

Ареальность во Всемирном атласе языковых структур



Александр Пиперски
(РГГУ / НИУ ВШЭ)

20.11.2020

Малый

АТЛАС МИРА



 **АИКА**
ИЗД-ВО «КАРТОГРАФИИ»

 **АСТ** ПРЕСС



МАЛЫЙ
АТЛАС
СССР

Feature 85A: Order of Adposition and Noun Phrase



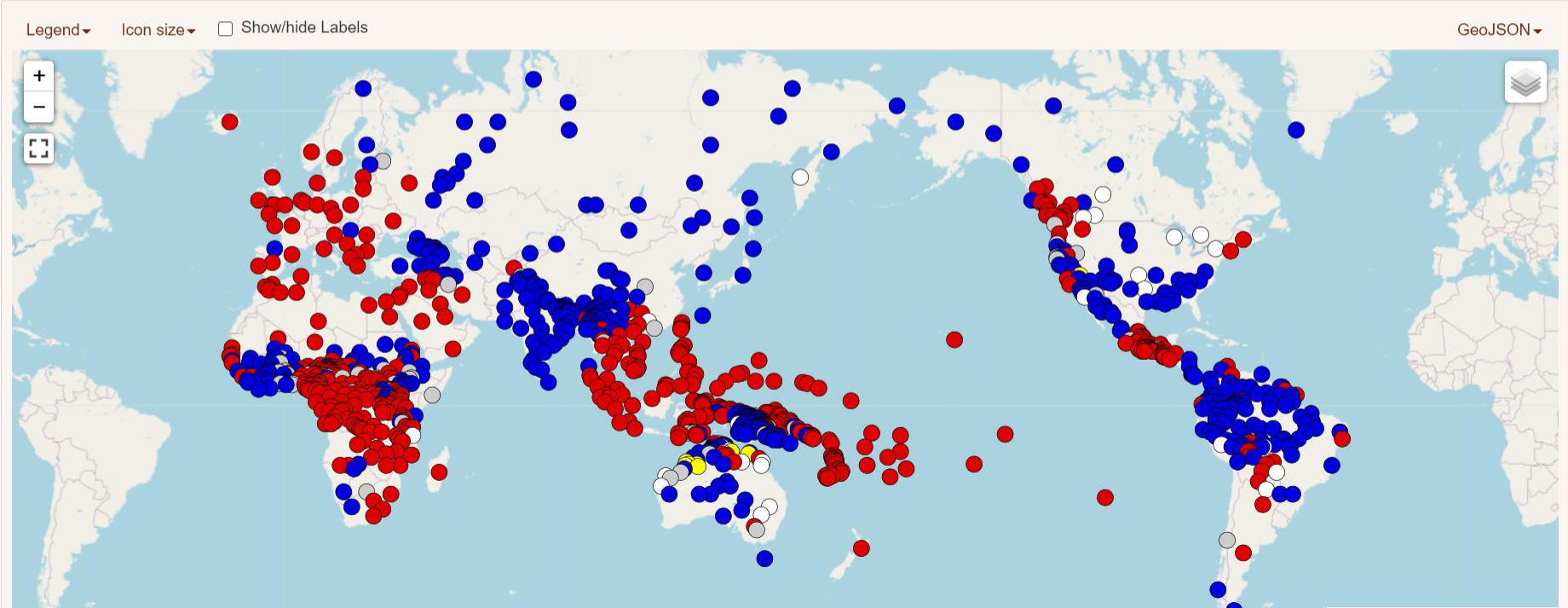
This feature is described in the text of chapter 85 [Order of Adposition and Noun Phrase](#) by Matthew S. Dryer [cite](#)

You may combine this feature with another one. Start typing the feature name or number in the field below.

Submit

Values

●	Postpositions	577
●	Prepositions	511
●	Inpositions	8
●	No dominant order	58
○	No adpositions	30



Feature 3A: Consonant-Vowel Ratio



This feature is described in the text of chapter 3 by [Ian Maddieson](#)

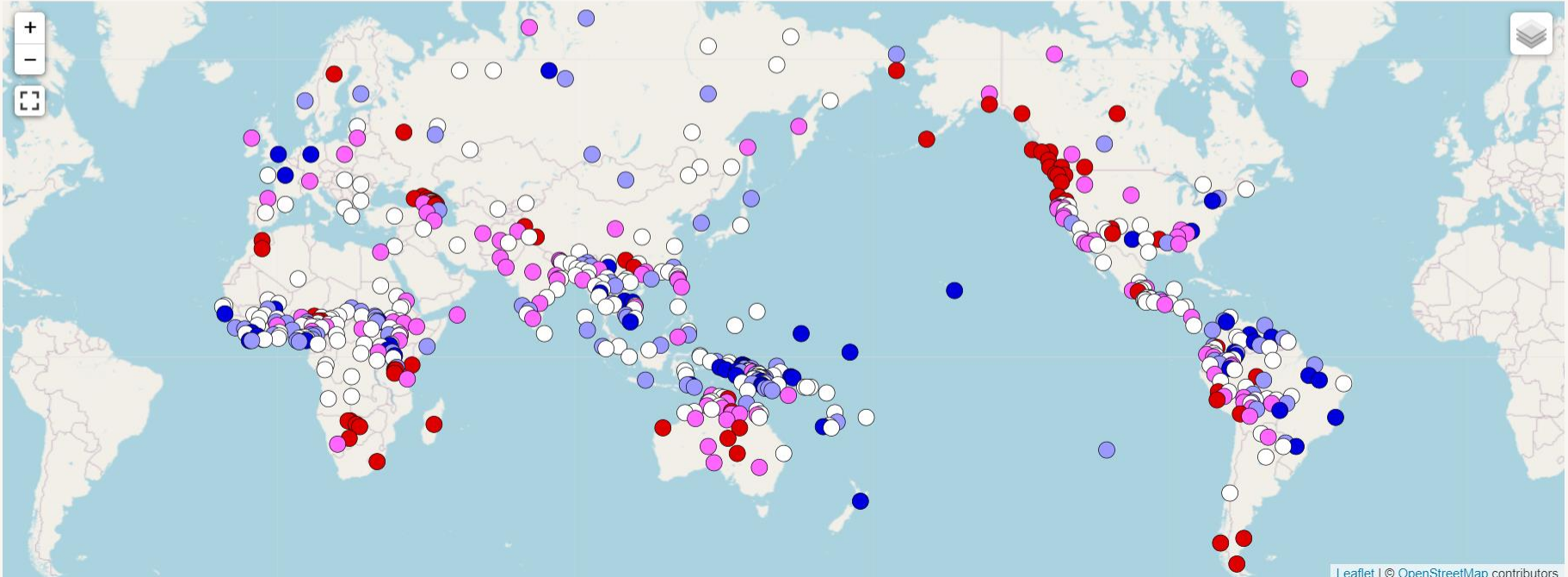
You may combine this feature with another one. Start typing the feature name or number in the field below.

Values

	Low	58
	Moderately low	101
	Average	234
	Moderately high	102
	High	69

Legend Show/hide Labels

GeoJSON



Feature 38A: Indefinite Articles

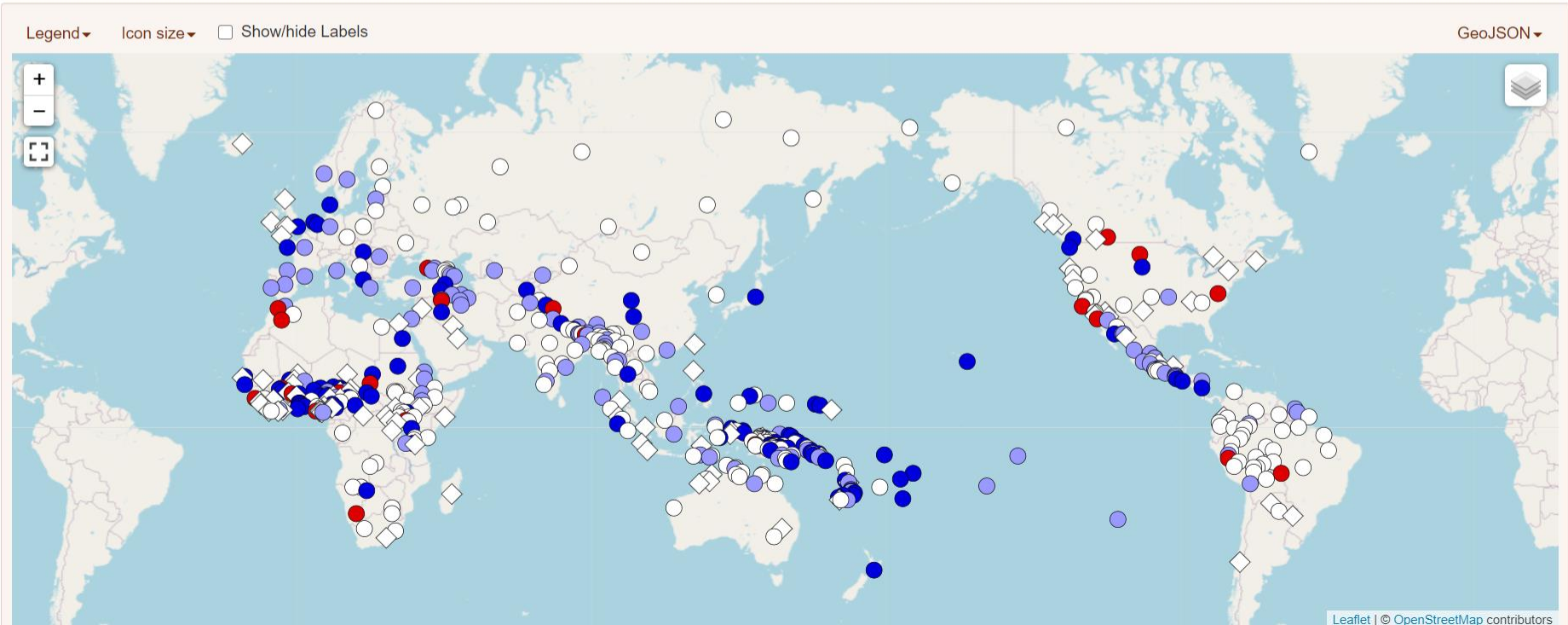


This feature is described in the text of chapter 38 [Indefinite Articles](#) by [Matthew S. Dryer](#) [cite](#)

You may combine this feature with another one. Start typing the feature name or number in the field below.

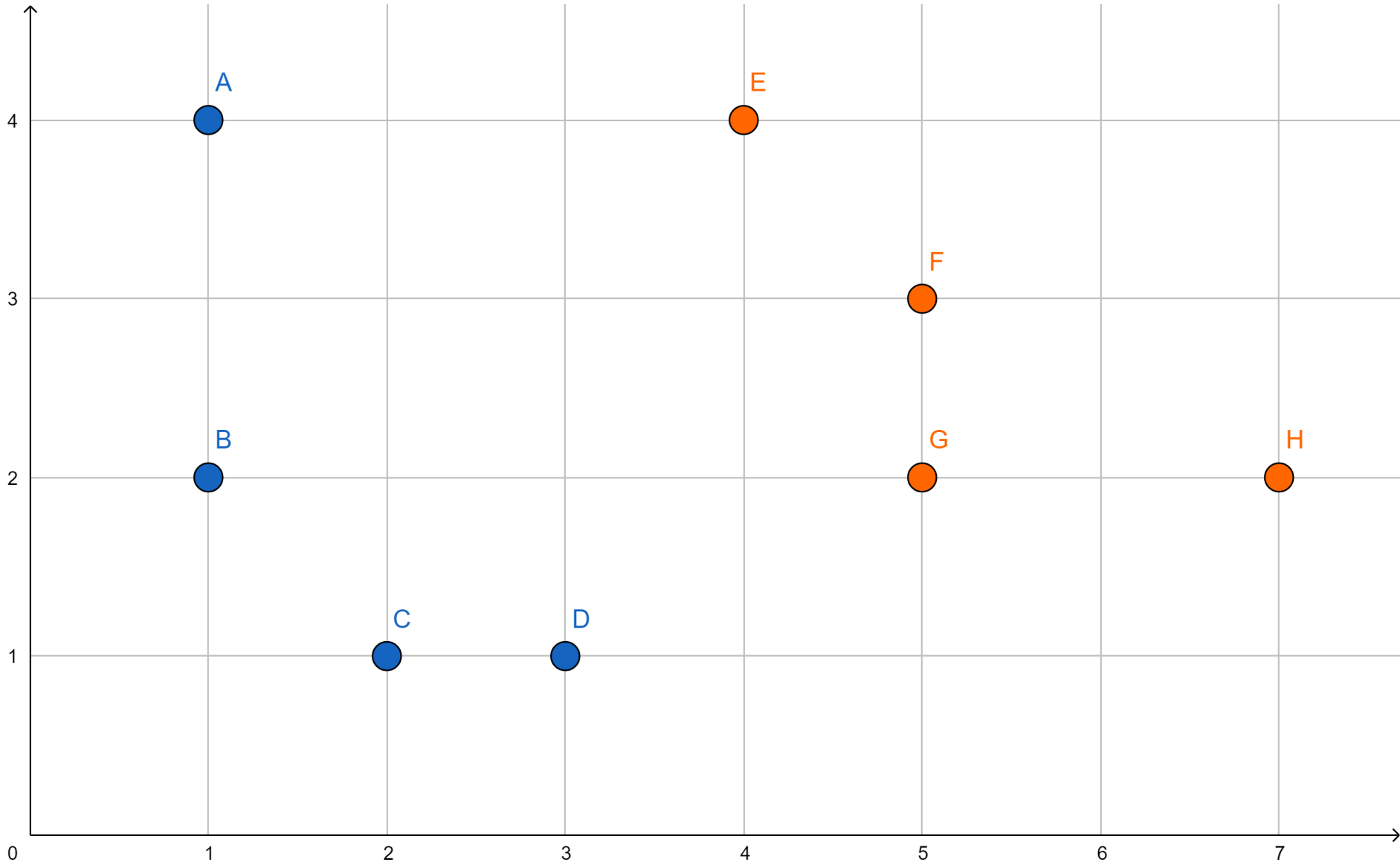
Values

●	Indefinite word distinct from 'one'	102
●	Indefinite word same as 'one'	112
●	Indefinite affix	24
◇	No indefinite, but definite article	98
○	No definite or indefinite article	198

