

IdeaPeak

# Bericht Experimente



Daniel Schulz, Martin Schüßler, Sascha Elsner, Markus Wirth

## Inhalt

Informationen zum Ablauf .....	3
Forschungsfrage .....	4
Welche Heuristik bietet in einem computerbasierten Bewertungsworkshop die besten Voraussetzungen für die Anwendung eines Stopp-Kriteriums? .....	4
Eigene Forschungsfrage .....	5
Wie wirken sich objektive und subjektive Bewertungskriterien auf den Rankingprozess aus? .....	5
Auswertungskriterien .....	6
Jeweils für objektive und subjektive Kriterien: Ab wie vielen Paarvergleichen erreichen die einzelnen Heuristiken die korrekte Platzierung der Top 5? .....	6
Im Vergleich dazu: Wann hätte das konkrete Stopp-Kriterium gegriffen? .....	6
Weitere Fragestellungen .....	6
Bei welcher Heuristik griff das Stopp-Kriterium nach korrekt ermittelter Platzierung der Top 5 am frühesten (Zeitersparnis unter Gewährleistung der Korrektheit)? .....	6
Welche Kriterien/Parameter würden sich noch für die Formulierung eines Stopp-Kriteriums eignen? .....	6
Benötigen die Heuristiken verhältnismäßig mehr Paarvergleiche für die Workshops die subjektive Kriterien enthalten? .....	7
Greift das Stopp-Kriterium genau so zuverlässig bei subjektiven, wie bei objektive Kriterien? .....	7
Reduziert sich die Zeit, die für einen Workshop mit aktiver optimaler Heuristik und Stoppkriterium benötigt wird, im selben Verhältnis wie die Anzahl der Paarvergleiche? ..	7
Offene Frage .....	7
Sind bei Überschneidungen im Rankingverlauf die beiden betroffenen Ideen selbst dafür die Ursache, weil mindestens eine der beiden durch Paarvergleich(e) an dieser Stelle als Sieger oder Verlierer bewertet wurde? .....	7
Rankings .....	8
Subjektiver Workshop .....	8
Objektiver Workshop .....	9
Gemischter Workshop .....	10
Ranking aller Workshops .....	11

Diagramme .....	11
Subjektiver Workshop .....	12
Euklid.....	12
Absolut = Manhattan.....	12
Fußball.....	14
Heumar .....	14
Objektiver Workshop .....	15
Euklid.....	16
Absolut = Manhattan.....	16
Fußball.....	17
Heumar .....	18
Gemischter Workshop.....	19
Euklid.....	20
Absolut = Manhattan.....	21
Fußball.....	22
Heumar .....	23
Varianzen der Lösungsvektoren .....	24
Subjektiver Workshop .....	24
Euklid.....	24
Manhattan.....	25
Fußball.....	26
Heumar .....	27
Objektiver Workshop .....	28
Euklid.....	28
Manhattan.....	29
Fußball.....	30
Heumar .....	31

## Informationen zum Ablauf

Wir haben zwei Experimente durchgeführt mit Probanden woraus drei Workshops entstanden sind. Die Probanden sollten paarweise Ideen (Urlaubsorte) unter jeweils fünf Kriterien paarweise vergleichen. Ziel war es, jeweils die besten fünf Ideen zu finden, wobei die Rangfolge der Top 5 unerheblich ist. Ein Experiment enthält rein subjektive Kriterien:

Subjektiver Workshop	Gewichtung in %
angenehmes Klima	53
Gastfreundschaft / Mentalität	20
Sauberkeit	9
reichhaltige Kultur und Geschichte	12
Eignung für mehrjährigen Aufenthalt	6

Das zweite Experiment enthält subjektive Kriterien die auch einen objektiven Charakter vorweisen, denn diese sind messbarer im Vergleich zu den obigen Kriterien. Nachfolgend wird deshalb die Bezeichnung Objektiver Workshop zur Unterscheidung genutzt.

Objektiver Workshop	Gewichtung in %
Anreisekosten (aufgrund der Entfernung)	45
Infrastruktur (ärztl. Versorgung, Verkehrsnetz, Kommunikationsmedien)	5
Sicherheit (im Sinne von Kriminalität & politische Stabilität)	26
Angebotsvielfalt (mögl. Aktivitäten im näheren Umfeld)	17
Lebensunterhaltungskosten	7

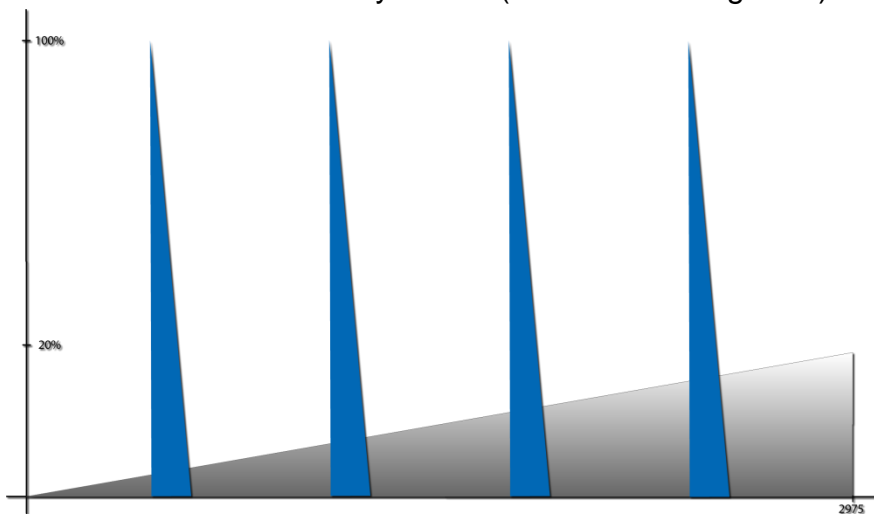
Der dritte Workshop enthält Kriterien aus beiden Experimenten und auch die dazugehörigen Nutzerentscheidungen.

Gemischter Workshop	Gewichtung in %
angenehmes Klima	25
Infrastruktur (ärztl. Versorgung, Verkehrsnetz, Kommunikationsmedien)	15
Gastfreundschaft / Mentalität	20
Sicherheit (im Sinne von Kriminalität & politische Stabilität)	20
reichhaltige Kultur und Geschichte	20

Die Heuristiken waren in den Experimenten nicht eingeschaltet. Wir haben mit RankIt 1.0 gearbeitet. Mit den Daten haben wir anschließend die Heuristiken gestartet. Zu beachten war hierbei folgendes: Wenn zwei Paarvergleiche als nächstes vorgeschlagen werden müssten, wird der Paarvergleich mit höchstem Kriteriengewicht ausgewählt. Die Daten sind nicht zu 100% exakt. RankIt 1.0 erfasst nicht alle Benutzerentscheidungen. Die fehlenden 558 (0,938 %) wurden durch eine Präferenzreihenfolge von Markus ersetzt, die wie folgt funktioniert: Den Ideen wurden Punkte zugeordnet zwischen 0 und 1000, für jedes Kriterium einzeln. Die Idee mit den meisten Punkten wurde Platz 1, die Idee mit den zweitmeisten Punkten wurde Platz 2 usw. Bei fehlenden Paarvergleichen in der Datenbank wurde diese Präferenzreihenfolge zu Hilfe genommen und für jeweils zwei Ideen daraus der Sieger und Gewinner ermittelt.

Beachtet wurde bei der Auswertung, dass die Nutzer während der Durchführung der Experimente nicht jeden Kriterienwechsel mitbekamen. Dieser fand alle 20% im

Experimentverlauf statt (viermal). Eine Befragung der Probanden ergab, dass der Wechsel erst einige Paarvergleiche später erkannt wurde. Wir gehen von jeweils 10 Vergleichen aus, die nicht unter dem neuen Kriterium stattfanden. Wenn der Benutzer bei der ersten Entscheidung nach Kriterienwechsel sich dem Wechsel nicht bewusst ist und spätestens bei der zehnten Idee den Wechsel mitbekommen hat, dann liegt der Erwartungswert der unbewussten Entscheidungen bei fünf Ideen. Die Wahrscheinlichkeit, dass jede dieser Entscheidungen falsch getroffen wurde beträgt 50 % also 10 Entscheidungen pro Benutzer sind pro Workshop wahrscheinlich falsch. Ebenso sank die Aufmerksamkeit der Probanden zeitlich linear. Wir gehen von 100%iger Aufmerksamkeit zu Beginn und 80% Aufmerksamkeit und damit Bewertungssicherheit am Ende der Experimente aus. Folgende Grafik dient zur Verdeutlichung. Die Anzahl der Paarvergleiche entspricht der x-Achse. Die Unaufmerksamkeit ist die y-Achse (nicht maßstabsgetreu).



Die Anzahl der unaufmerksamen Entscheidungen berechnet sich nach folgender Formel: Summe von  $x=1$  bis 2975 aus der Gleichung  $(0,2/2975)*x$ . 50 % davon sind wahrscheinlich falsch (148). Die Gesamtfehlerrate für die Entscheidungen liegt damit bei 5,78 %.

## Forschungsfrage

### ***Welche Heuristik bietet in einem computerbasierten Bewertungsworkshop die besten Voraussetzungen für die Anwendung eines Stopp-Kriteriums?***

Hierbei unbeachtet bleibt, ob sich das Ranking innerhalb der Top 5 Ideen im Verlauf ändert. Entscheidungsgrundlage ist, zu welchem Zeitpunkt die fünf besten Ideen absehbar feststehen. Aus den Diagrammen zur Fußballheuristik lässt die Erkenntnis gewinnen, dass sich diese nicht für ein Stopp-Kriterium eignet.

Im subjektiven Workshop zeigt die Heuristik Euklid im Vergleich zur absoluten Heuristik und der Heuristik Heumar den geradlinigsten und glattesten Verlauf. Die Top 5 stehen erst im letzten Drittel des Lösungsvektorverlaufs fest. Grund hierfür ist die starke Ähnlichkeit der 5. mit der 6. Idee. Die Top 6 würden sonst schon bei der Hälfte aller Paarvergleiche feststehen.

Im objektiven Workshop stehen die Top 5 unter allen Heuristiken erst im letzten Drittel des Lösungsvektorverlaufs fest. Die sechs besten Ideen für Euklid und Heumar stehen bereits

ab 6000 Paarvergleichen fest. Euklid zeigt nach 6000 von 29751 Paarvergleichen einen annähernd linearen Verlauf.

