



**ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ**  
**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ**  
**ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ**

**Τεχνικές Εξόρυξης Δεδομένων**

**ΕΡΓΑΣΙΑ №2**

**Γαλούνη Κωνσταντίνα  
Δελβινιώτης Σπύρος**

## Αναλύσεις

Μετά την υλοποίηση του αλγορίθμου LCSS έγιναν αναλύσεις πάνω στα δεδομένα που μας είχαν δοθεί. Τα αποτελέσματα είχαν μεγάλο ενδιαφέρον και ακολουθεί η ανάλυσή τους.

Πιο συγκεκριμένα, για τον φάκελο των 1000 σημείων παρατηρήθηκαν τα μικρότερα ποσοστά ομοιότητας, σε σχέση με τα δεδομένα των υπόλοιπων φακέλων. Τα αποτελέσματα είναι αναμενόμενα, αν παρατηρήσουμε τις ακολουθίες που συγκρίνουμε στο χάρτη ή στην αναπαράσταση σε άξονες x,y.

Η επιλογή του «ε» έγινε μετά από δοκιμές σε συγκεκριμένα δεδομένα. Πιο ειδικά τα αποτελέσματα βρίσκονται παρακάτω:

Αρχείο A	Αρχείο B	$\epsilon$	Similarity	Elapsed Time (sec)
366.txt	3557.txt	0.1	0.2 %	0.0965
366.txt	3557.txt	0.05	0.1 %	0.1478
366.txt	3557.txt	0.005	0.0 %	0.0214

Παρατηρούμε ότι καθώς το  $\epsilon$  μειώνεται, η ομοιότητα αυξάνεται. Αυτό είναι αναμενόμενο αφού αυξάνοντας το  $\epsilon$ , αυξάνουμε και την ανεκτικότητα σε σφάλμα. Επιλέξαμε το 0.05 ως ένα ικανό σφάλμα για την ανάλυση των δεδομένων.

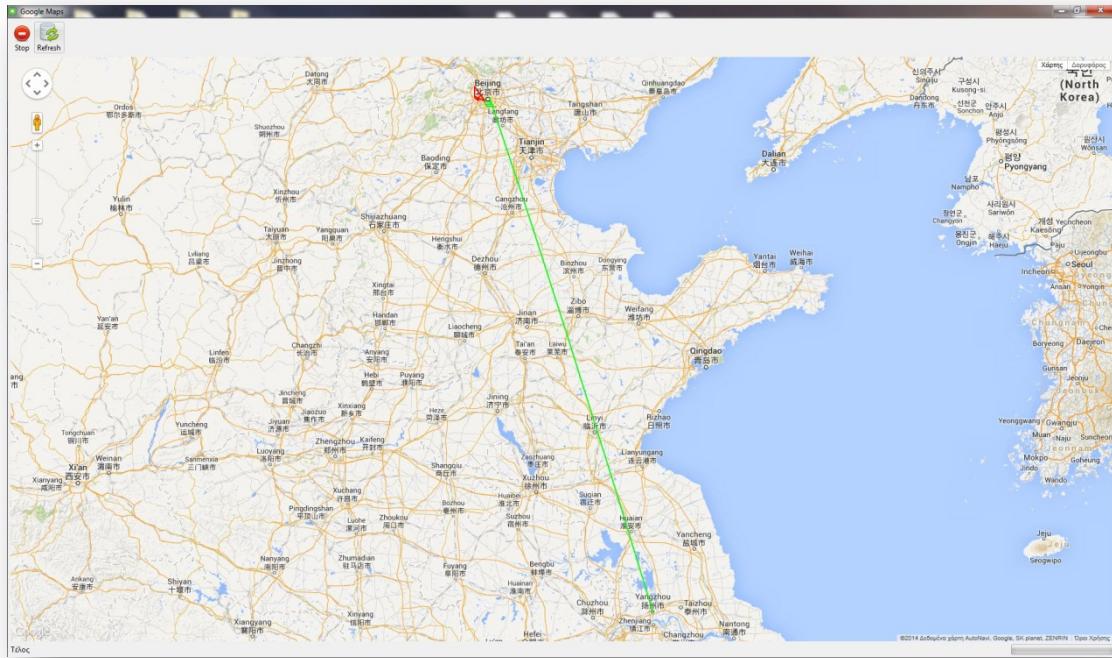
Έτσι τα αποτελέσματα των αναλύσεων διαμορφώνονται ως εξής:

- **Για τα 1000 Points:**

Αρχείο A	Αρχείο B	$\epsilon$	Similarity
366.txt	3557.txt	0.05	0.1 %
366.txt	5099.txt	0.05	3.2 %
366.txt	6665.txt	0.05	29.9 %
3557.txt	5099.txt	0.05	14.89 %
3557.txt	6665.txt	0.05	0.2 %
5099.txt	6665.txt	0.05	17.5 %