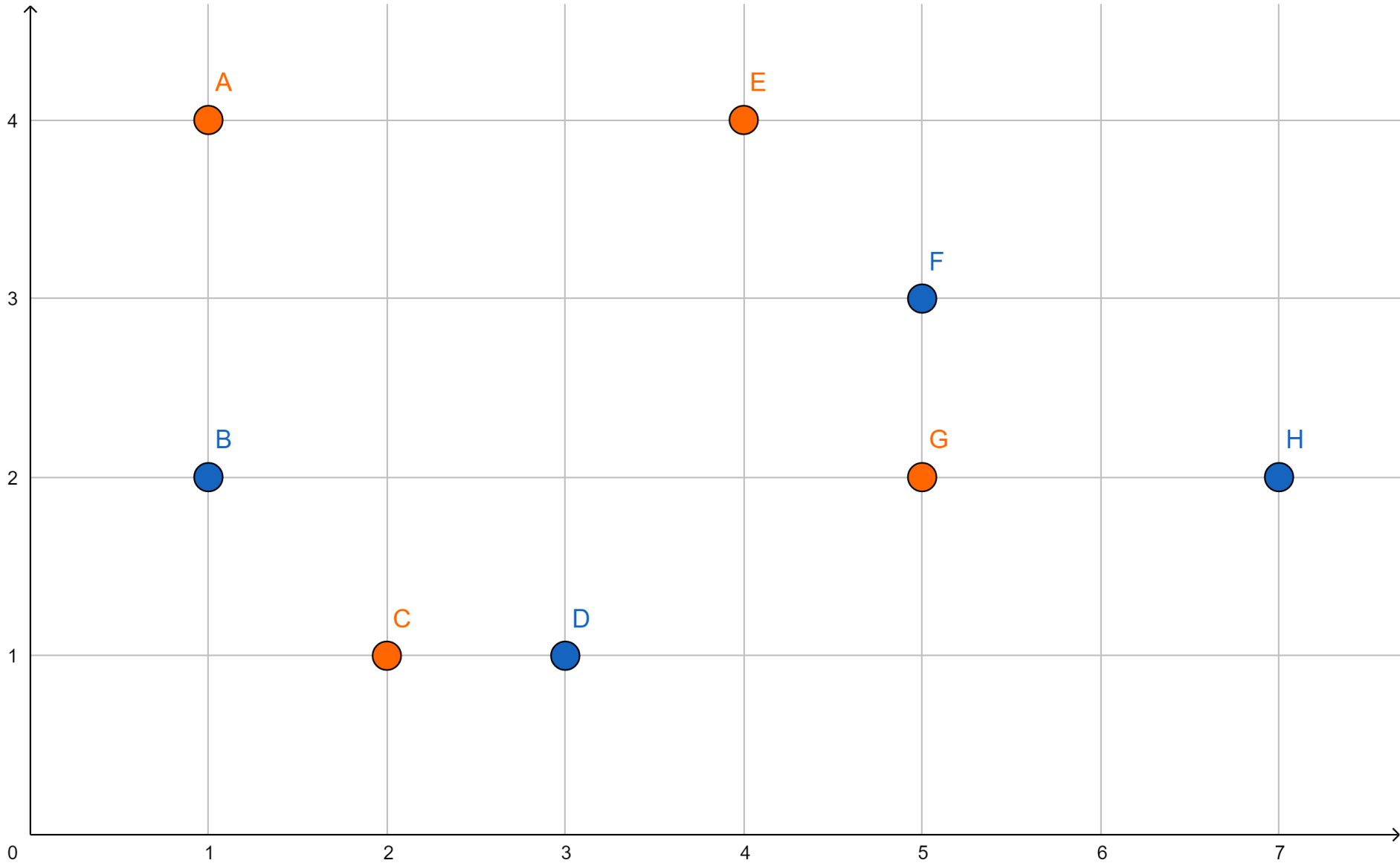


Ареальность на картах Всемирного атласа языковых структур

- На некоторых картах значки явно более сильнее группируются ареально
- Как это квантифицировать?



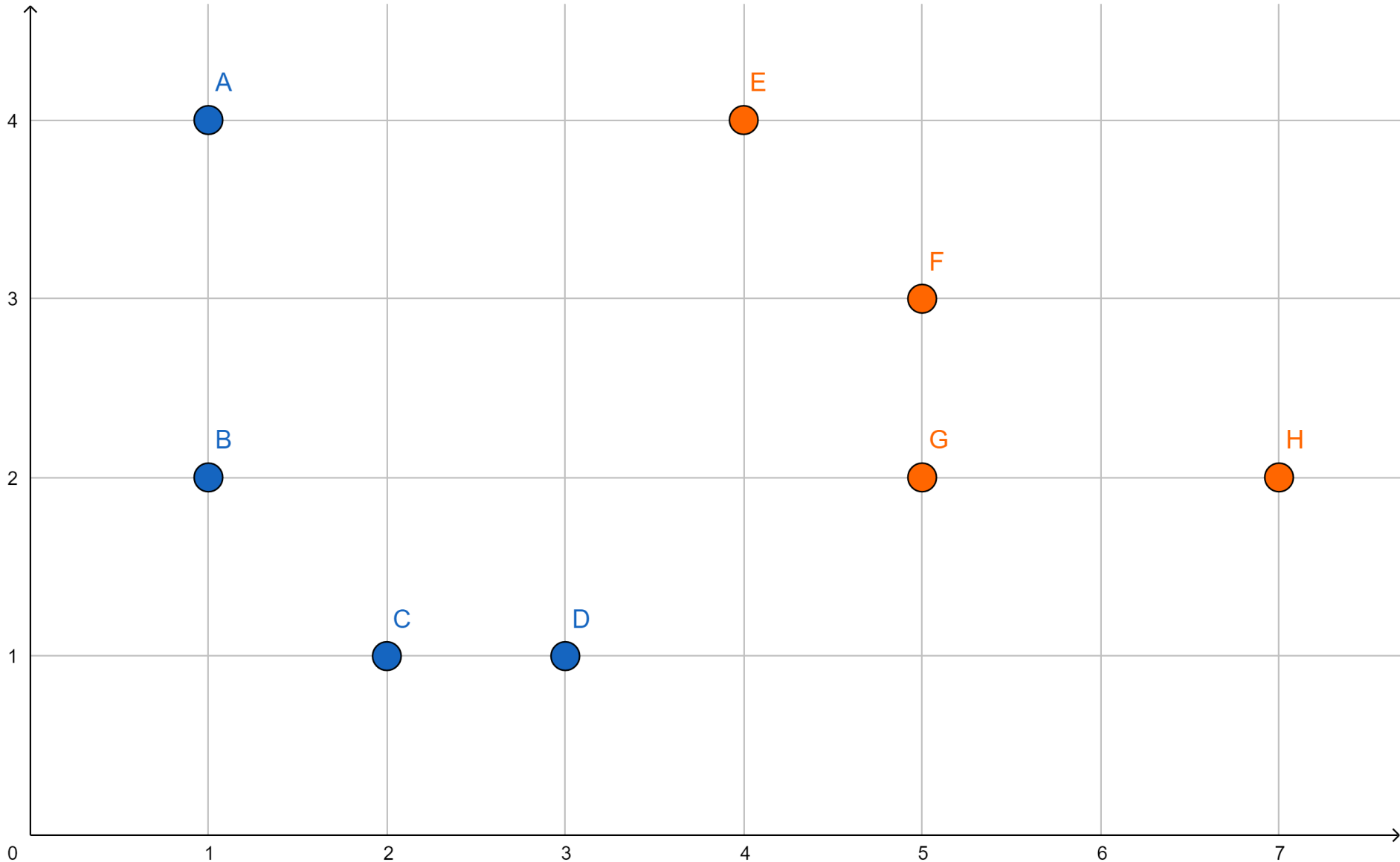
Карта 1 (ареальная)



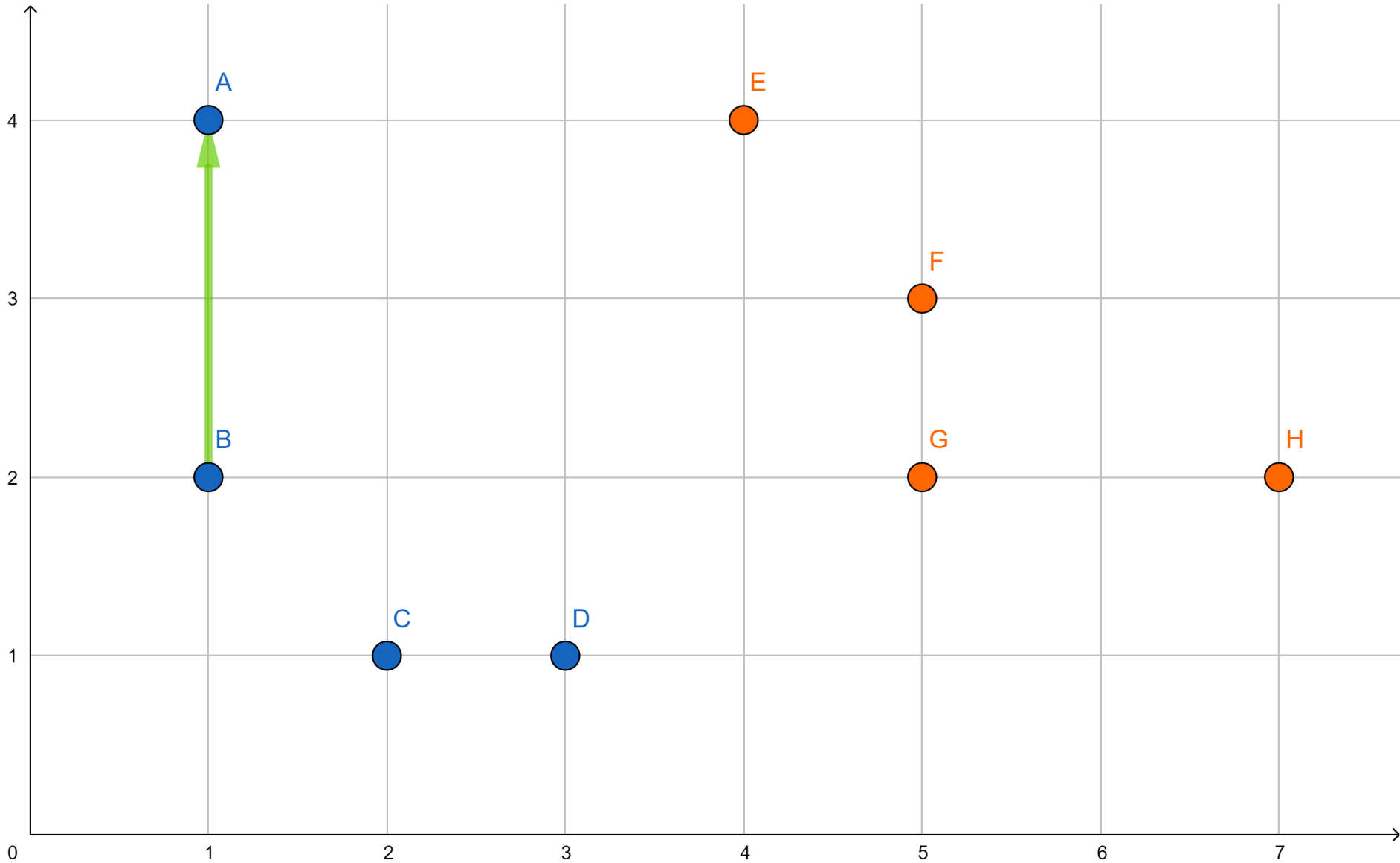
Карта 2 (неареальная)