Im gemischten Workshop sind die Kriteriengewichte gleichmäßiger verteilt. Euklid zeigt unter dieser Bedingung das beste Ergebnis. Die Top 5 Ideen stehen nach 13000 Paarvergleichen ( $\frac{1}{3}$ ) fest. Würde hier schon ein Stopp-Kriterium greifen, ließe sich noch keine sichere Aussage darüber treffen, ob das Ranking im späteren Verlauf noch Schwankungen unterliegt.

Zusammenfassend: Die Heuristik Euklid bietet die besten Voraussetzungen für die Anwendung eines Stopp-Kriteriums. In jedem Workshop ist nach 33 % ein Trend des weiteren Verlaufs erkennbar. Signifikante Änderungen treten danach nichtmehr auf.

Zur weiteren Analyse sei auf die Diagramme ab Seite 12 verwiesen.

## **Eigene Forschungsfrage**

### ***Wie wirken sich objektive und subjektive Bewertungskriterien auf den Rankingprozess aus?***

Da die Fußball-Heuristik den stärksten Schwankungen im Lösungsvektor unterliegt, wird diese nichtmehr betrachtet.

Die Probanden zeigten bei subjektiveren Kriterien mehr Schwankungen in ihren Entscheidungen, was sich in dem Verlauf des Lösungsvektors widerspiegelt. Nach 6000 bis 8000 Paarvergleichen lassen sich die Ideen in Gruppen einteilen, innerhalb derer es noch zu Überschneidungen kommt. Jedoch wechselt eine Idee unter den 30% der Besten nichtmehr in eine Verlierergruppe unterhalb von 50% der besten Ideen. Der weitere Rankingprozess ist zu diesem Zeitpunkt bereits absehbar, aber noch nicht entschieden.

Unter objektiveren Kriterien waren die Probanden sicherer in ihren getroffenen Entscheidungen. Nach 3000 bis 4000 Entscheidungen (von 29751) zeichnen sich schon deutlich die Gewinner ab. Es sind im Vergleich zu subjektiven Kriterien die Hälfte der Paarvergleiche nötig.

Die Top 5 Ideen ändern sich unter beiden Kriterien zwischen dem erwähnten Intervall und dem Ende noch mehrmals. Das liegt an der Ähnlichkeit einiger Ideen. Würde ein Workshop durchgeführt werden, mit mehr Unterschieden der Ideen in ihren Eigenschaften und echt objektiven Kriterien, würde sich ein feststehendes Ranking möglicherweise schon eher abzeichnen.

## Auswertungskriterien

### ***Jeweils für objektive und subjektive Kriterien: Ab wie vielen Paarvergleichen erreichen die einzelnen Heuristiken die korrekte Platzierung der Top 5?***

Die Werte sind jeweils auf volle 1000 Paarvergleiche aufgerundet. In den Diagrammen ist die zugehörige Stelle mit einem roten Pfeil darunter markiert.

Heuristiken	Subjektiver Workshop	Objektiver Workshop
Euklid	20000 (67 %)	27000 (91 %)
Absolut	25000 (84 %)	25000 (84 %)
Fußball	ca. 29750 (100 %)	29000 (97 %)
Heumar	28000 (94 %)	28000 (94 %)

### ***Im Vergleich dazu: Wann hätte das konkrete Stopp-Kriterium gegriffen?***

René Chelvier'swissenschaftlich fundiertes Stopp-Kriterium ist zum Zeitpunkt des Projektabschlusses noch nicht vollständig erarbeitet. An Stelle seines Kriteriums, haben wir selbst ein Stopp-Kriterium entwickelt. Es wird nach 20 % der Ideenbewertung gestartet und arbeitet nach folgendem Prinzip. Fortlaufend nach jedem getätigten Paarvergleich wird im Bereich der letzten 10% bis zum aktuellen Paarvergleich überprüft, ob im gesamten Ranking Platzwechsel stattfanden. Hat sich das Ranking während dieser letzten 10% nicht mindestens einmal geändert, kann der Bewertungsprozess abgebrochen werden. Dieser Fall ist bei keiner Heuristik eingetreten.

## Weitere Fragestellungen

### ***Bei welcher Heuristik griff das Stopp-Kriterium nach korrekt ermittelter Platzierung der Top 5 am frühesten (Zeitersparnis unter Gewährleistung der Korrektheit)?***

Diese Frage lässt sich nicht beantworten, da unser Stopp-Kriterium nicht griff.

### ***Welche Kriterien/Parameter würden sich noch für die Formulierung eines Stopp-Kriteriums eignen?***

Transitivität ließe sich nutzen um Paarvergleiche einzusparen und deren Entscheidung abzuleiten. Wenn Idee C gegen B gewinnt und Idee B gegen A gewinnt, dann gewinnt auch C gegen A. Dieser letzte Paarvergleich müsste nicht durchgeführt werden. Beachtet werden muss allerdings, dass Menschen sich weniger transitiv entscheiden, je subjektiver das Kriterium ist unter dem die Entscheidung gefällt werden soll oder je mehr Ähnlichkeiten die betroffenen Ideen zueinander aufweisen.

***Benötigen die Heuristiken verhältnismäßig mehr Paarvergleiche für die Workshops die subjektive Kriterien enthalten?***

Eine genaue Aussage lässt sich nicht treffen ohne Vorliegen des wissenschaftlichen Stoppkriteriums. Jedes Stoppkriterium, dass sich an den Varianzen im Lösungsvektor orientiert, sollte jedoch unter subjektiveren Kriterien mehr Paarvergleiche benötigen. Zum Vergleich siehe oben: Um einen annähernd glatten Verlauf der Lösungsvektoren unter den sechs besten Ideen zu erzielen benötigt Euklid bei subjektiven Kriterien 2,5x so viele Paarvergleiche wie unter objektiveren Kriterien.

***Greift das Stopp-Kriterium genau so zuverlässig bei subjektiven, wie bei objektive Kriterien?***

Darüber lässt sich keine Aussage treffen, aus oben genanntem Grund. Anhand der Verläufe der Lösungsvektoren lässt sich allerdings vermuten, dass bei objektiveren Kriterien ein Stopp-Kriterium zuverlässiger greifen würde.

***Reduziert sich die Zeit, die für einen Workshop mit aktiver optimaler Heuristik und Stoppkriterium benötigt wird, im selben Verhältnis wie die Anzahl der Paarvergleiche?***

Keine Aussage möglich.

***Offene Frage***

**Sind bei Überschneidungen im Rankingverlauf die beiden betroffenen Ideen selbst dafür die Ursache, weil mindestens eine der beiden durch Paarvergleich(e) an dieser Stelle als Sieger oder Verlierer bewertet wurde?**



# Rankings

## *Subjektiver Workshop*

Reihenfolge der Urlaubsorte (Ideen) nach absteigendem Lösungsvektorwert sortiert:  
Die Top 5 sind: Fiji, Athen, Dubai, Rom und Sydney.

ID	Urlaubsort	Lösungsvektorwert
8	Fiji	7,793%
2	Athen	6,001%
7	Dubai	5,679%
24	Rom	5,496%
29	Sydney	4,537%
15	Madrid	4,385%
17	Miami	4,383%
27	San Francisco	4,221%
9	Havana	4,138%
34	Venedig	3,456%
13	Kapstadt	3,441%
1	Amsterdam	3,373%
35	Wien	3,020%
21	Paris	2,922%
19	New York	2,794%
14	London	2,705%
23	Prag	2,385%
12	Kairo	2,286%
28	Stockholm	2,248%
16	Mexico City	2,242%
6	Budapest	2,214%
5	Bodensee	2,186%
3	Bagdad	1,932%
25	Rügen	1,894%
11	Jerusalem	1,888%
20	Normandie	1,879%
22	Peking	1,788%
4	Berlin	1,566%
10	Hong Kong	1,541%
33	Vancouver	1,517%
31	Tokio	1,434%
30	Tibet	1,011%
18	Moskau	0,964%
26	Sahara	0,406%
32	Tschernobyl	0,274%

## **Objektiver Workshop**

Reihenfolge der Urlaubsorte (Ideen) nach absteigendem Lösungsvektorwert sortiert:  
Die Top 5 sind: Berlin, Rügen, Wien, Prag und London.

ID	Urlaubsort	Lösungsvektorwert
4	Berlin	9,760%
25	Rügen	7,284%
35	Wien	6,985%
23	Prag	6,707%
14	London	6,251%
5	Bodensee	5,961%
21	Paris	5,392%
28	Stockholm	5,226%
1	Amsterdam	5,171%
24	Rom	4,156%
34	Venedig	3,679%
20	Normandie	3,015%
15	Madrid	2,993%
6	Budapest	2,580%
19	New York	2,095%
29	Sydney	1,804%
27	San Francisco	1,747%
2	Athen	1,735%
33	Vancouver	1,599%
7	Dubai	1,590%
17	Miami	1,431%
18	Moskau	1,414%
31	Tokio	1,149%
12	Kairo	0,996%
13	Kapstadt	0,989%
22	Peking	0,976%
26	Sahara	0,941%
16	Mexico City	0,918%
10	Hong Kong	0,875%
11	Jerusalem	0,859%
32	Tschernobyl	0,793%
3	Bagdad	0,773%
9	Havana	0,744%
8	Fiji	0,734%
30	Tibet	0,677%

## ***Gemischter Workshop***

Reihenfolge der Urlaubsorte (Ideen) nach absteigendem Lösungsvektorwert sortiert:  
Die Top 5 sind: Rom, Sydney, Wien, San Francisco und Madrid.

