Versiegelung)

1. **Jaccard Index:** Schnittmenge/Vereinigung

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

2. **Sørensen–Dice Koeffizient:** 2\*Schnittmenge/Addition der Anzahl der Elemente

beider Objekt

Jaccard = SDC/(2-SDC)

$$DSC(A, B) = \frac{2|A \cap B|}{|A| + |B|}$$

3. **Hamming-Distanz:** Wie viele Stellen muss ich verändern, um von Objekt A zu Objekt B zu kommen? Bsp.:  $HD(110, 111) = 1$        $HD(101, 010) = 3$

# Danke für eure Aufmerksamkeit!

## Unsere Quelle:

- Spatial relations between lines and regions  
(<https://pdfs.semanticscholar.org/62e2/96f9e81f175751907f3e2c919dffdcaae56b.pdf> )
- Relation between spatial similarity and map scale change in mulit-scale maps  
([https://ac.els-cdn.com/S1674984715000191/1-s2.0-S1674984715000191-main.pdf?tid=ace53621-8259-44eb-bd32-e80ecb415137&acdnat=1538735023\\_3c975278e6a2171fa660b98d014922d8](https://ac.els-cdn.com/S1674984715000191/1-s2.0-S1674984715000191-main.pdf?tid=ace53621-8259-44eb-bd32-e80ecb415137&acdnat=1538735023_3c975278e6a2171fa660b98d014922d8) )
- Spatial Similarity Relations in automated map generalization  
(<https://uwspace.uwaterloo.ca/handle/10012/8317> )
- Directional relations  
([https://link.springer.com/referenceworkentry/10.1007/978-3-319-23519-6\\_1539-1](https://link.springer.com/referenceworkentry/10.1007/978-3-319-23519-6_1539-1) )
- How to statistically compare two maps  
([https://www.researchgate.net/post/How to statistically compare two maps](https://www.researchgate.net/post/How_to_statistically_compare_two_maps) )
- Map Comparison Kit (<http://mck.riks.nl/> )

# Beispiel metrische Distanz: Geohashes

- Idee: Standorte teilen wird immer populärer
- Kodieren der Koordinaten in einen String
  - z.B. 51.969173, 7.59578 -> u1jrsuen6
- je länger der gemeinsame Prefix ist, desto näher sind sich die beiden Objekte

## Grenzen:

- Randfälle
- nicht linear, verschieden große Fläche eines Geohash Gebietes am Äquator im Vergleich zu einem am Nordpol