Ареальность

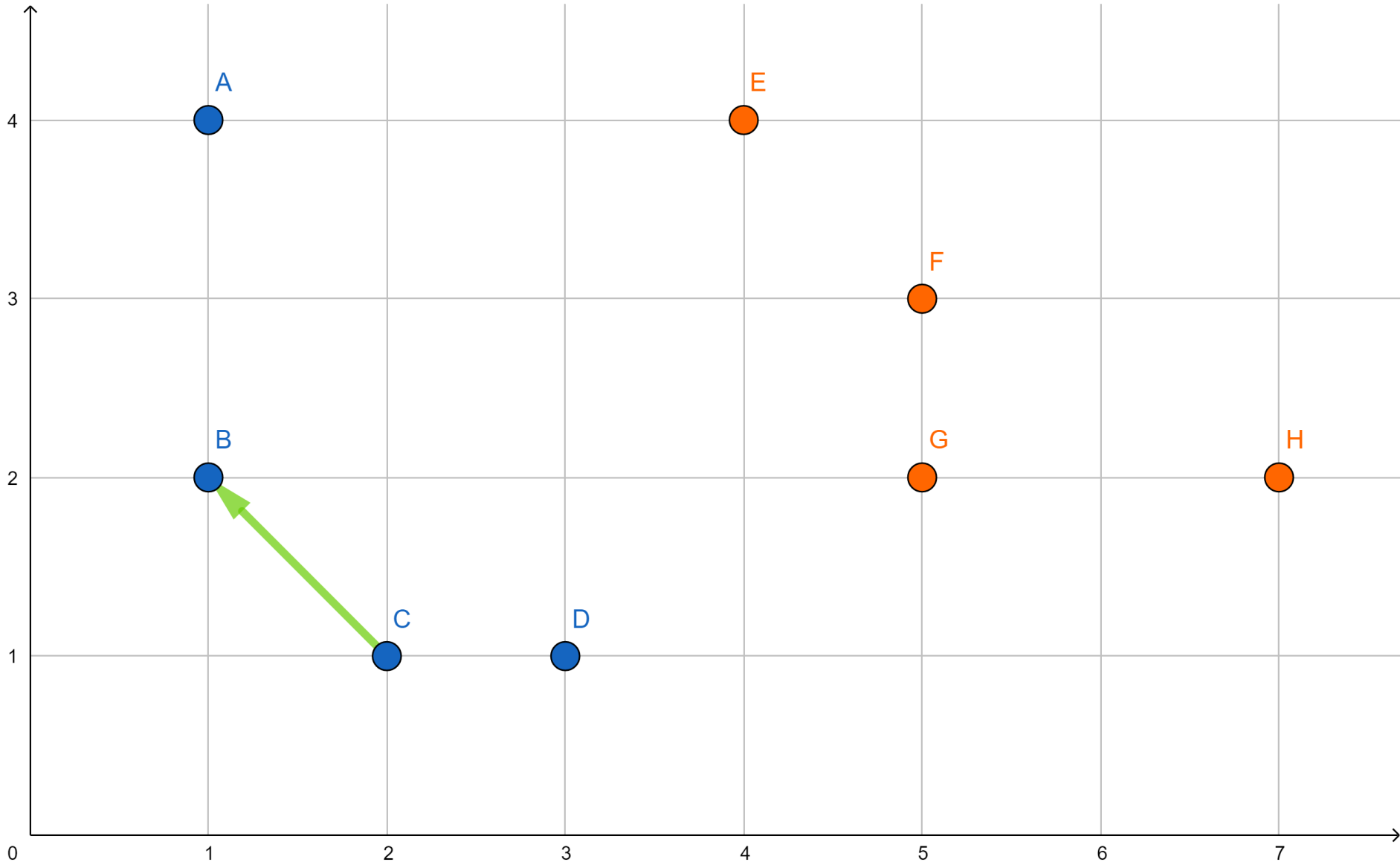
- Основная идея: на карте с явно выраженной ареальностью значки похожи на своих соседей



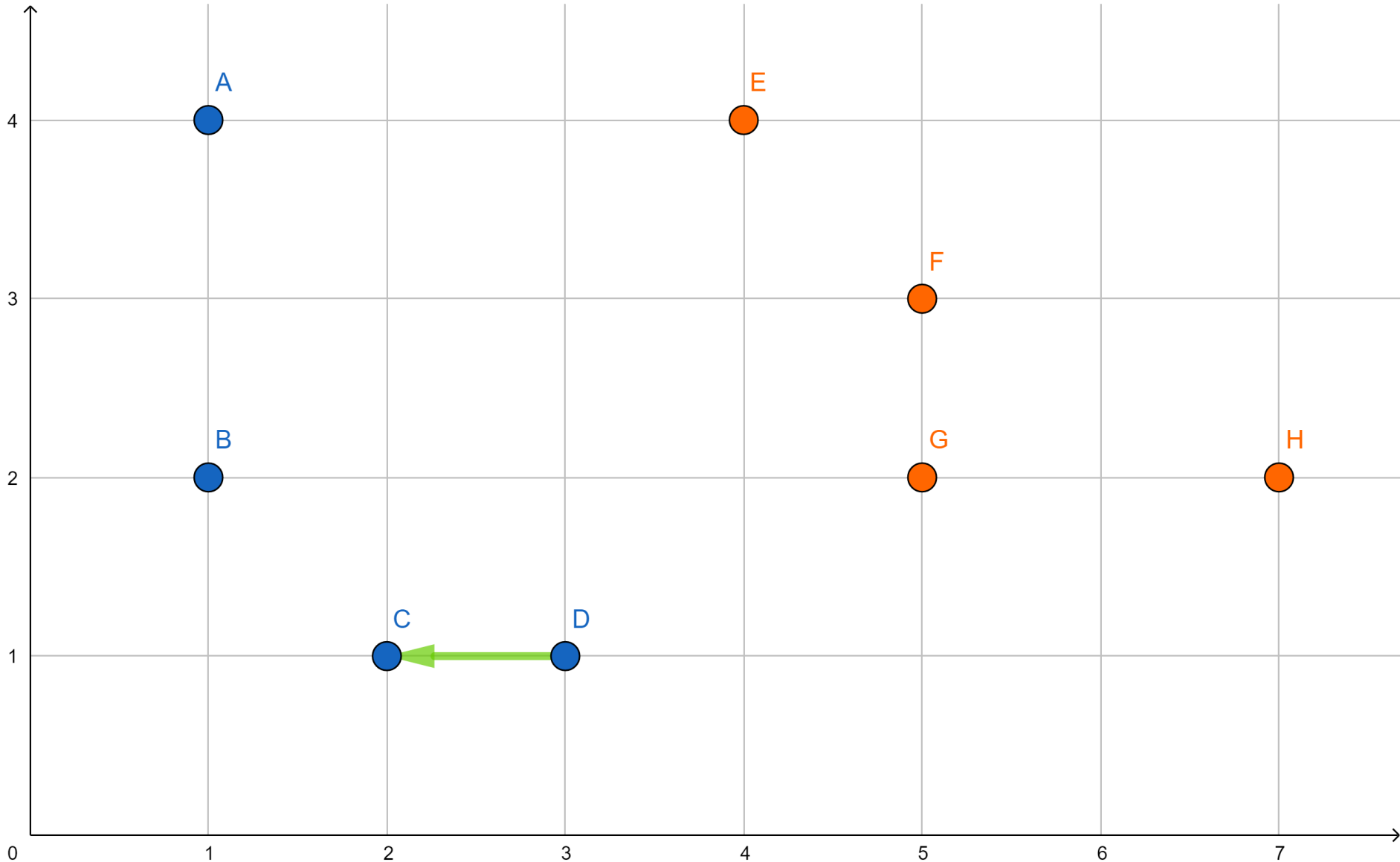
Карта 1 (ареальная)



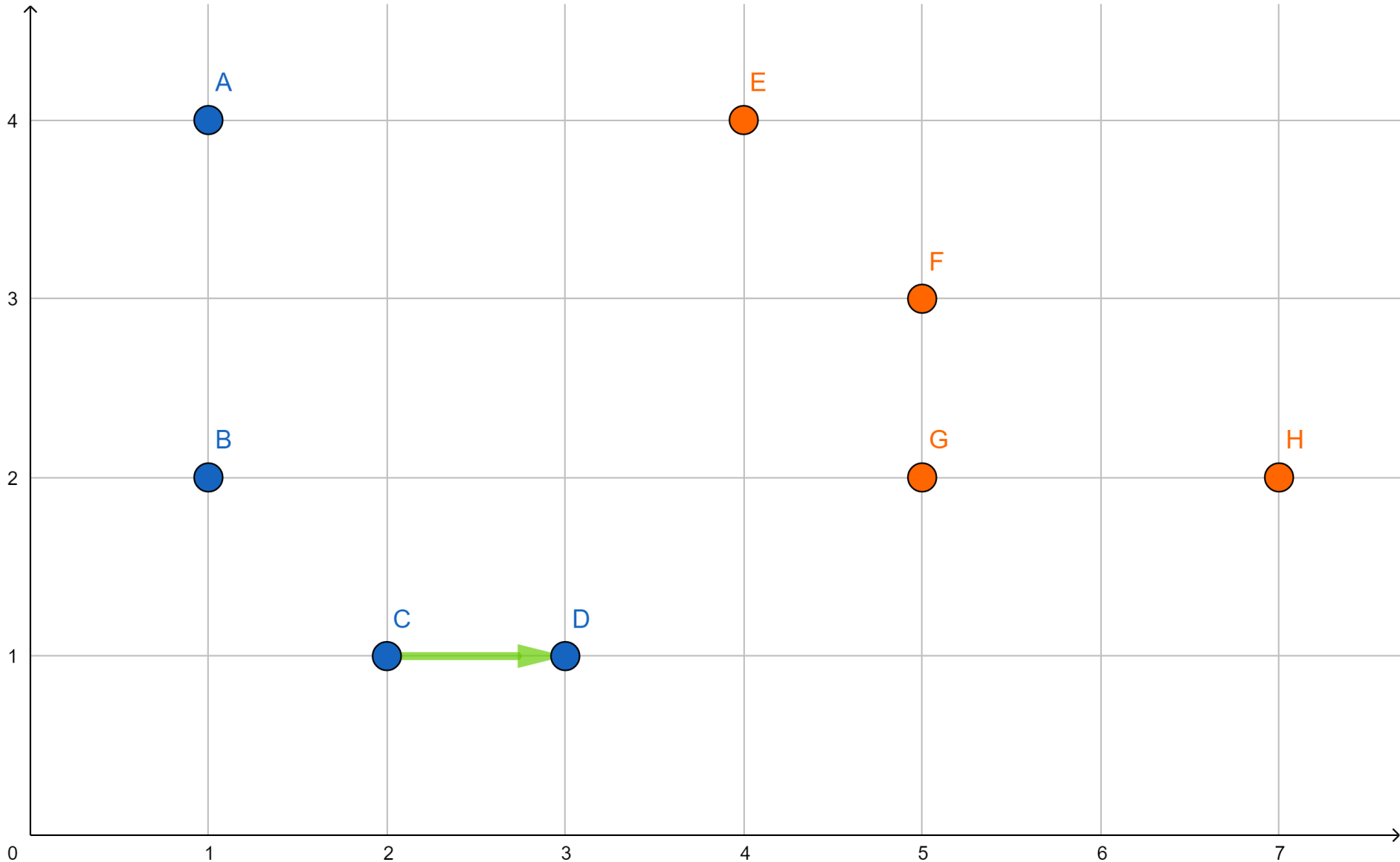
Карта 1 (ареальная)



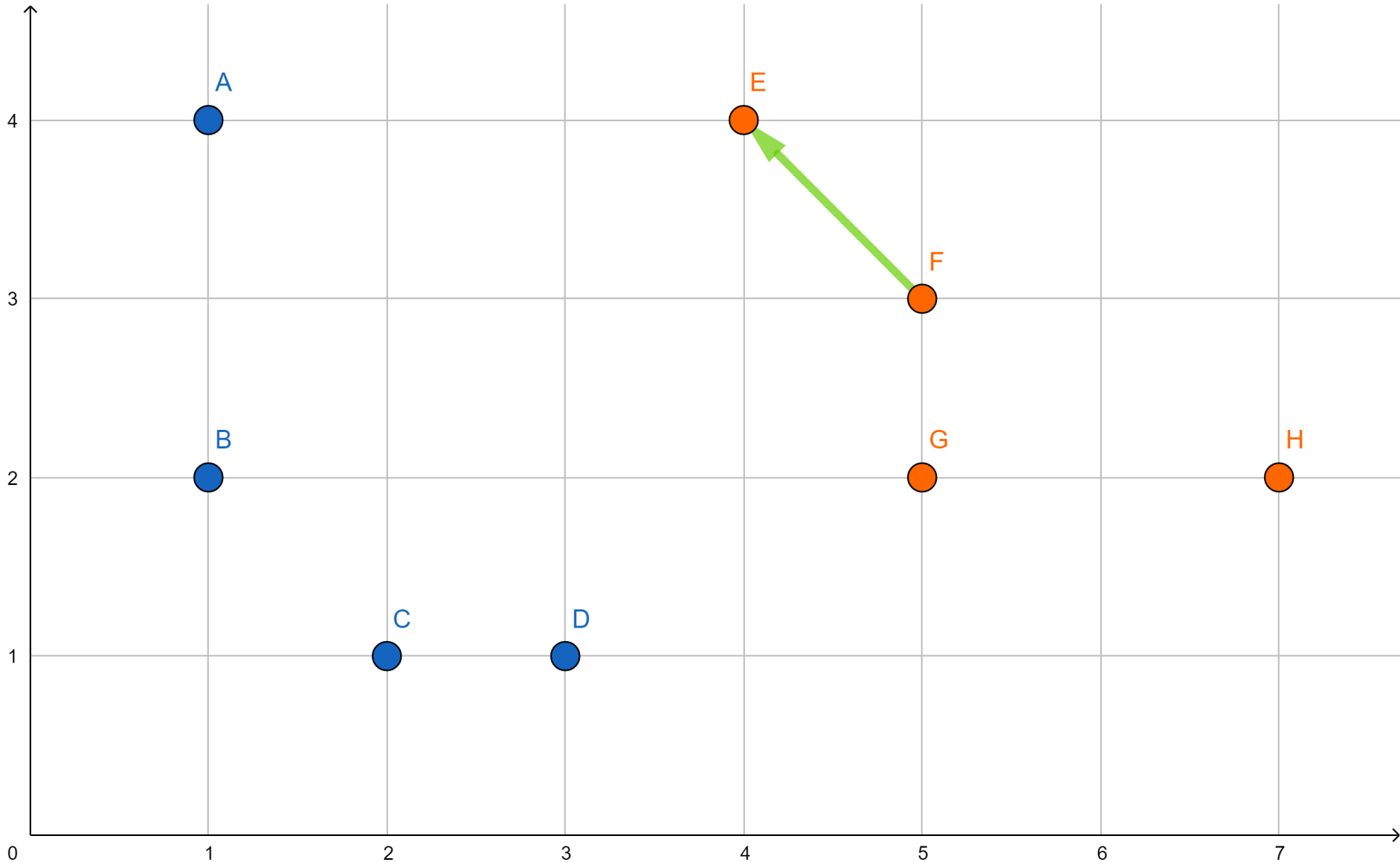
Карта 1 (ареальная)