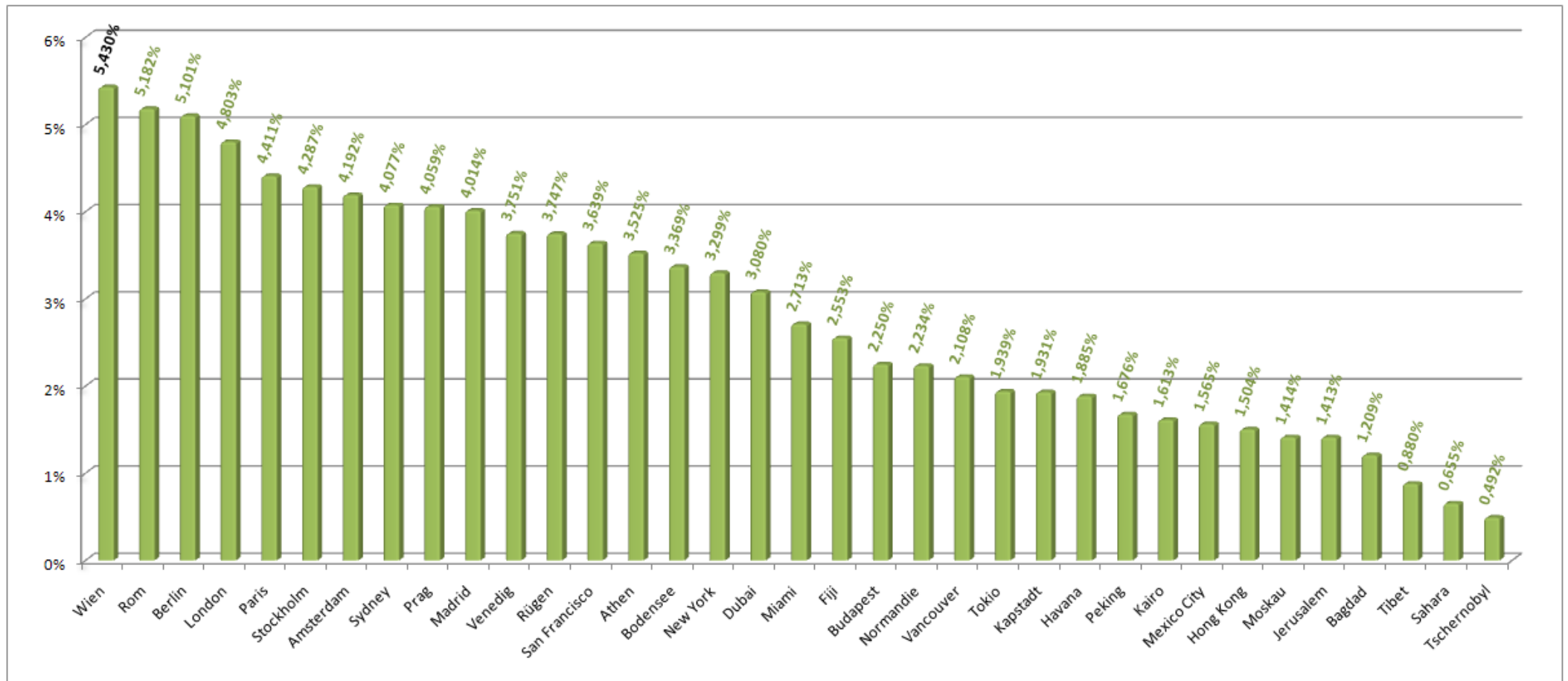
ID	Urlaubsort	Lösungsvektorwert
24	Rom	5,735%
29	Sydney	5,349%
35	Wien	4,979%
27	San Francisco	4,658%
15	Madrid	4,553%
14	London	4,354%
28	Stockholm	4,131%
19	New York	4,122%
21	Paris	4,120%
2	Athen	4,104%
34	Venedig	3,853%
1	Amsterdam	3,769%
7	Dubai	3,454%
17	Miami	3,149%
4	Berlin	3,012%
23	Prag	2,805%
8	Fiji	2,596%
33	Vancouver	2,531%
31	Tokio	2,514%
13	Kapstadt	2,194%
9	Havana	2,119%
22	Peking	2,073%
6	Budapest	2,059%
5	Bodensee	2,050%
25	Rügen	1,995%
10	Hong Kong	1,874%
12	Kairo	1,849%
20	Normandie	1,836%
16	Mexico City	1,818%
11	Jerusalem	1,651%
18	Moskau	1,504%
3	Bagdad	1,325%
30	Tibet	0,976%
26	Sahara	0,533%
32	Tschernobyl	0,355%

## Ranking aller Workshops

Reihenfolge der Urlaubsorte (Ideen) nach absteigendem Lösungsvektorwert sortiert:

Die Top 5 sind: Wien, Rom, Berlin, London und Paris.

Die Werte sind aus den Workshops aggregiert. Der subjektive Workshop geht dabei mit fünffacher Wertigkeit ein; der objektive mit dreifacher und der gemischte mit einfacher Wertigkeit.



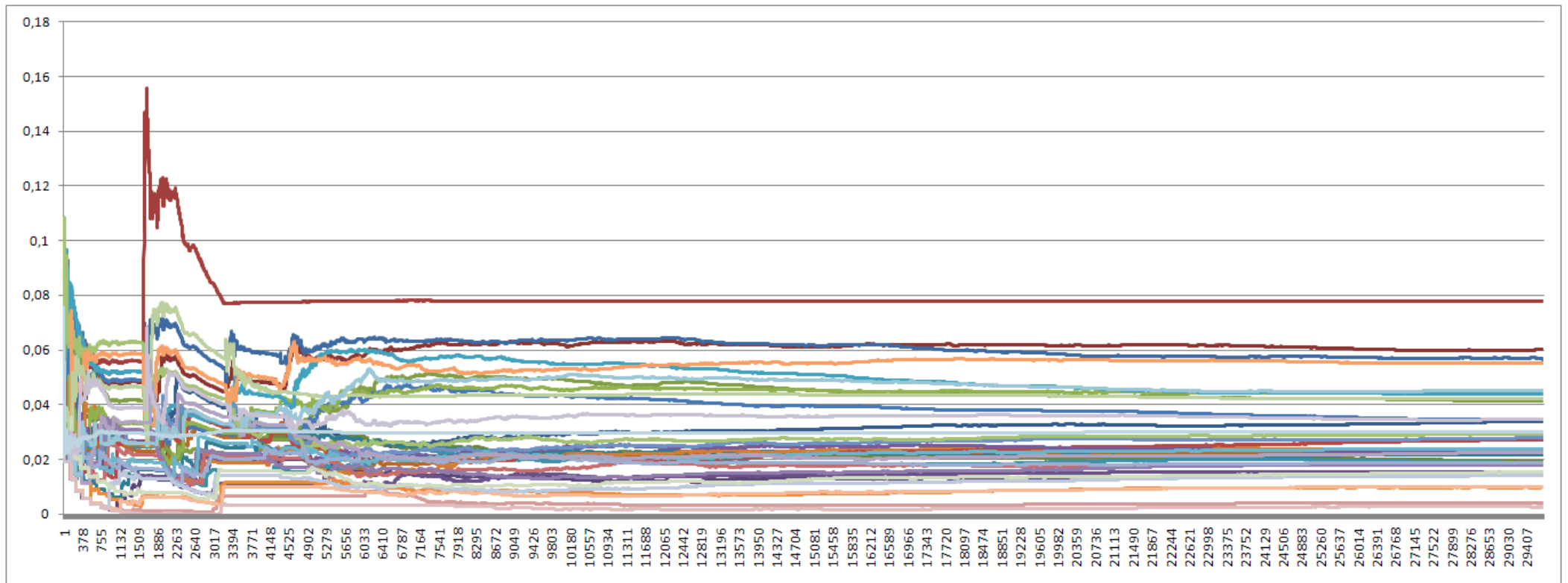
# Diagramme

## Subjektiver Workshop

### Euklid

x-Achse – Anzahl Paarvergleiche

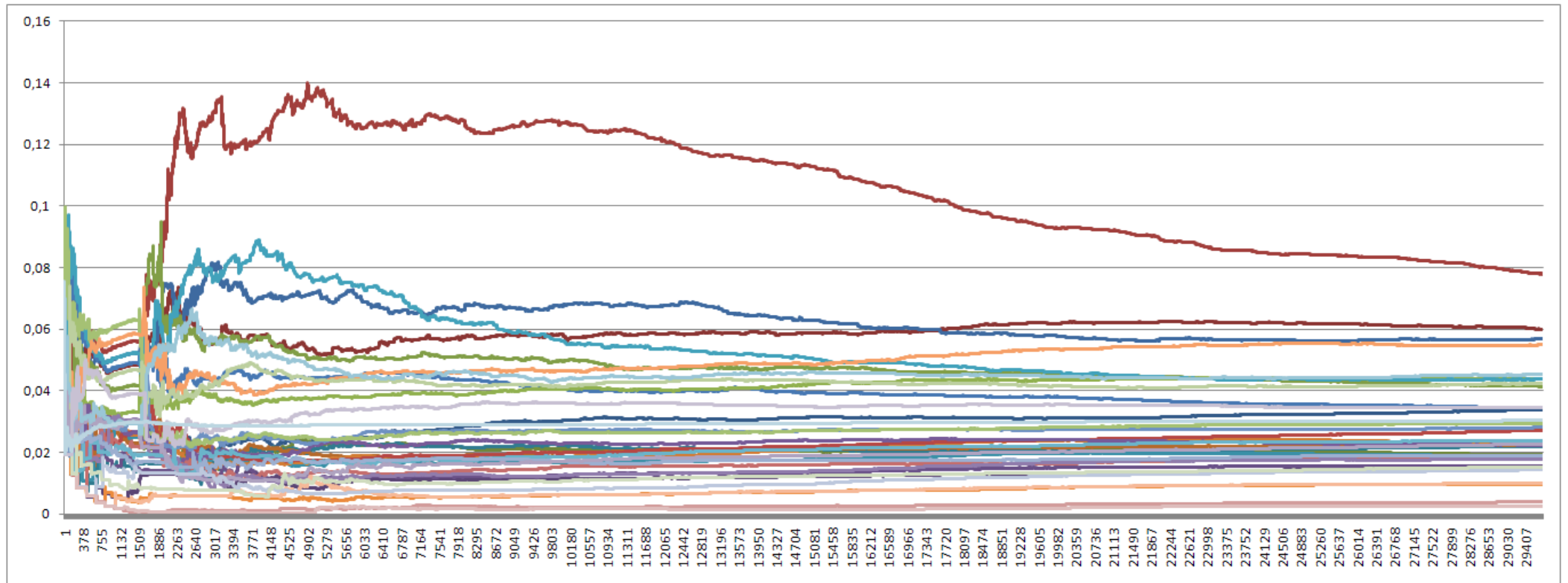
y-Achse – prozentualer Rankinganteil



## Absolut = Manhattan

x-Achse – Anzahl Paarvergleiche

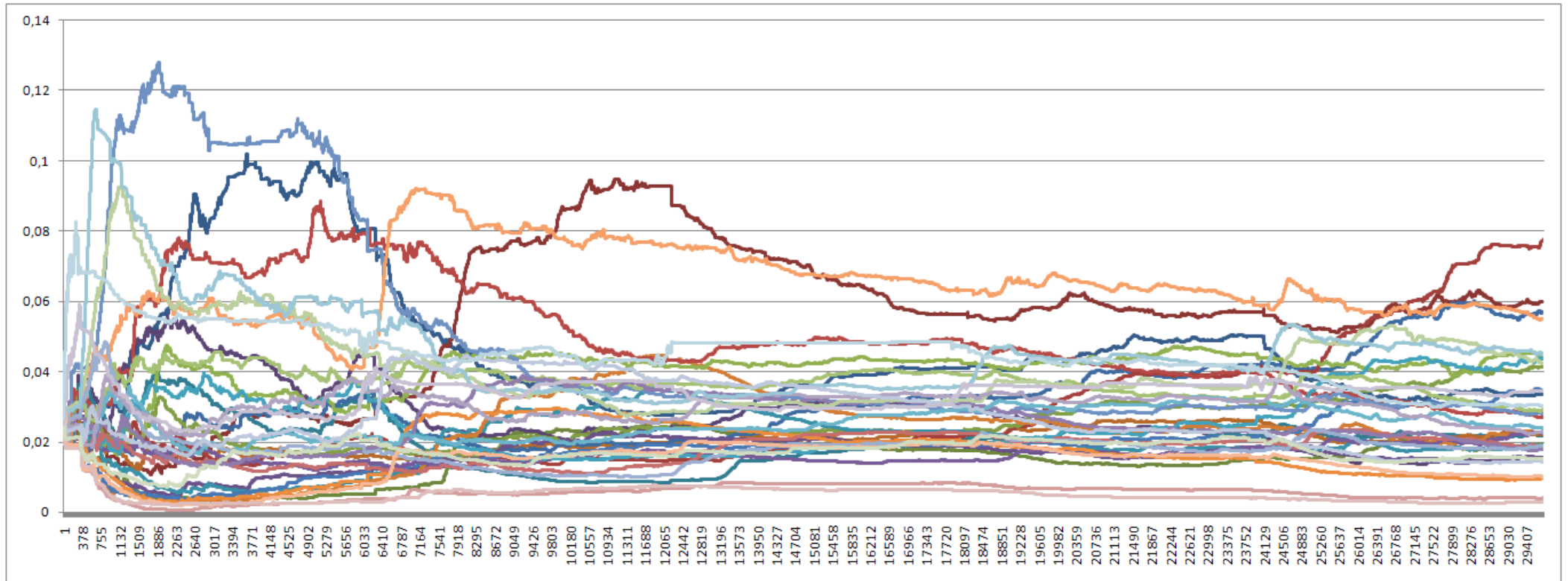
y-Achse – prozentualer Rankinganteil



# Fußball

x-Achse – Anzahl Paarvergleiche

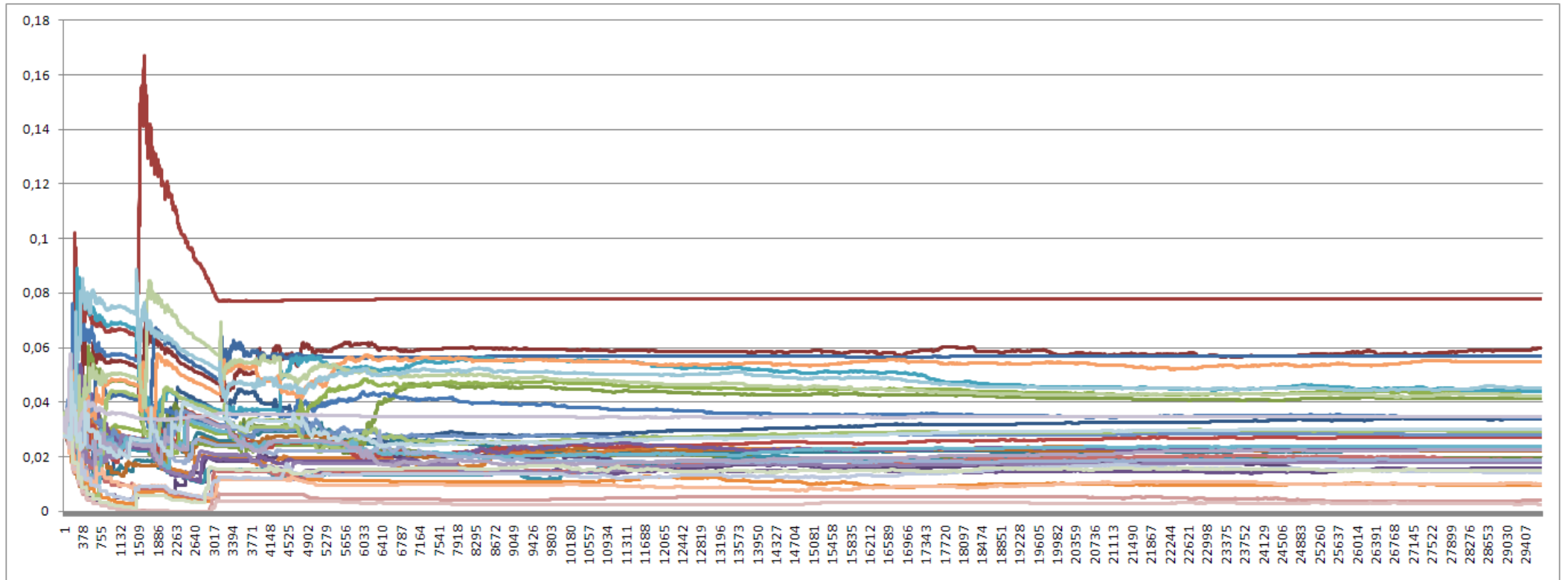
y-Achse – prozentualer Rankinganteil



# Heumar

x-Achse – Anzahl Paarvergleiche