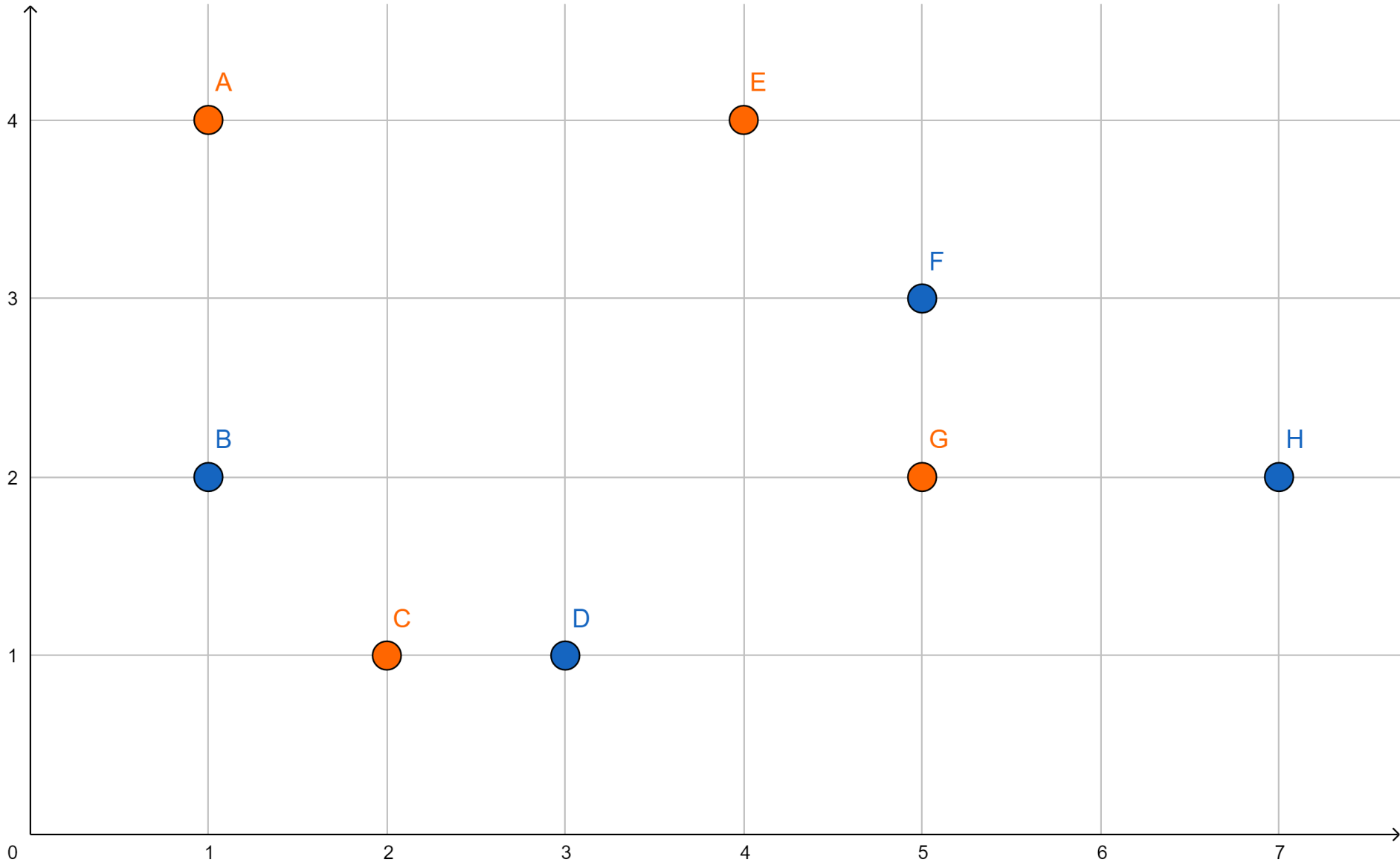
Карта 1 (ареальная)

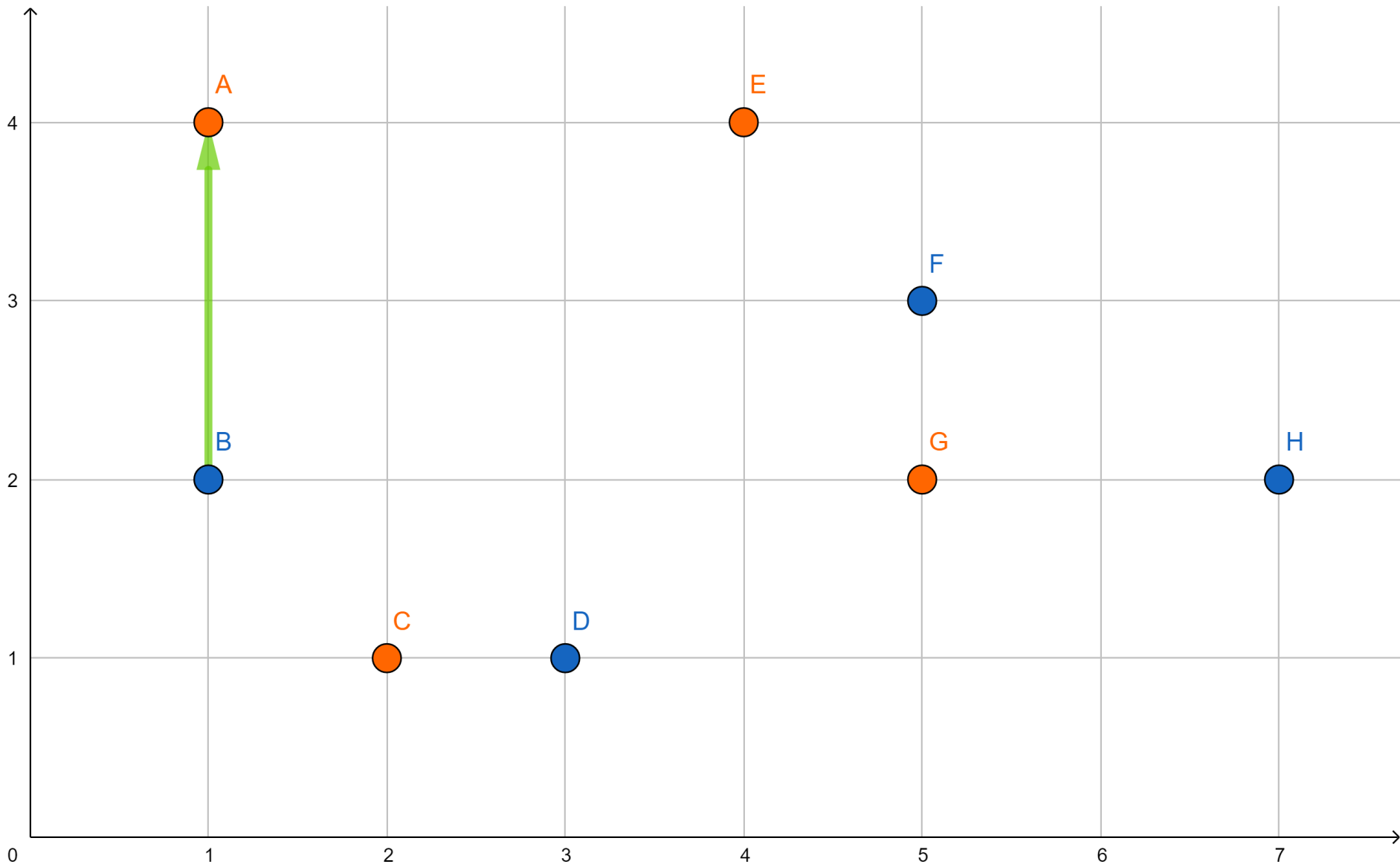


Карта 1 (ареальная)

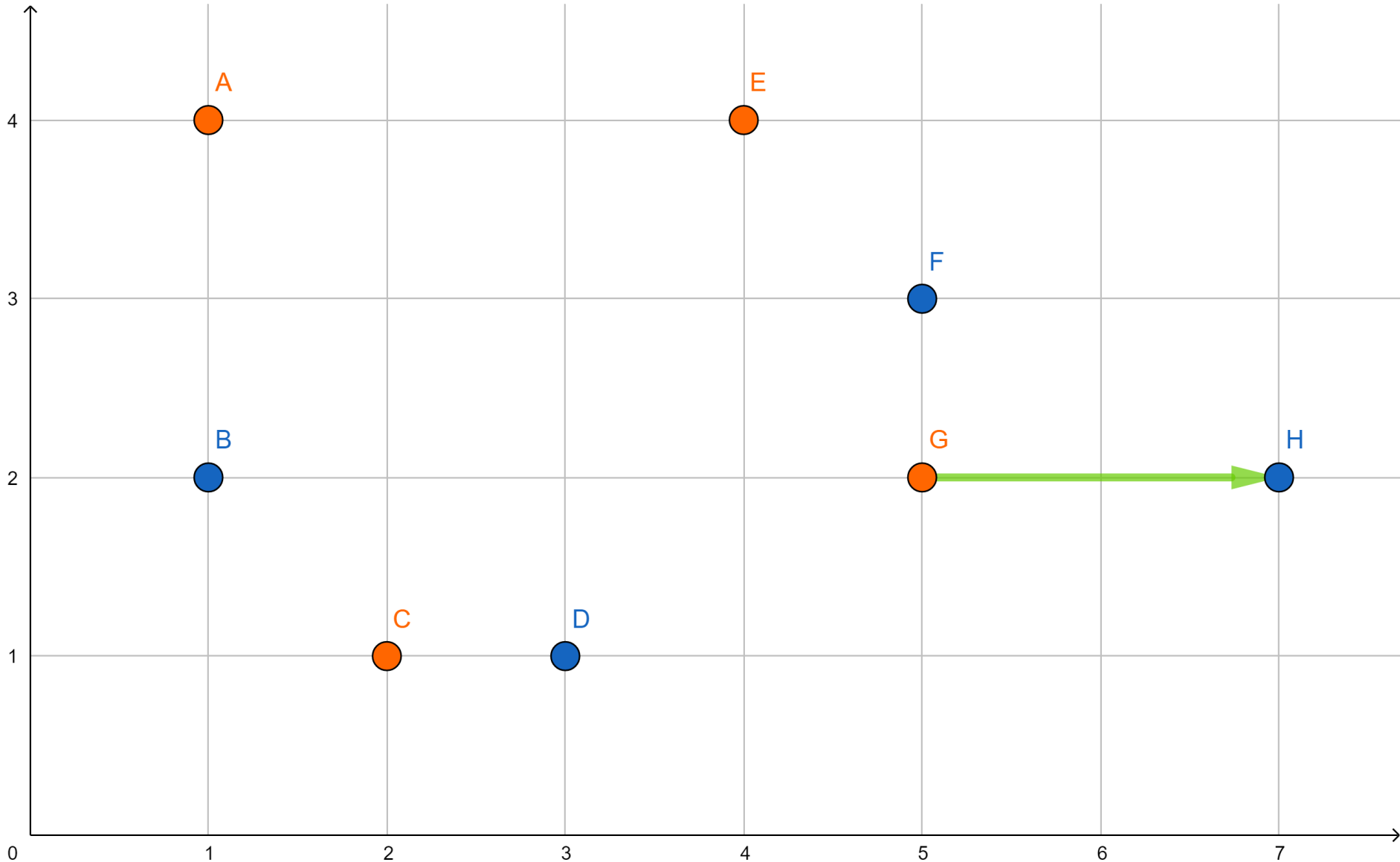


Карта 1 (ареальная)





Карта 2 (неореальная)



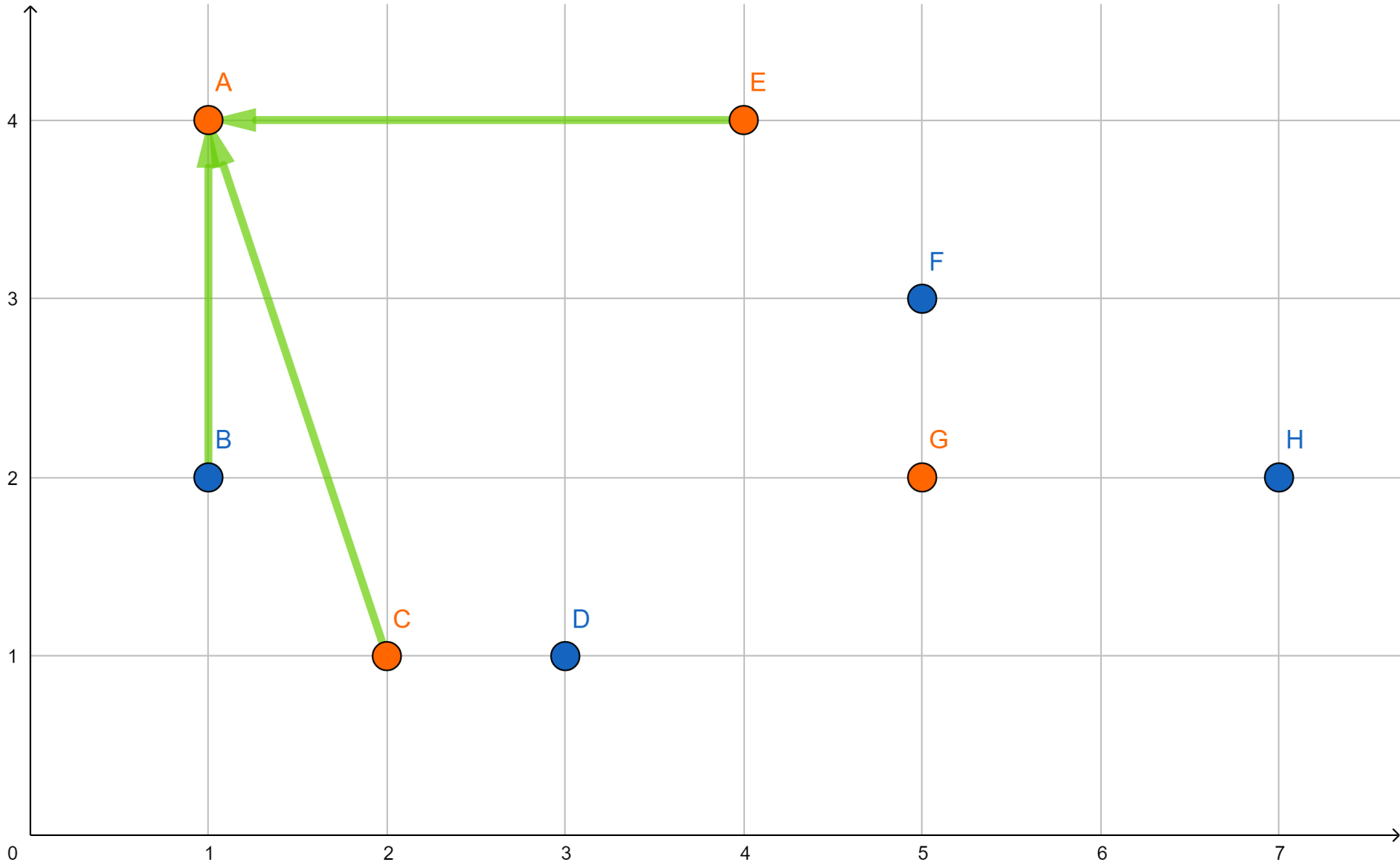
Карта 2 (нереальная)