y-Achse – prozentualer Rankinganteil



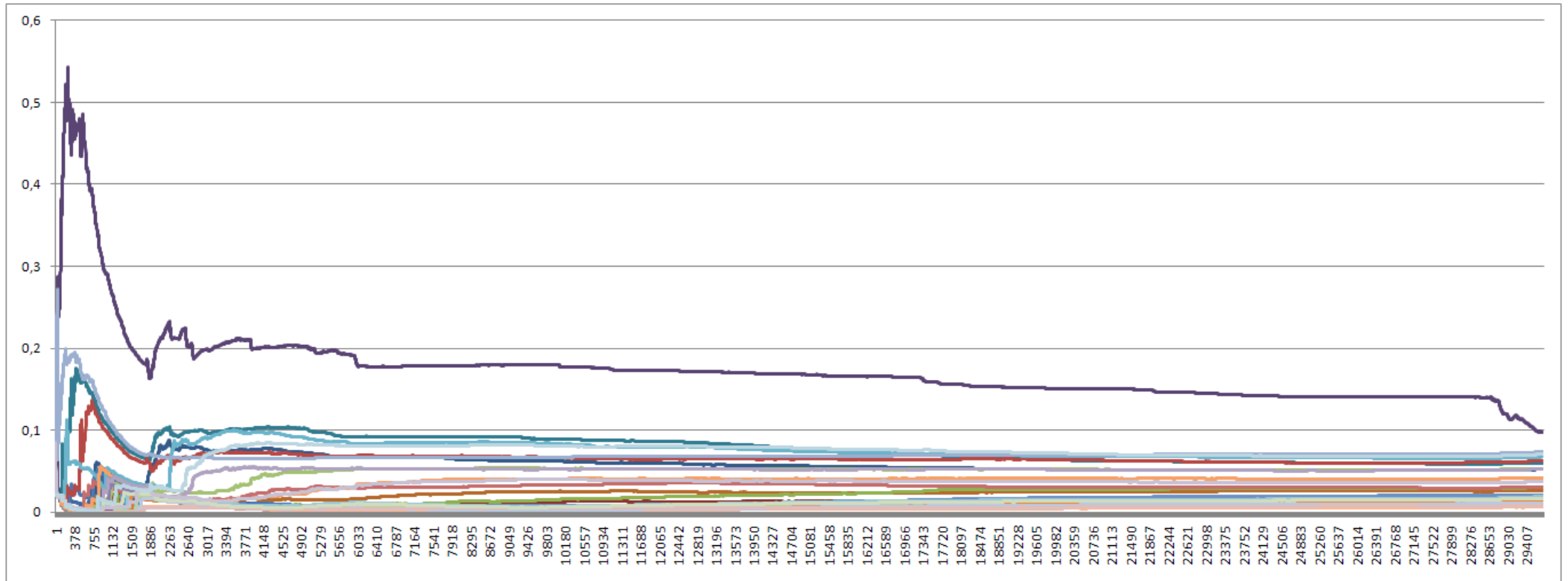


# Objektiver Workshop

## Euklid

x-Achse – Anzahl Paarvergleiche

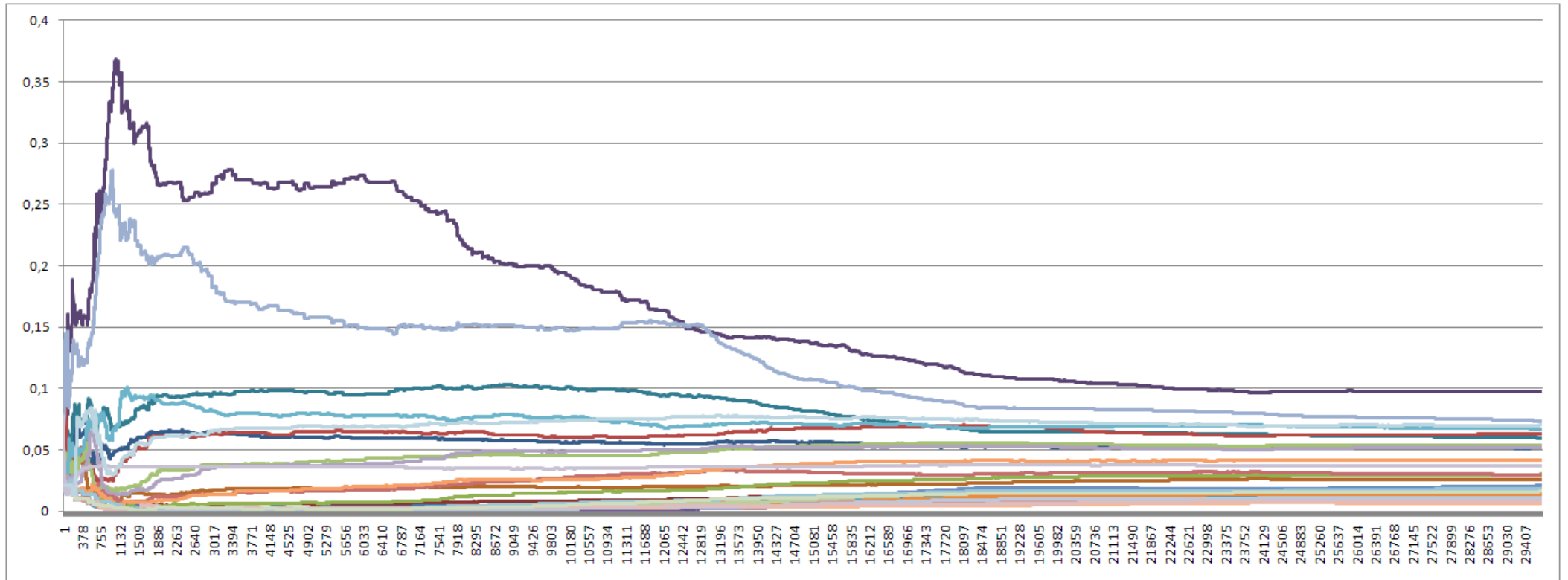
y-Achse – prozentualer Rankinganteil



## Absolut = Manhattan

x-Achse – Anzahl Paarvergleiche

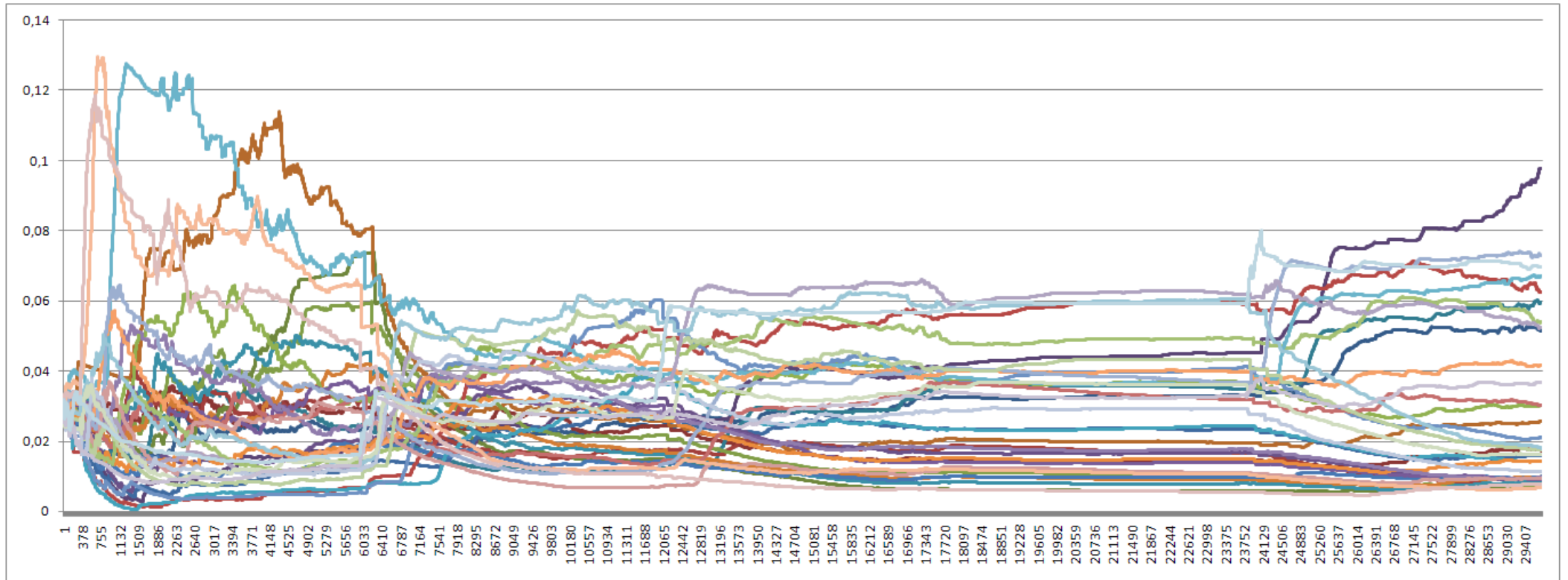
y-Achse – prozentualer Rankinganteil



# Fußball

x-Achse – Anzahl Paarvergleiche

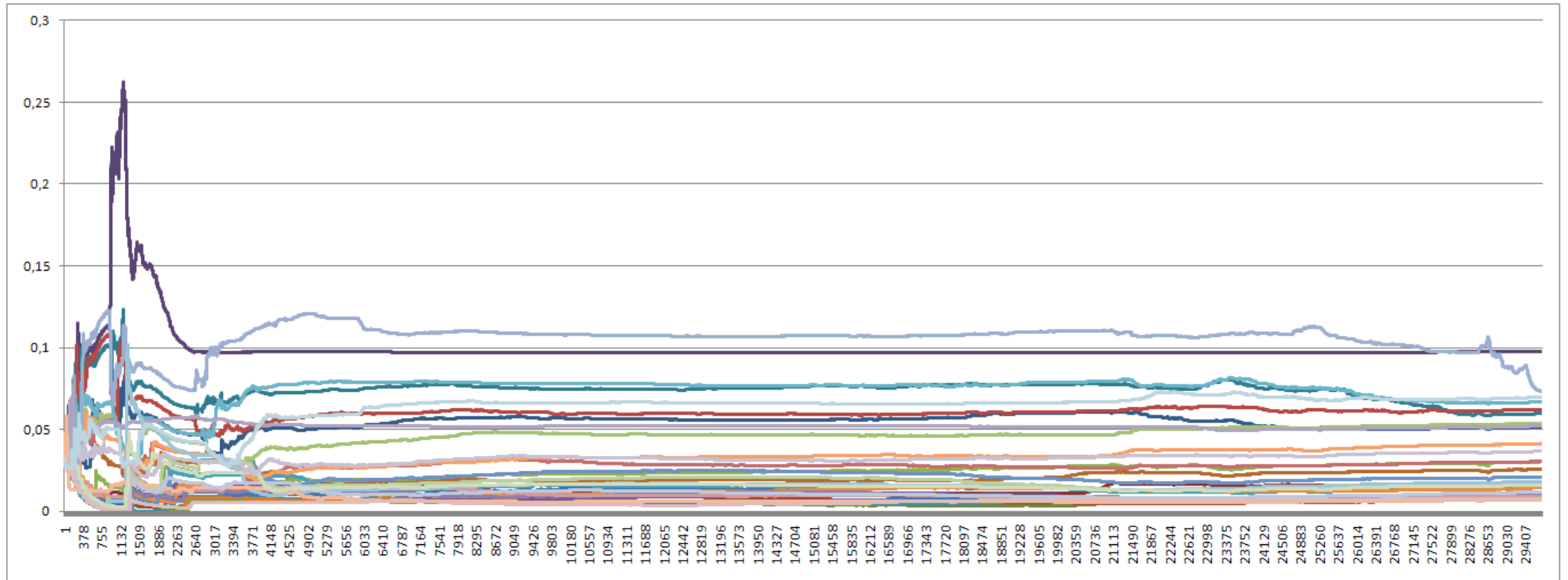
y-Achse – prozentualer Rankinganteil