Ареальность: ближайший сосед

- Карта 1 (ареальная):
8 точек из 8 совпадают по цвету
с ближайшим соседом
- Карта 2 (неареальная):
0 точек из 8 совпадают по цвету
с ближайшим соседом

Ареальность

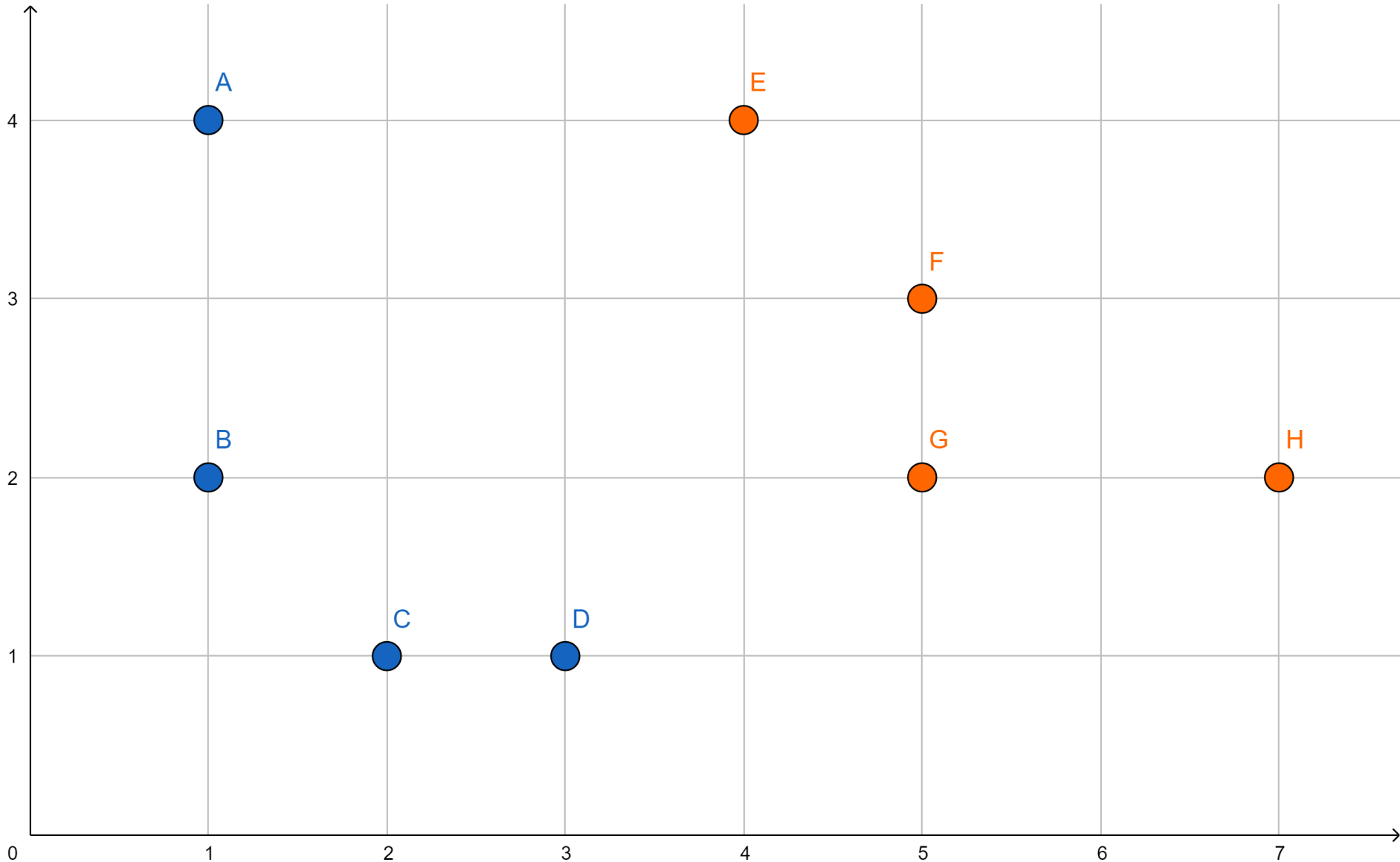
- Можно рассматривать больше соседей



Карта 2 (неореальная)

Ареальность: 3 ближайших соседа

- Карта 1 (ареальная):
8 точек из 8 совпадают по цвету
с большинством из 3 ближайших соседей

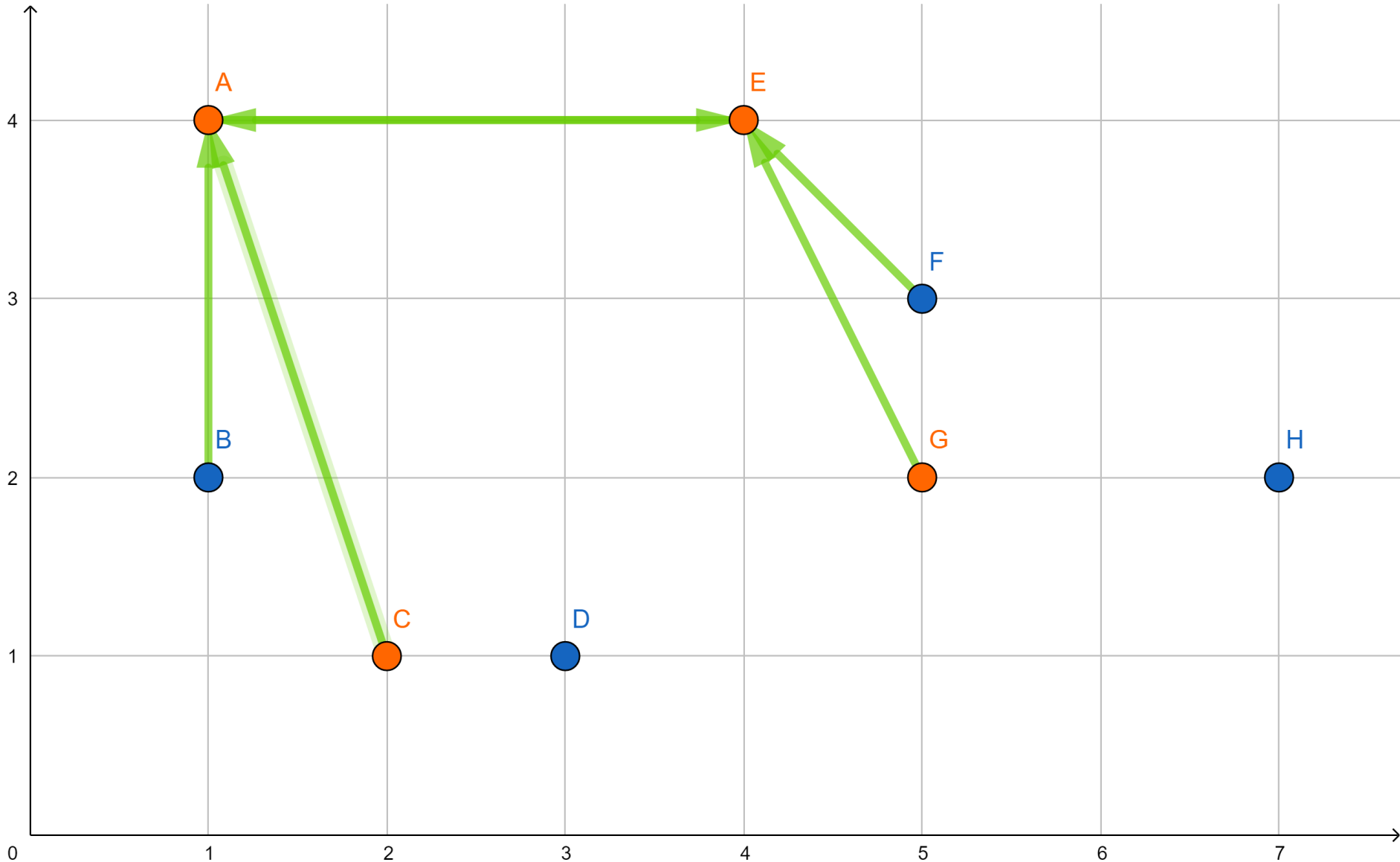


Карта 1 (ареальная)

Ареальность:

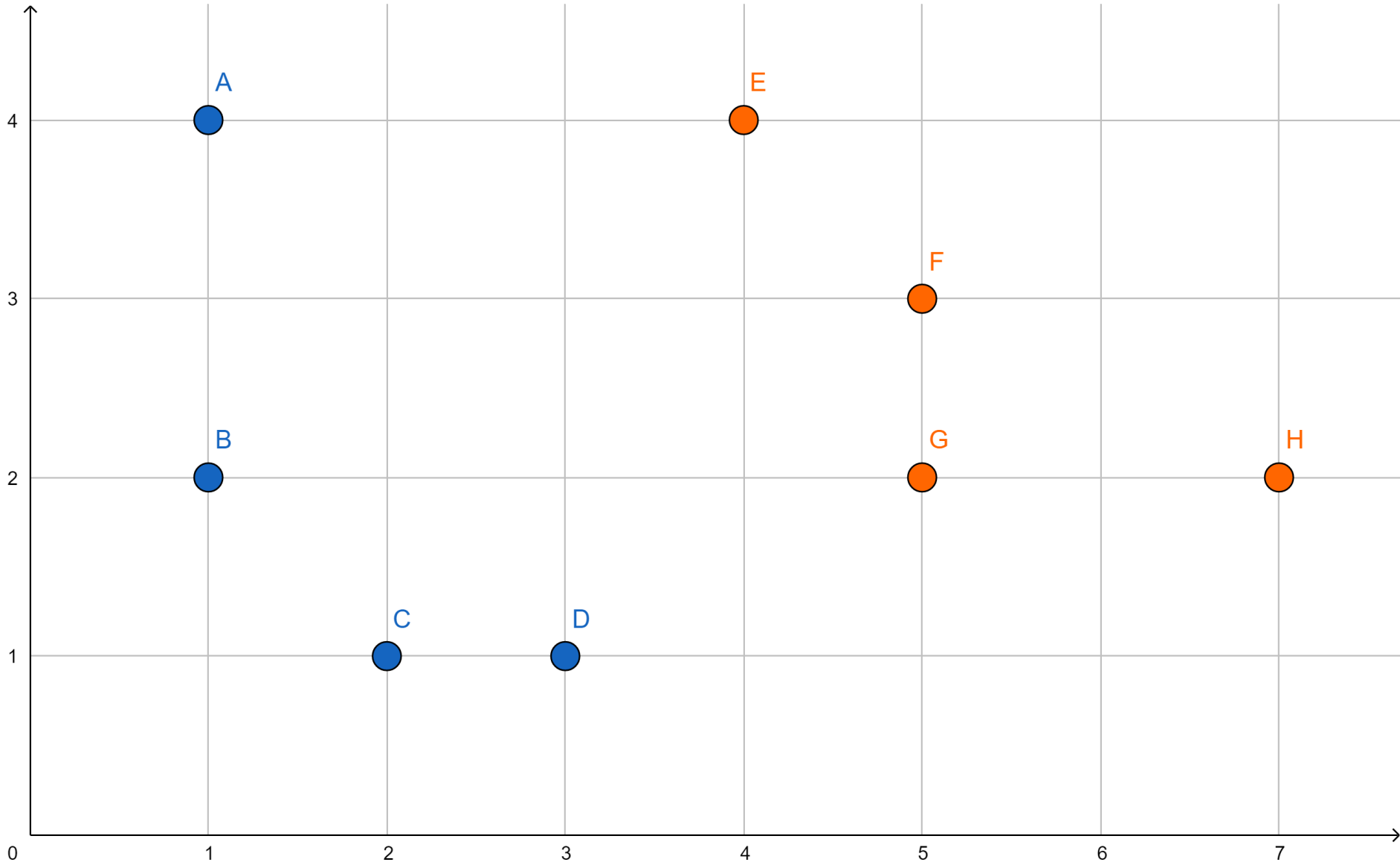
3 ближайших соседа

- Карта 1 (ареальная):
8 точек из 8 совпадают по цвету с большинством из 3 ближайших соседей
- Карта 2 (неареальная):
2 точек из 8 совпадают по цвету с большинством из 3 ближайших соседей



Карта 2 (неореальная)

Ареальность:
7 ближайших соседей



Карта 1 (ареальная)

Ареальность:
7 ближайших соседей



Ареальность: k ближайших соседей

k	Карта 1 (ареальная)	Карта 2 (неареальная)
1	8 (100%)	0 (0%)
3	8 (100%)	2 (25%)
5	8 (100%)	0 (0%)
7	0 (0%)	0 (0%)