# Heumar

x-Achse – Anzahl Paarvergleiche

y-Achse – prozentualer Rankinganteil

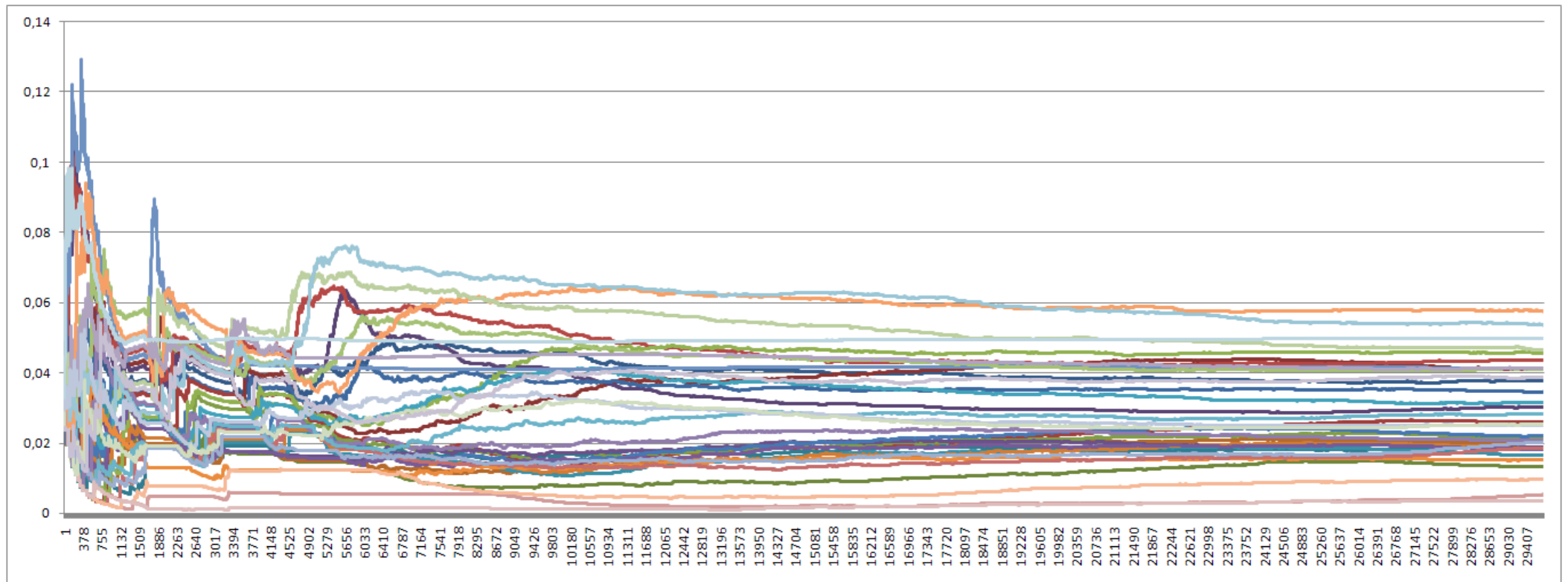


## Gemischter Workshop

### Euklid

x-Achse – Anzahl Paarvergleiche

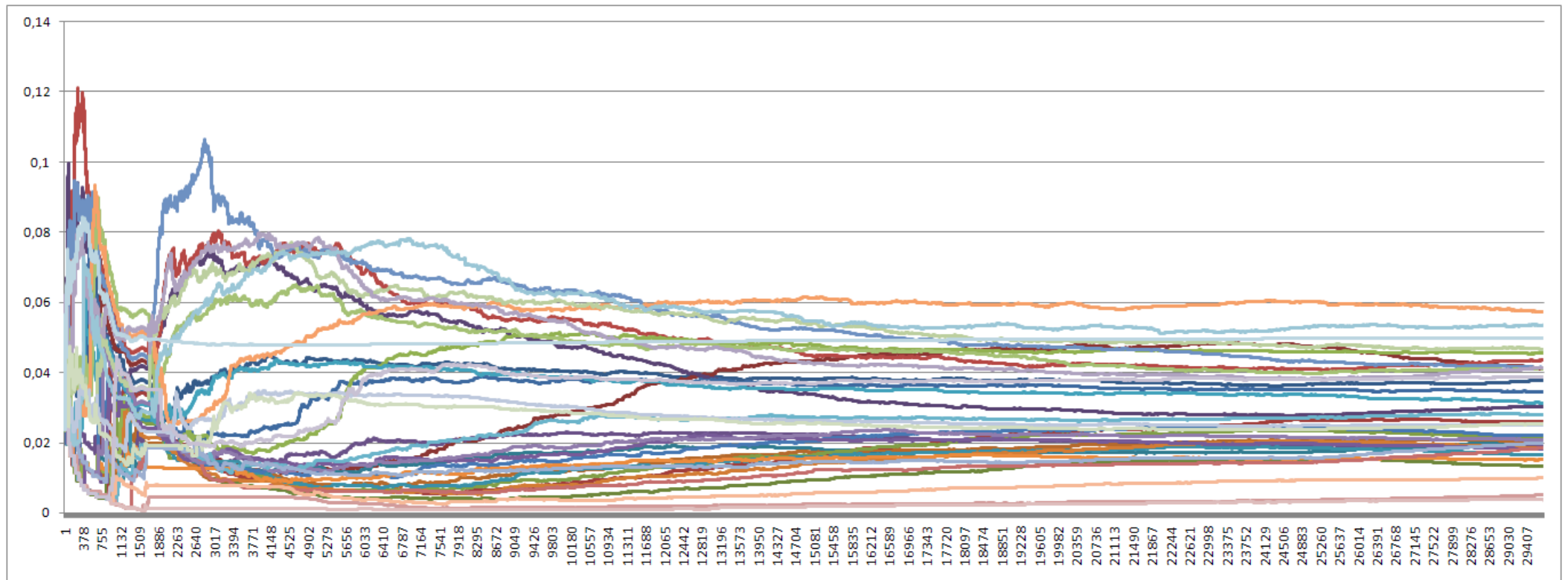
y-Achse – prozentualer Rankinganteil



## Absolut = Manhattan

x-Achse – Anzahl Paarvergleiche

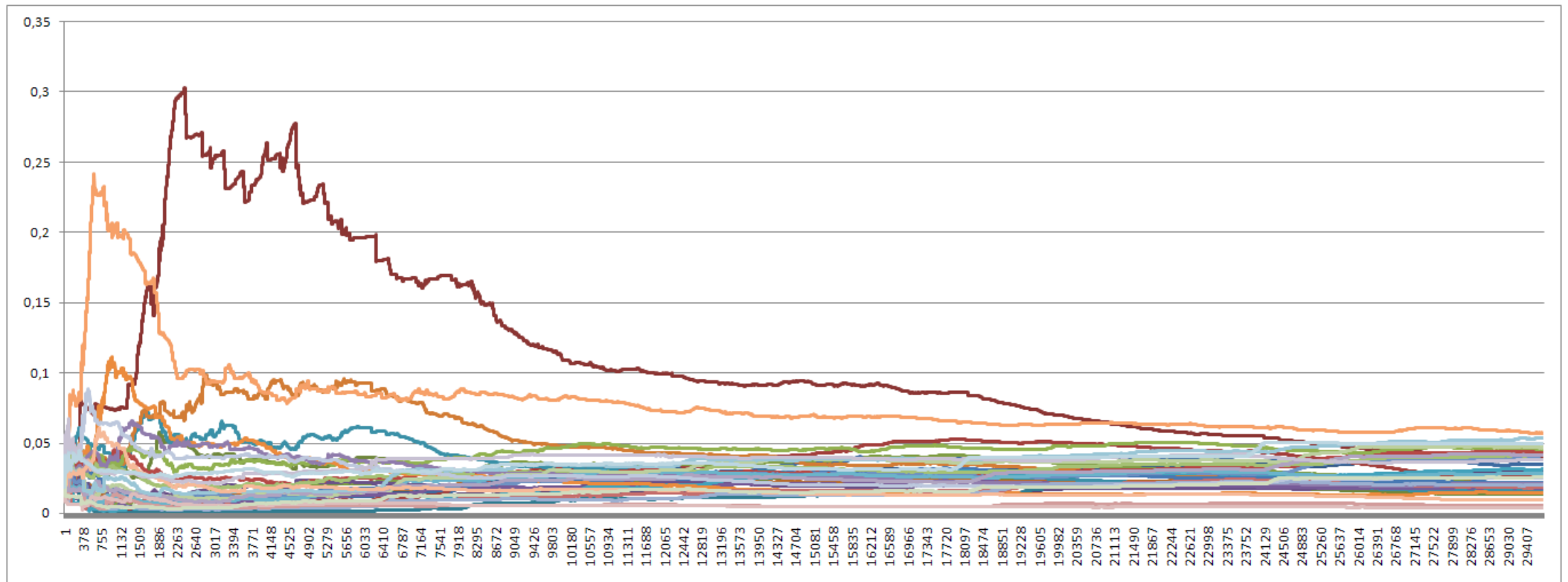
y-Achse – prozentualer Rankinganteil



# Fußball

x-Achse – Anzahl Paarvergleiche

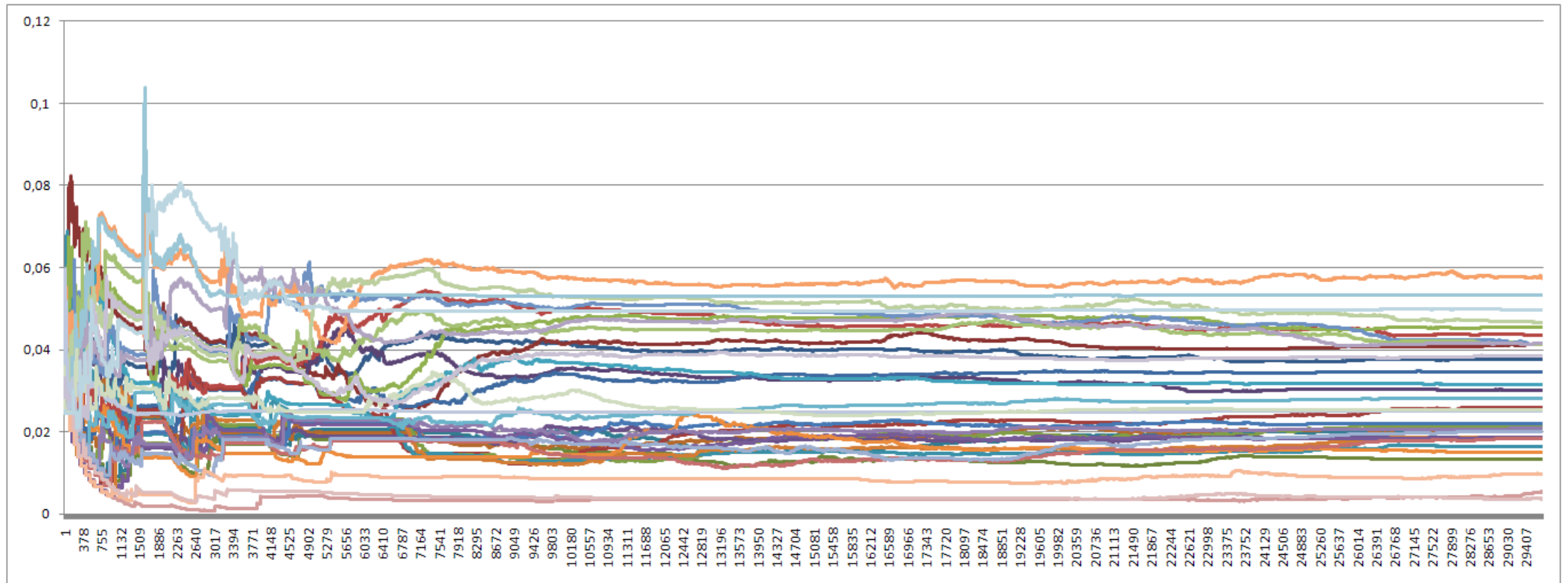
y-Achse – prozentualer Rankinganteil



# Heumar

x-Achse – Anzahl Paarvergleiche

y-Achse – prozentualer Rankinganteil





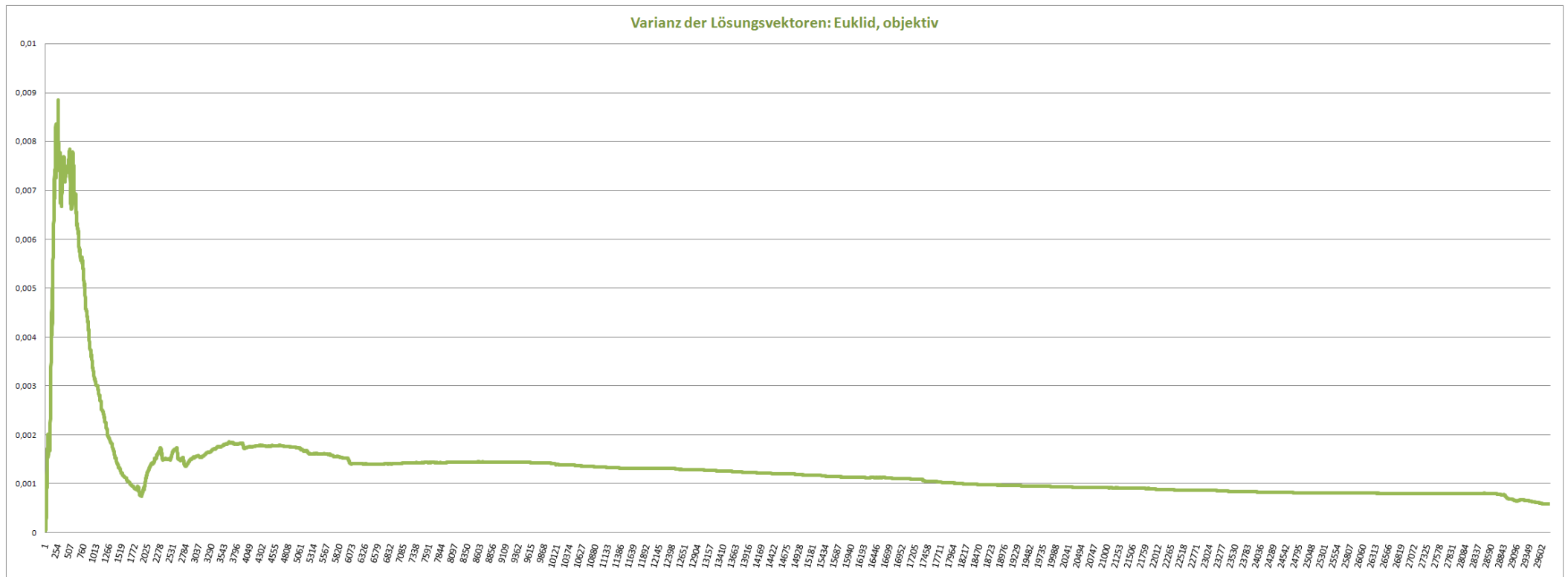
# Varianzen der Lösungsvektoren

## Subjektiver Workshop

### Euklid

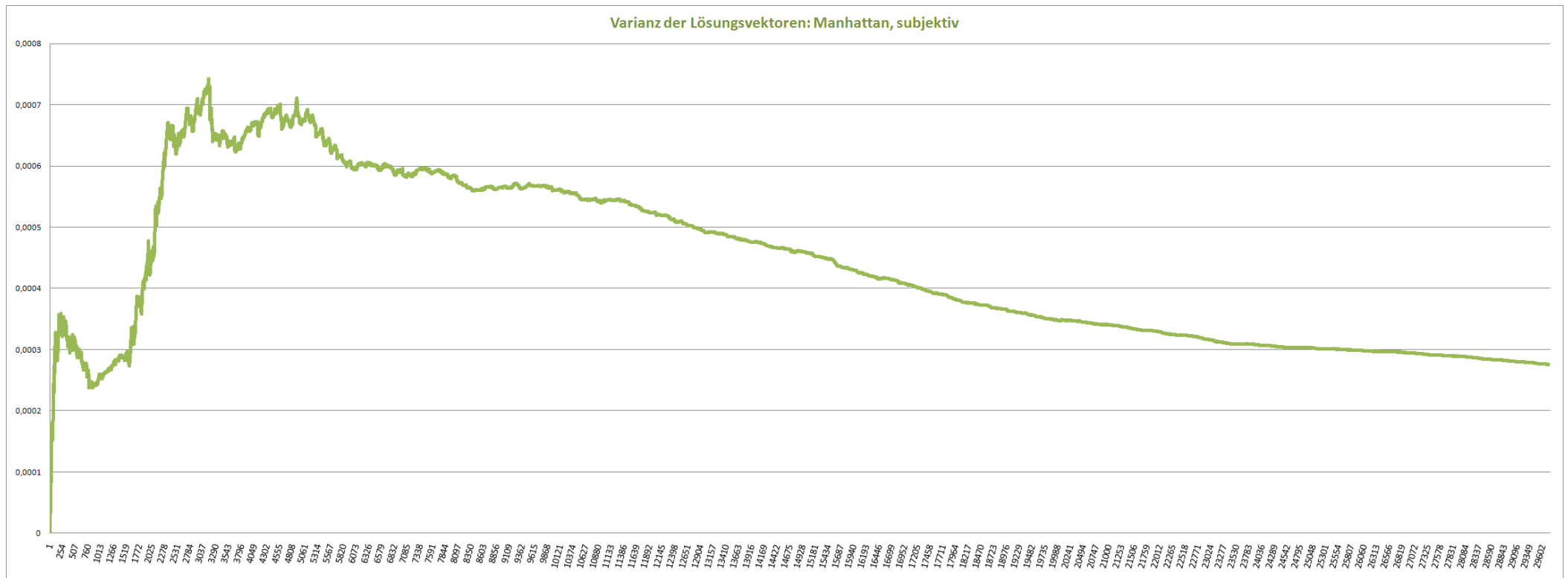
x-Achse – Anzahl Paarvergleiche

y-Achse – Varianz



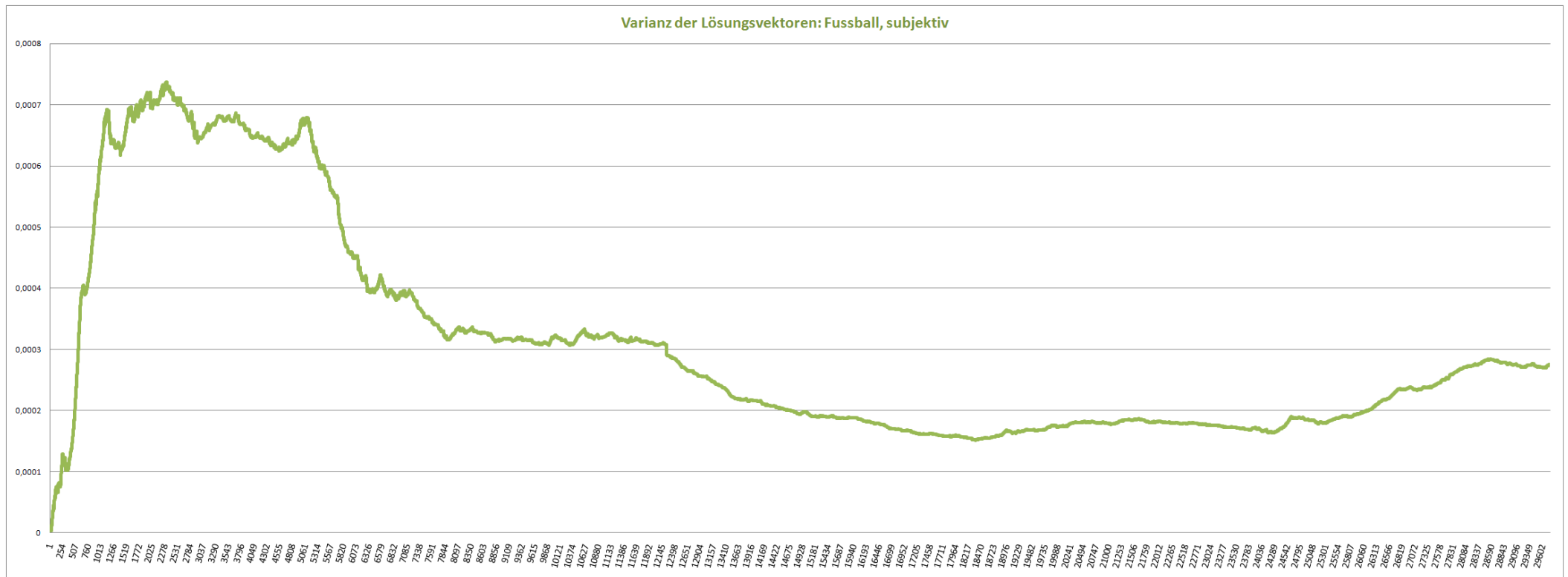
# Manhattan

x-Achse – Anzahl Paarvergleiche  
y-Achse – Varianz



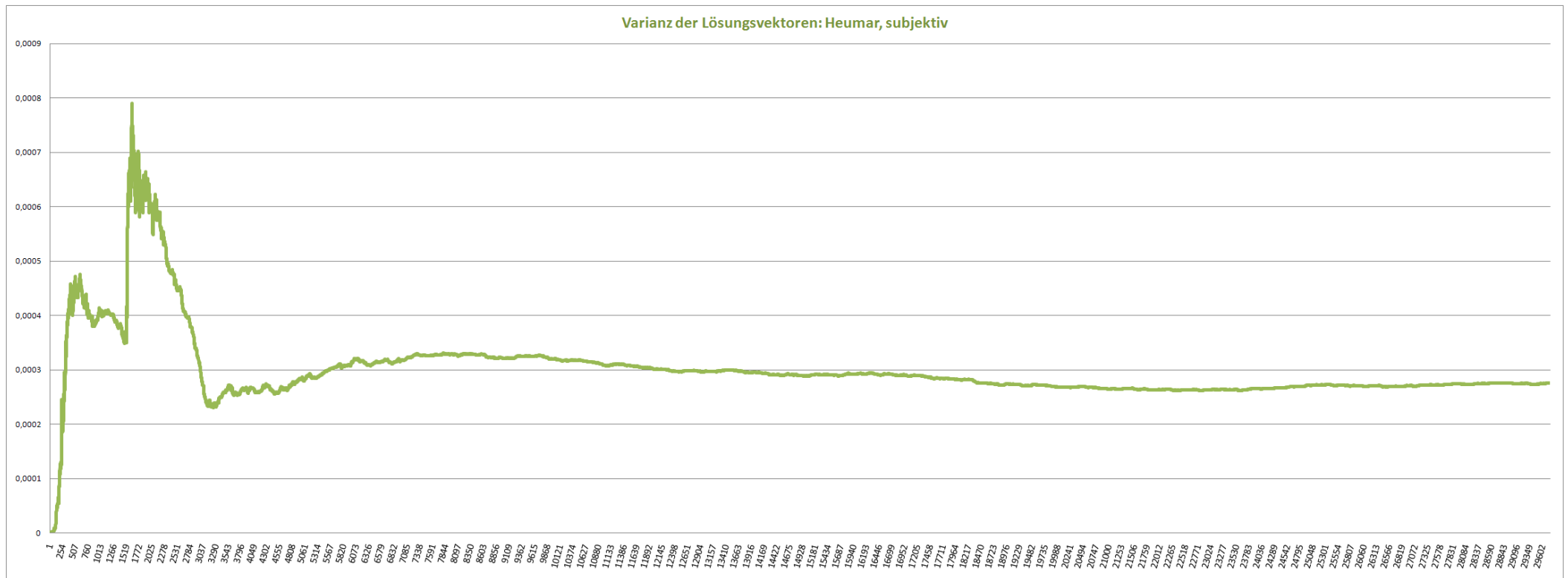
# Fußball

x-Achse – Anzahl Paarvergleiche  