Ареальность на картах WALS

- Проблемы:
 - разная выборка языков по количеству и составу на разных картах

Showing 1 to 100 of 192 entries



Id	Name	Authors	Area	Languages	Details
<input type="text" value="Search"/>	<input type="text" value="Search"/>		--any--	<input type="text" value="Search"/>	
83A	Order of Object and Verb	Matthew S. Dryer	Word Order	1519	Values
82A	Order of Subject and Verb	Matthew S. Dryer	Word Order	1497	Values
81A	Order of Subject, Object and Verb	Matthew S. Dryer	Word Order	1377	Values
87A	Order of Adjective and Noun	Matthew S. Dryer	Word Order	1366	Values
143F	Postverbal Negative Morphemes	Matthew S. Dryer	Word Order	1324	Values
143E	Preverbal Negative Morphemes	Matthew S. Dryer	Word Order	1324	Values
143A	Order of Negative Morpheme and Verb	Matthew S. Dryer	Word Order	1324	Values
143G	Minor morphological means of signaling negation	Matthew S. Dryer	Word Order	1324	Values
97A	Relationship between the Order of Object and Verb and the Order of Adjective and Noun	Matthew S. Dryer	Word Order	1316	Values
86A	Order of Genitive and Noun	Matthew S. Dryer	Word Order	1249	Values
88A	Order of Demonstrative and Noun	Matthew S. Dryer	Word Order	1224	Values
144A	Position of Negative Word With Respect to Subject, Object, and Verb	Matthew S. Dryer	Word Order	1189	Values
85A	Order of Adposition and Noun Phrase	Matthew S. Dryer	Word Order	1183	Values
112A	Negative Morphemes	Matthew S. Dryer	Simple Clauses	1157	Values

Showing 1 to 100 of 192 entries



Id	Name	Authors	Area	Languages	Details
<input type="text" value="Search"/>	<input type="text" value="Search"/>		--any--	<input type="text" value="Search"/>	
90G	Double-headed relative clauses	Matthew S. Dryer	Word Order	5	Values
141A	Writing Systems	Bernard Comrie	Other	6	Values
143D	Optional Triple Negation	Matthew S. Dryer	Word Order	6	Values
90F	Adjoined relative clauses	Matthew S. Dryer	Word Order	10	Values
144Y	The Position of Negative Morphemes in Object-Initial Languages	Matthew S. Dryer	Word Order	16	Values
144U	Double negation in verb-initial languages	Matthew S. Dryer	Word Order	17	Values
90E	Correlative relative clauses	Matthew S. Dryer	Word Order	23	Values
144C	Languages with different word order in negative clauses	Matthew S. Dryer	Word Order	28	Values
144O	Optional Double Negation in SOV languages	Matthew S. Dryer	Word Order	31	Values
139A	Irregular Negatives in Sign Languages	Ulrike Zeshan	Sign Languages	35	Values
144G	Optional Double Negation in SVO languages	Matthew S. Dryer	Word Order	35	Values
140A	Question Particles in Sign Languages	Ulrike Zeshan	Sign Languages	38	Values
10B	Nasal Vowels in West Africa	John Hajek	Phonology	40	Values
144N	Obligatory Double Negation in SOV languages	Matthew S. Dryer	Word Order	45	Values

Ареальность на картах WALS

- Проблемы:
 - разная выборка языков по количеству и составу на разных картах
 - разное количество значений признаков



Feature 107A: Passive Constructions



This feature is described in the text of chapter 107 [Passive Constructions](#) by Anna Siewierska [cite](#)

You may combine this feature with another one. Start typing the feature name or number in the field below.

Values

	Present	162
	Absent	211



Feature 33A: Coding of Nominal Plurality



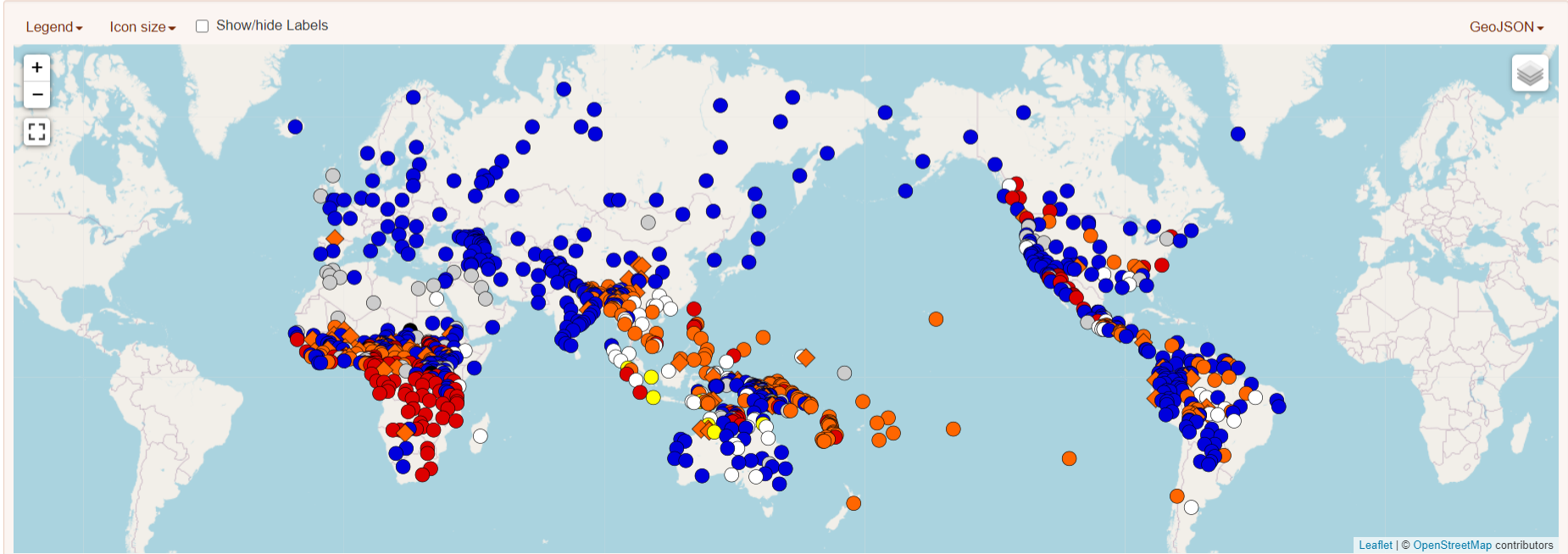
This feature is described in the text of chapter 33 [Coding of Nominal Plurality](#) by [Matthew S. Dryer](#) [cite](#)

You may combine this feature with another one. Start typing the feature name or number in the field below.

Submit

Values

●	Plural prefix	126
●	Plural suffix	513
●	Plural stem change	6
●	Plural tone	4
●	Plural complete reduplication	8
●	Mixed morphological plural	60
●	Plural word	170
◆	Plural clitic	81
○	No plural	98



Ареальность на картах WALS

- Проблемы:
 - разная выборка языков по количеству и составу на разных картах
 - разное количество значений признаков
 - разное частотное распределение значений признаков

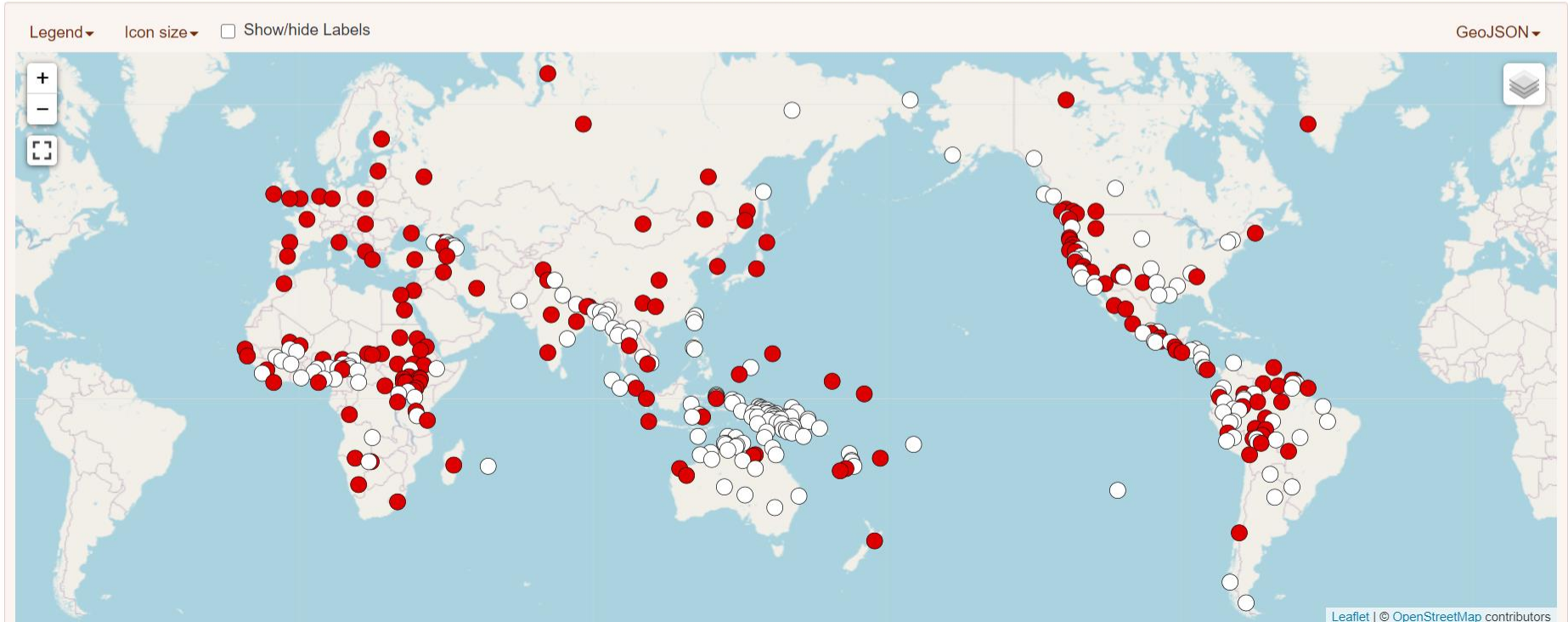
Feature 107A: Passive Constructions



This feature is described in the text of chapter 107 [Passive Constructions](#) by Anna Siewierska [cite](#)

You may combine this feature with another one. Start typing the feature name or number in the field below.

Values		
●	Present	162
	Absent	211



Feature 73A: The Optative

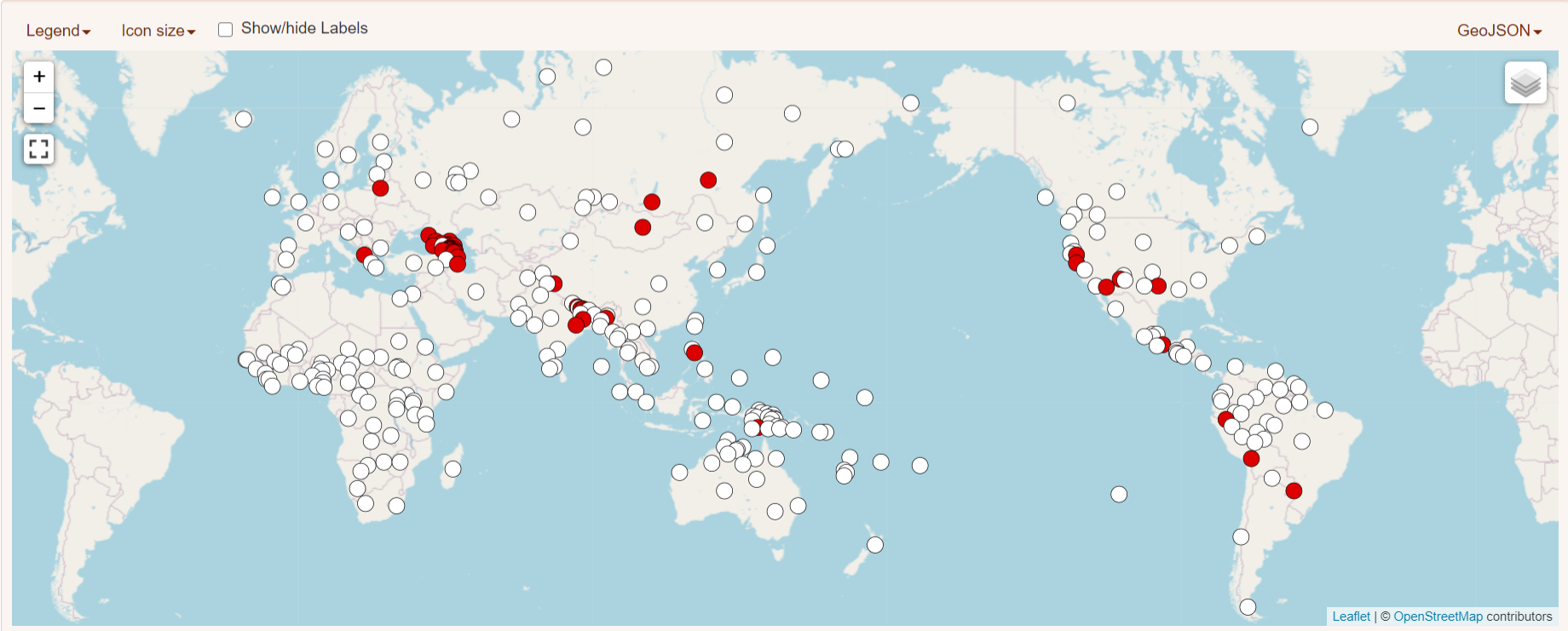


This feature is described in the text of chapter 73 [The Optative](#) by Nina Dobrushina and Johan van der Auwera and Valentin Goussev [cite](#)

You may combine this feature with another one. Start typing the feature name or number in the field below.