y-Achse – Varianz



# Heumar

x-Achse – Anzahl Paarvergleiche  
y-Achse – Varianz

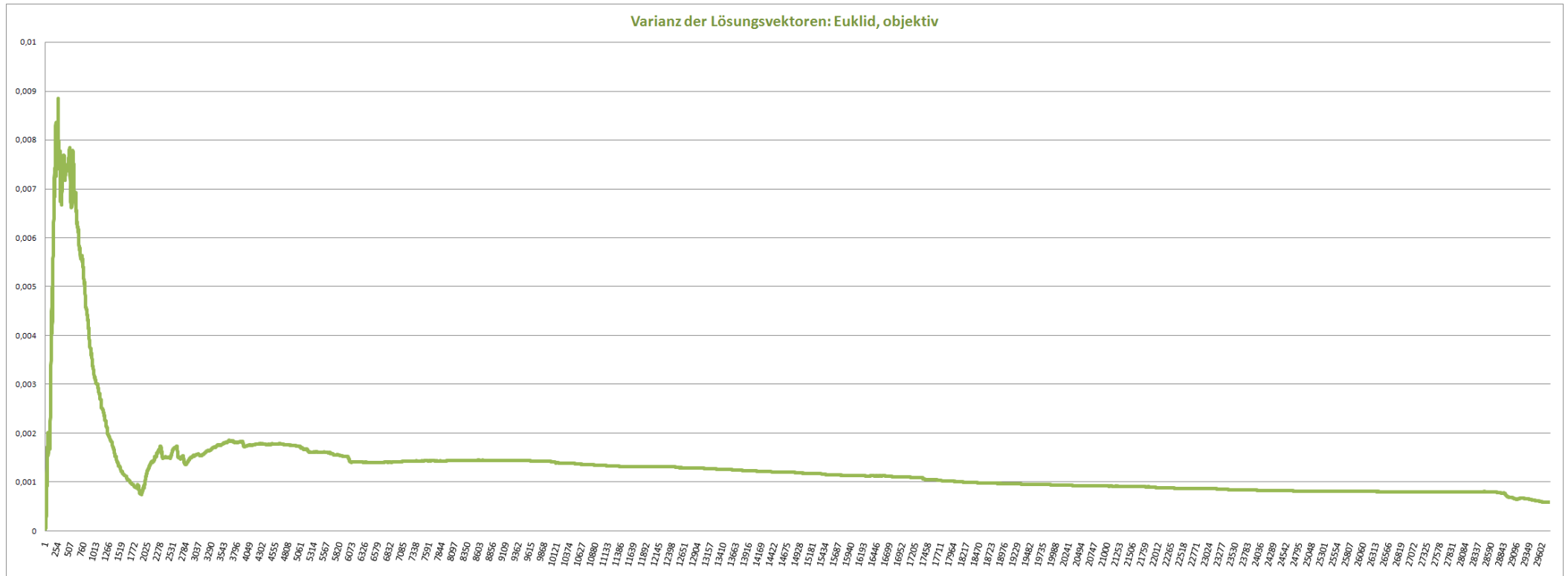


# Objektiver Workshop

## Euklid

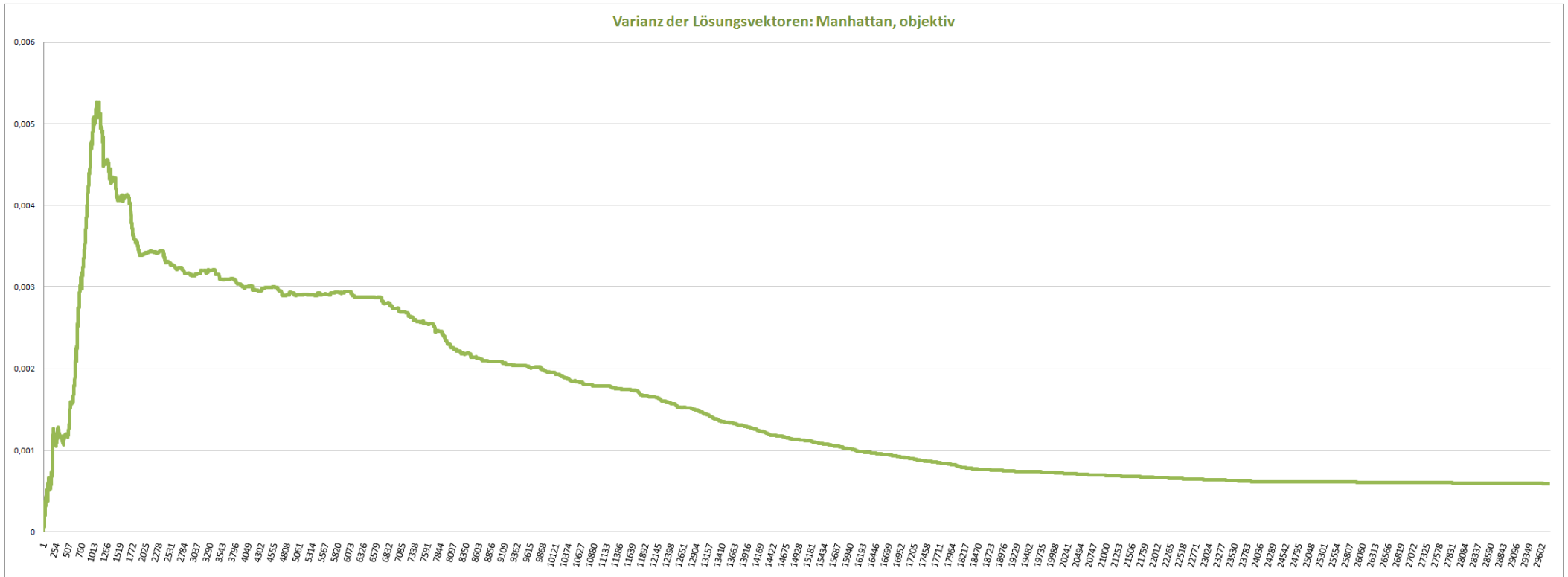
x-Achse – Anzahl Paarvergleiche

y-Achse – Varianz



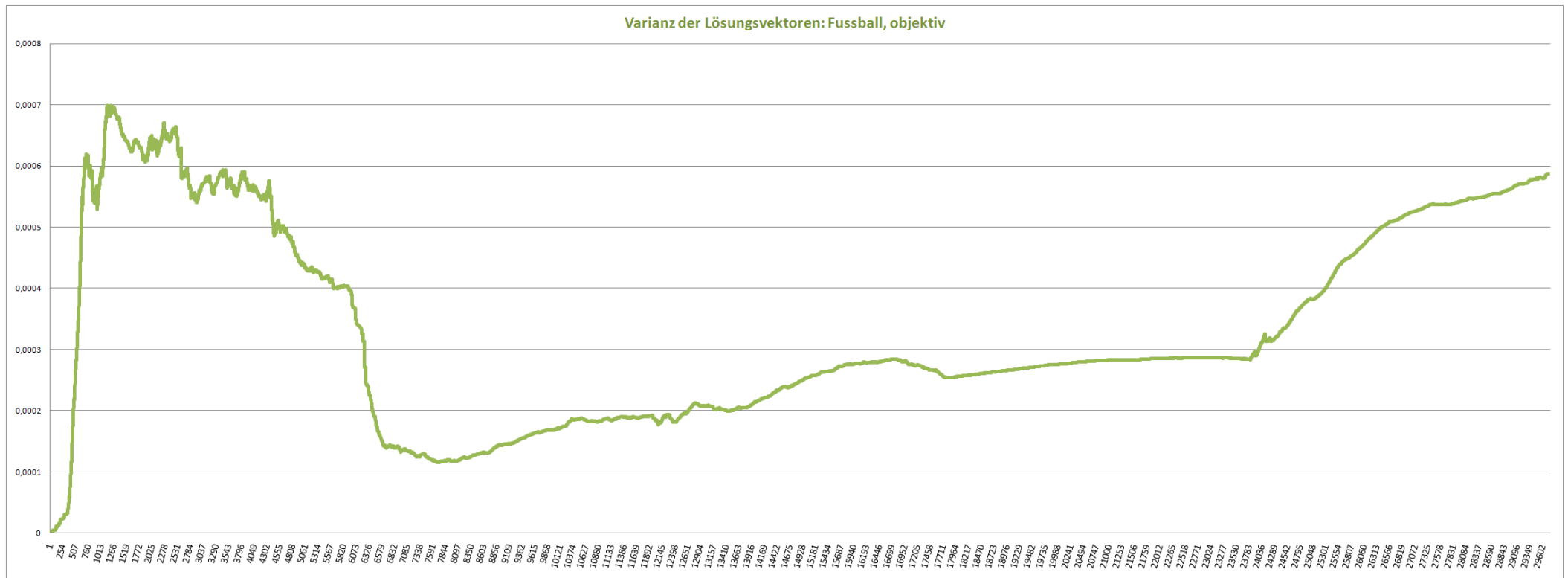
# Manhattan

x-Achse – Anzahl Paarvergleiche  
y-Achse – Varianz



# Fußball

x-Achse – Anzahl Paarvergleiche  
y-Achse – Varianz



# Heumar

x-Achse – Anzahl Paarvergleiche  
y-Achse – Varianz