Values

<input checked="" type="radio"/>	Inflectional optative present	48
<input type="radio"/>	Inflectional optative absent	271



Методика

- Отобрать признаки, для которых есть данные по ≥ 300 языков
- Для каждого признака сгенерировать случайную выборку из 300 языков
- Для каждой выборки вычислить долю верных предсказаний методом $k = 1, 2, 3, \dots, 30$ ближайших соседей
 - При ничьей — случайный выбор
- Выбрать k , которое в среднем даёт наилучший результат для данного признака

Feature 73A: The Optative

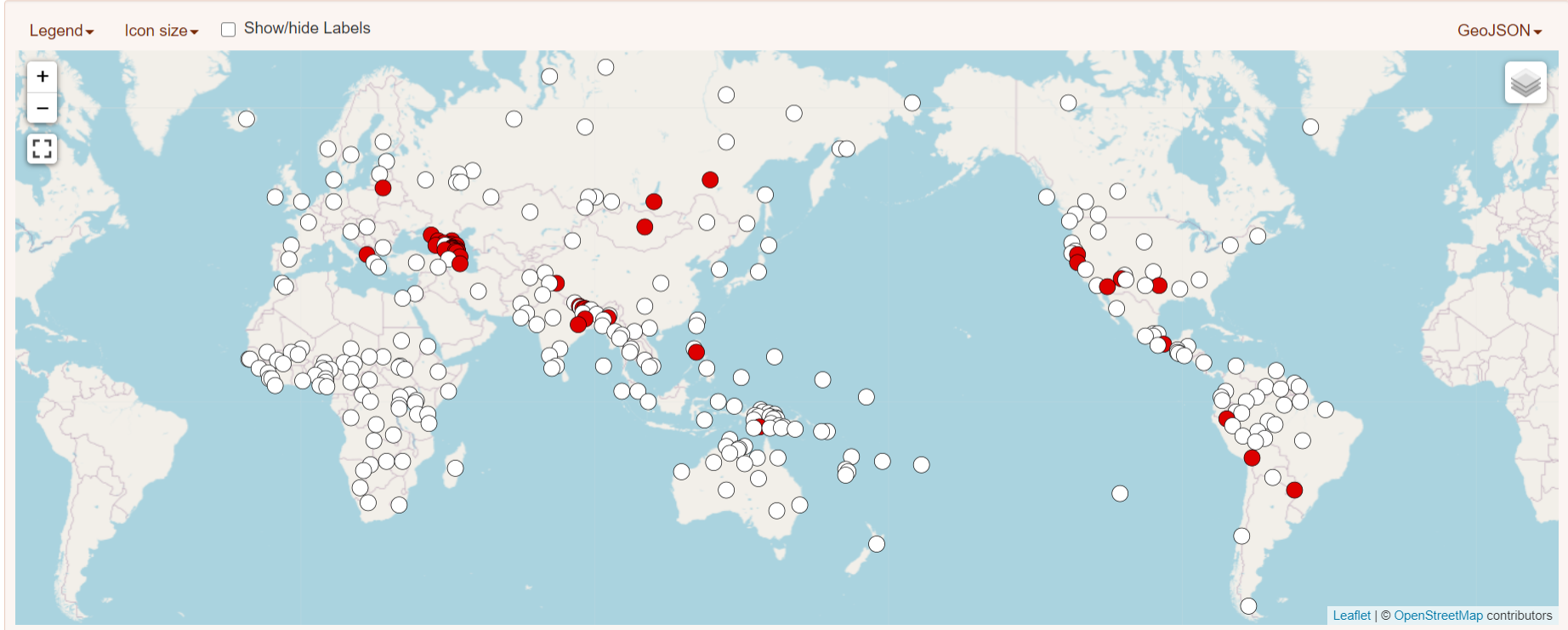


This feature is described in the text of chapter 73 [The Optative](#) by Nina Dobrushina and Johan van der Auwera and Valentin Goussev [cite](#)

You may combine this feature with another one. Start typing the feature name or number in the field below.

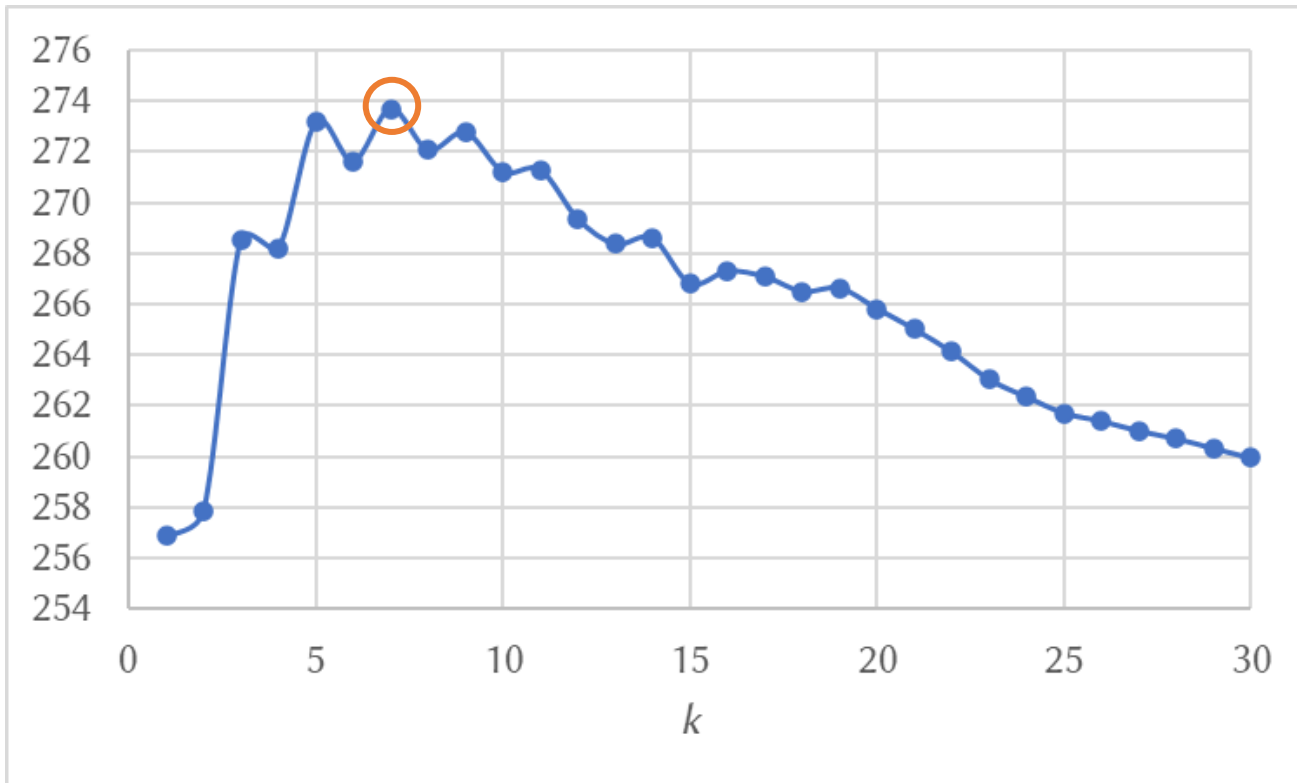
Values

<input checked="" type="radio"/>	Inflectional optative present	48
<input type="radio"/>	Inflectional optative absent	271



73A. The Optative

- Лучший результат при $k = 7$
- 273,7 (91,2%) верных предсказаний



Feature 84A: Order of Object, Oblique, and Verb



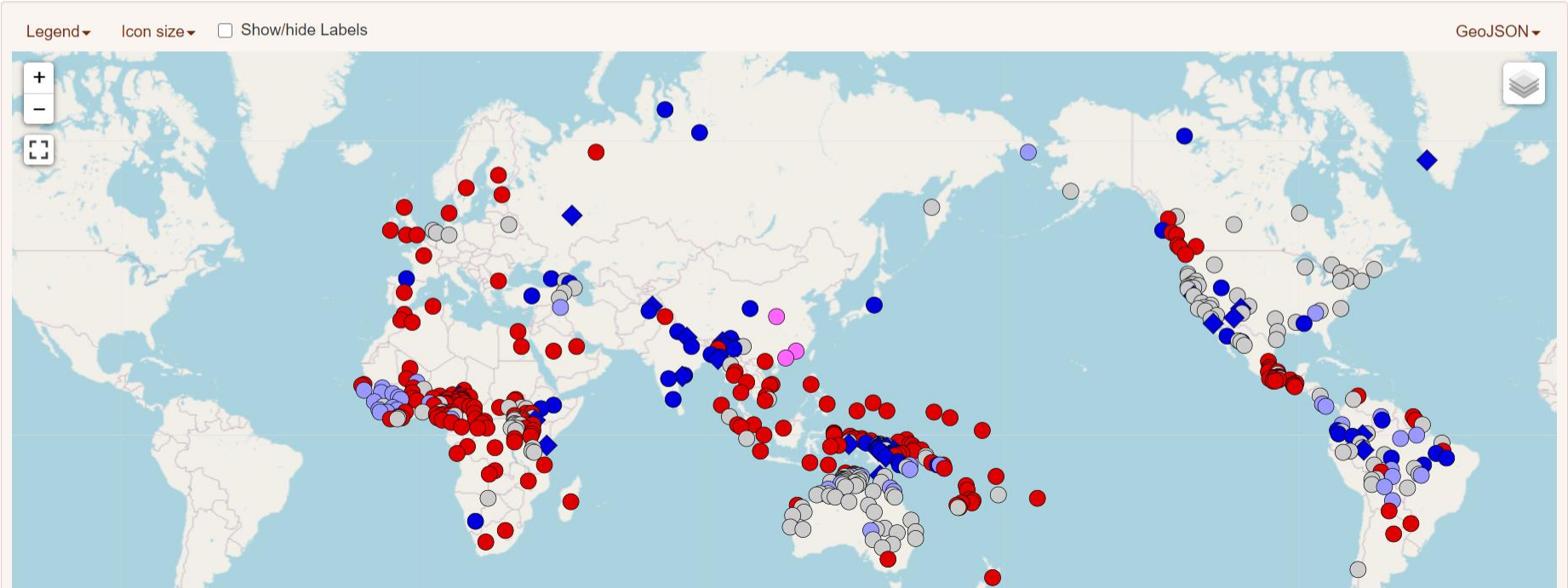
This feature is described in the text of chapter 84 [Order of Object, Oblique, and Verb](#) by Matthew S. Dryer with Orin D. Gensler [cite](#)

You may combine this feature with another one. Start typing the feature name or number in the field below.

Submit

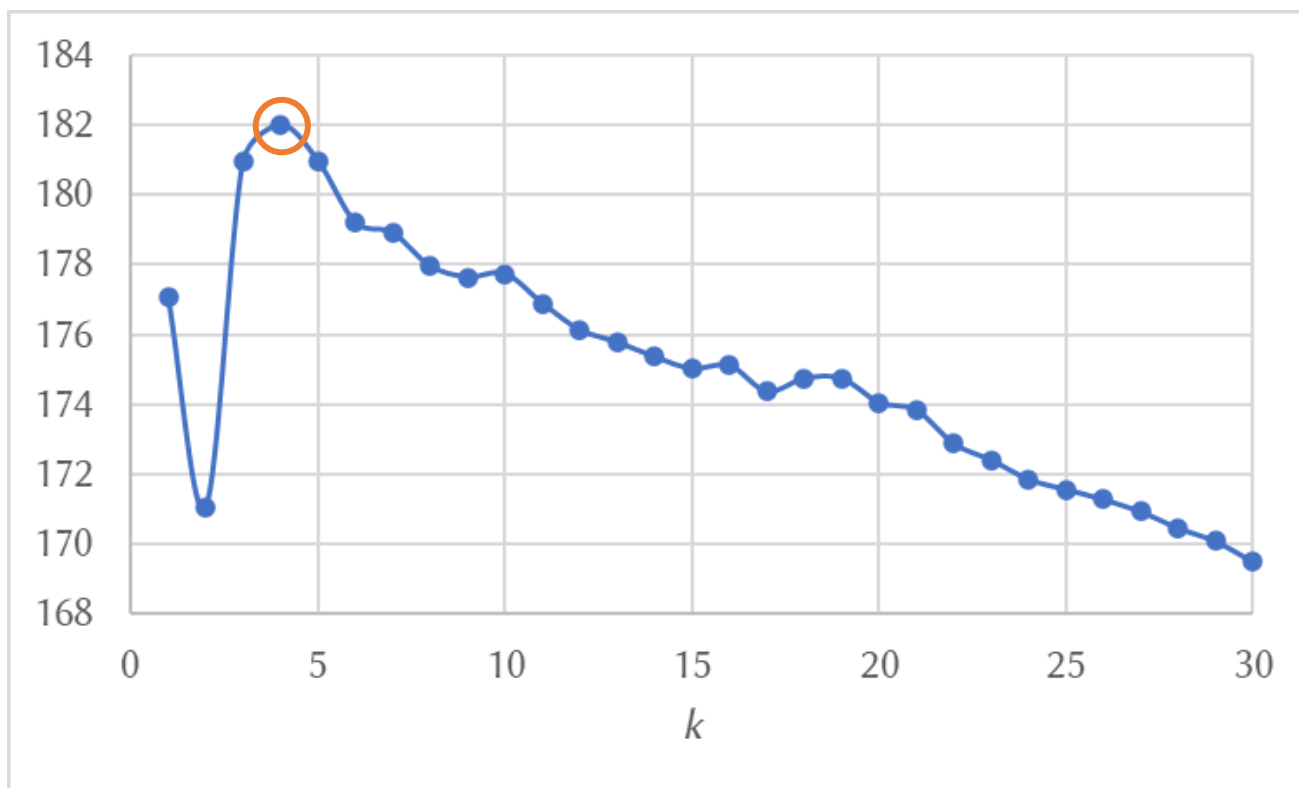
Values

●	VOX	210
●	XVO	3
●	XOV	48
◆	OXV	27
●	OVX	45
●	No dominant order	167



84A. Order of Object, Oblique and Verb

- Лучший результат при $k = 4$
- 182,0 (60,7%) верных предсказаний



Оценка успешности предсказаний

- Нужен **бейслайн**
- Самый частый класс
 - 73А. The Optative:
Inflectional Optative Absent (271 / 319 = 85,0%)
 - 84А. Order of Object, Oblique and Verb
VOX (210 / 500 = 42,0%)
- Успешность (s): на какую долю возможного улучшения метод ближайших соседей (n) лучше бейслайна (b):

$$s = (n - b) / (1 - b)$$

Оценка успешности предсказаний

- 73A. The Optative:

$$n = 91,2\%; b = 85,0\%$$

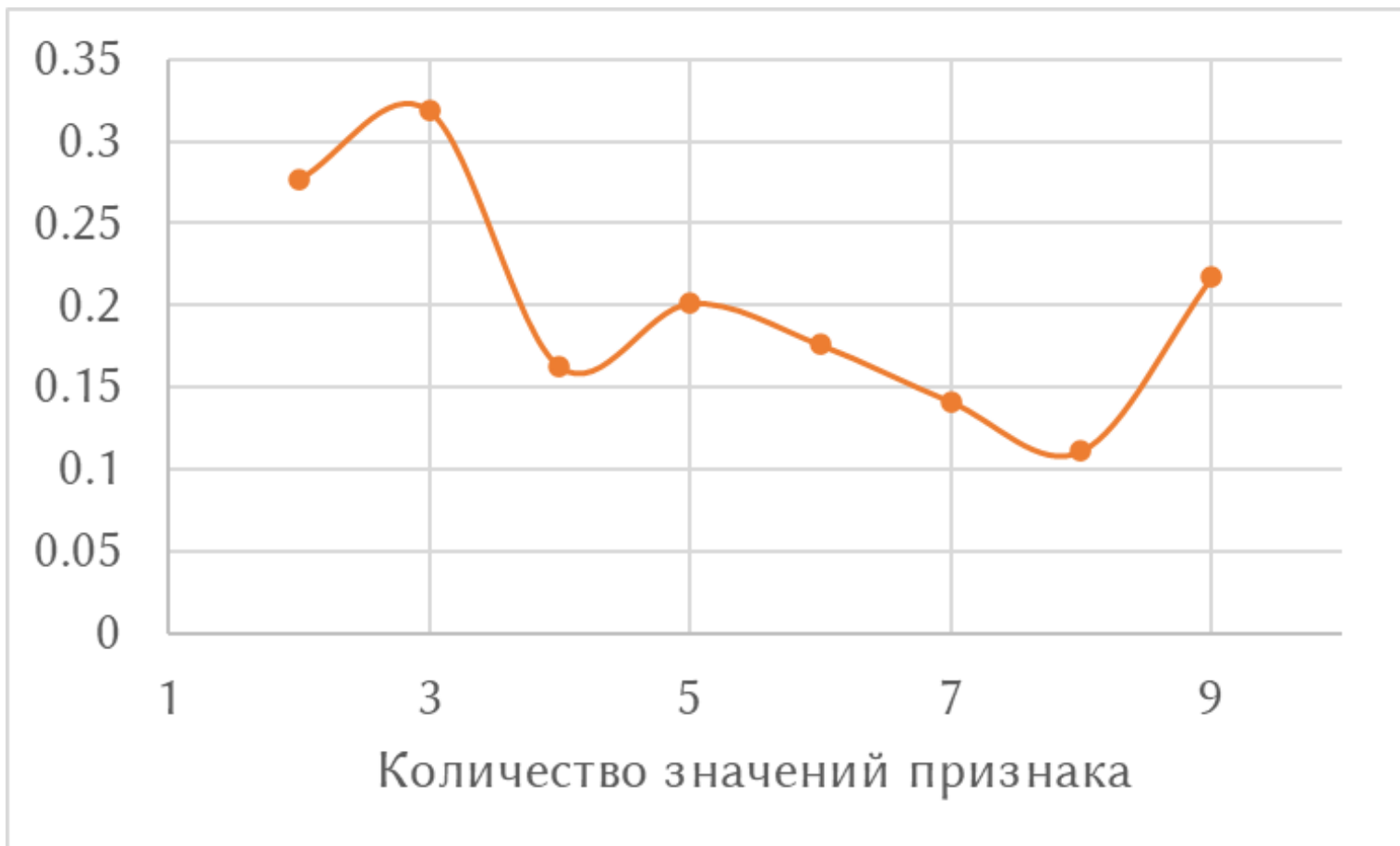
$$s = (91,2\% - 85,0\%) / (100\% - 85,0\%) = \\ = 6,2\% / 15,0\% = 0,41$$

- 84A. Order of Object, Oblique and Verb

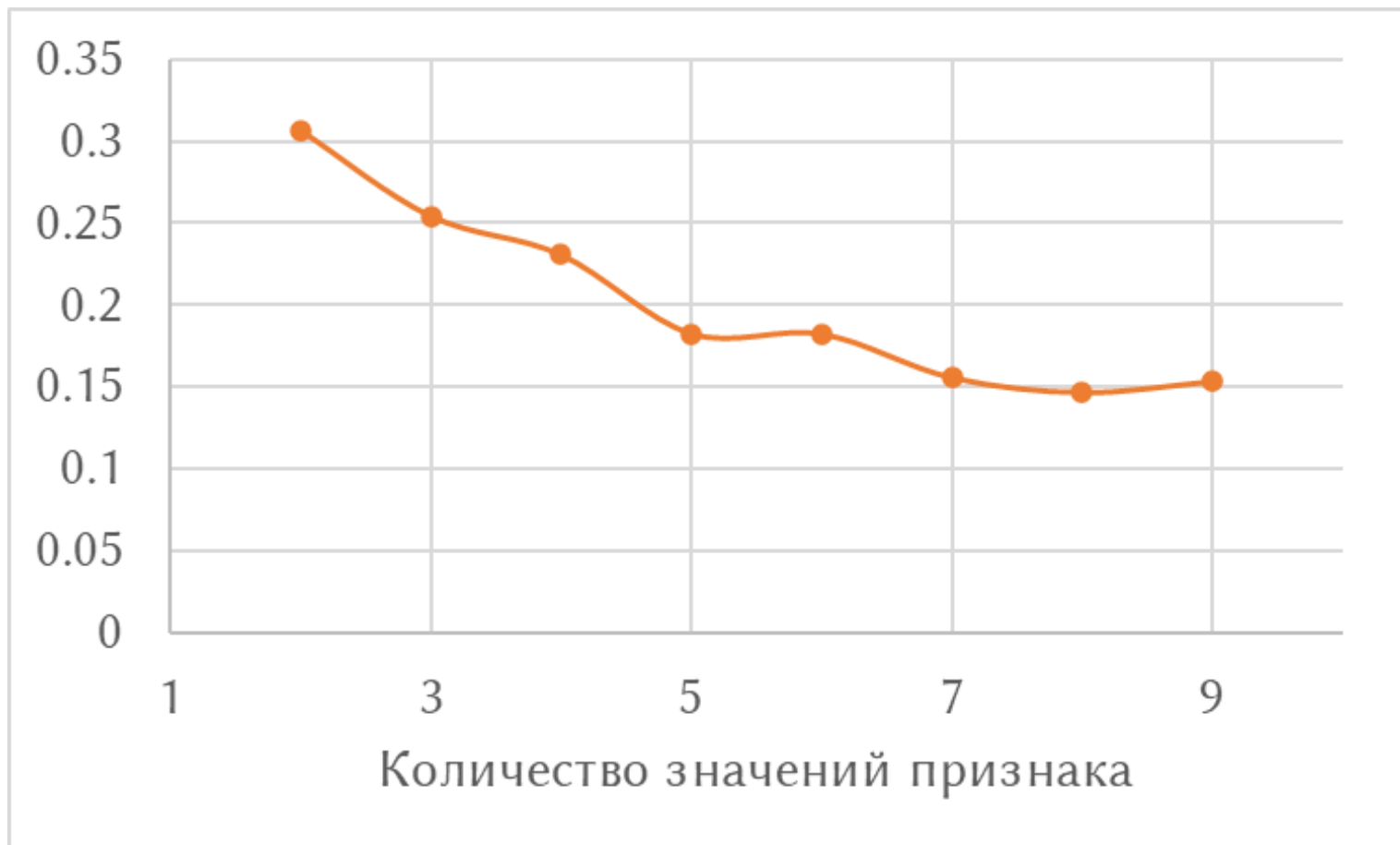
$$n = 60,7\%; b = 42,0\%$$

$$s = (60,7\% - 42,0\%) / (100\% - 42,0\%) = \\ = 18,7\% / 58,0\% = 0,32$$

Средняя успешность предсказаний



Средняя успешность предсказаний (сглаженная)



Оценка успешности предсказаний

- Итоговая оценка успешности:

$$S_{norm} = s / \bar{s}_v$$

(\bar{s}_f — средняя успешность по признакам с $v - 1$, v и $v + 1$ значениями)

Самые ареальные признаки

ID	Признак	Категория	Кол-во значений признака	Наиболее частотный класс	$s (s_{norm})$	k	Предсказание по k ближайшим соседям
85A	Order of Adp and NP	Word Order	5	146,07	55% (3,0)	5	231,32
95A	Relationship between the Order of O and V and the Order of Adp and NP	Word Order	5	123,99	51% (2,8)	5	212,93
88A	Order of Dem and N	Word Order	6	137,50	47% (2,6)	17	214,48
81A	Order of S, O and V	Word Order	7	123,09	39% (2,5)	5	192,74
46A	Indefinite Pronouns	Nominal Categories	5	178,53	43% (2,4)	5	231,21
89A	Order of Num and N	Word Order	4	157,94	52% (2,3)	6	232,30
83A	Order of O and V	Word Order	3	140,82	55% (2,2)	3	228,14
5A	Voicing and Gaps in Plosive Systems	Phonology	5	134,92	38% (2,1)	15	198,19
51A	Position of Case Affixes	Nominal Categories	9	131,52	32% (2,1)	7	184,80
86A	Order of Gen and N	Word Order	3	164,53	50% (1,9)	3	231,66

Feature 85A: Order of Adposition and Noun Phrase

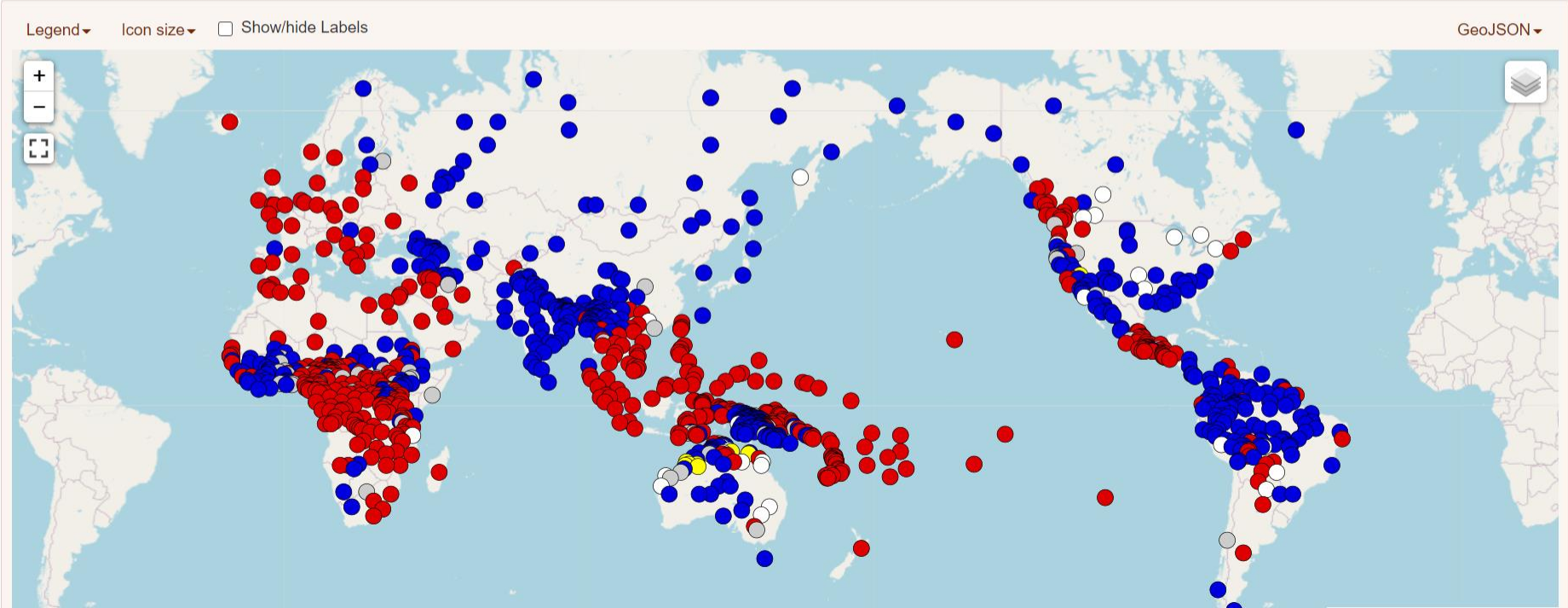


This feature is described in the text of chapter 85 by Matthew S. Dryer

You may combine this feature with another one. Start typing the feature name or number in the field below.

Values

●	Postpositions	577
●	Prepositions	511
●	Inpositions	8
●	No dominant order	58
○	No adpositions	30



Самые ареальные уровни языка

Уровень языка	Количество признаков	Среднее s_{norm}
Word Order	18	1,67
Nominal Categories	11	1,08
Nominal Syntax	1	1,02
Phonology	18	0,81
Verbal Categories	7	0,66
Simple Clauses	13	0,62
Morphology	2	0,59
Lexicon	2	0,51

Самые ареальные уровни языка

Уровень языка	Количество признаков	Среднее s_{norm}
Word Order	18	1,67
Nominal Categories	11	1,08
Phonology	18	0,81
Verbal Categories	7	0,66
Simple Clauses	13	0,62

- **Наибольшая ареальность наблюдается в порядке слов**

Интерпретация

- Ареальность на картах может быть связана с тем, что значения признака:
 - хорошо сохраняются
(а родственные языки часто расположены рядом)
 - легко заимствуются
- Хочется построить модель, которая будет измерять вклад двух этих факторов

Интерпретация



Н е l p !

Спасибо
за внимание!