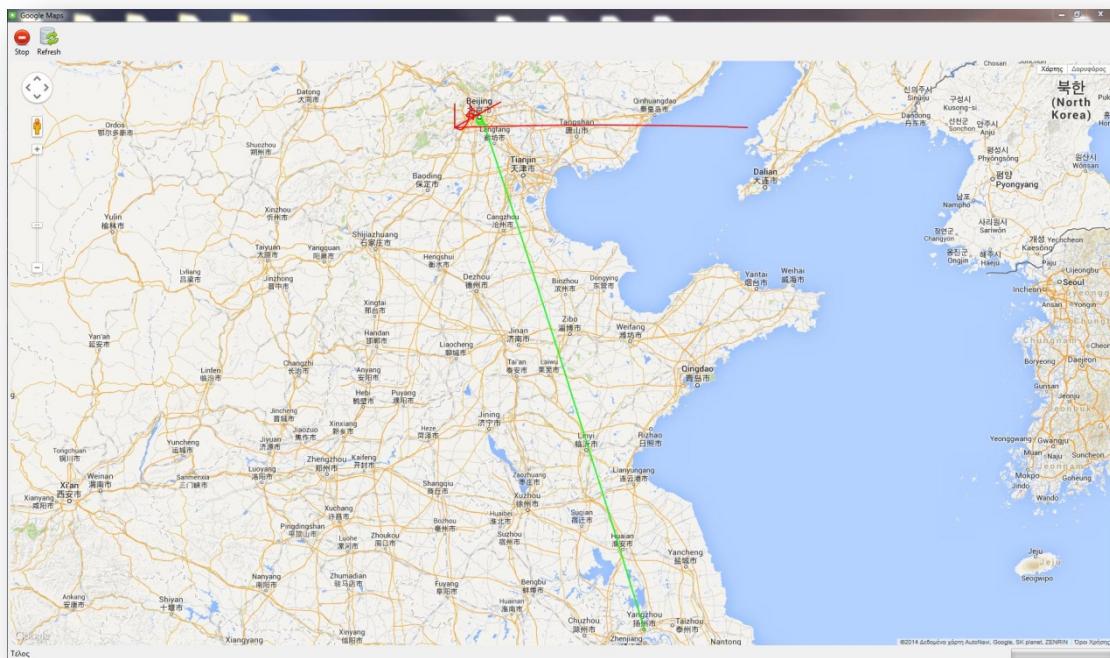
## 366.txt - 3557.txt

Similarity: 0.1 %



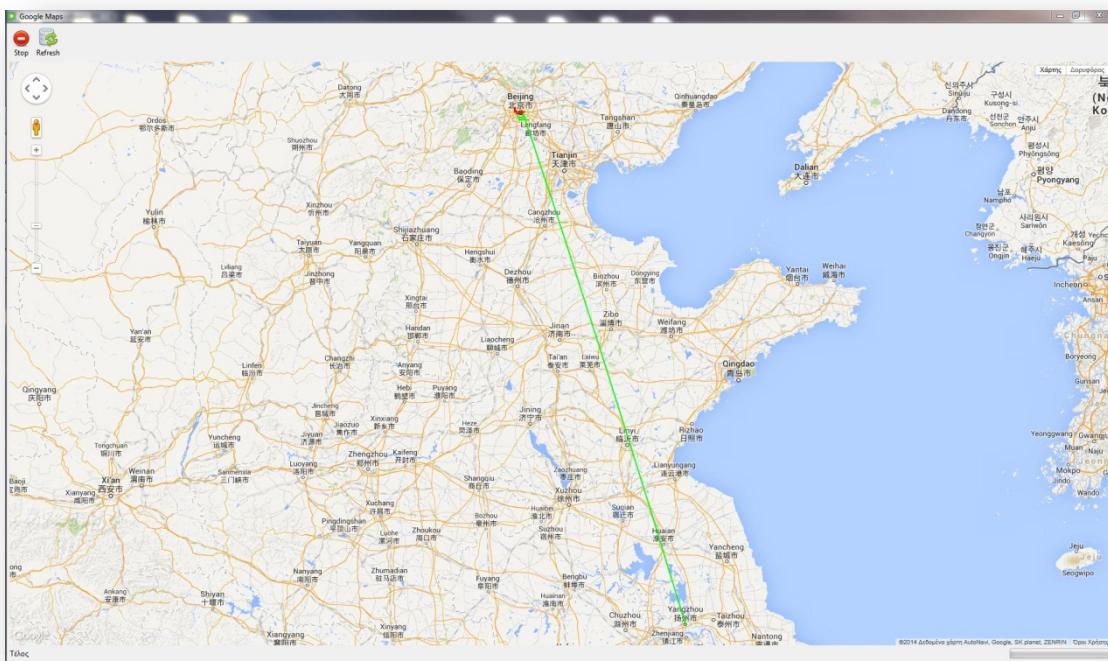
## 366.txt - 5099.txt

Similarity: 3.2 %



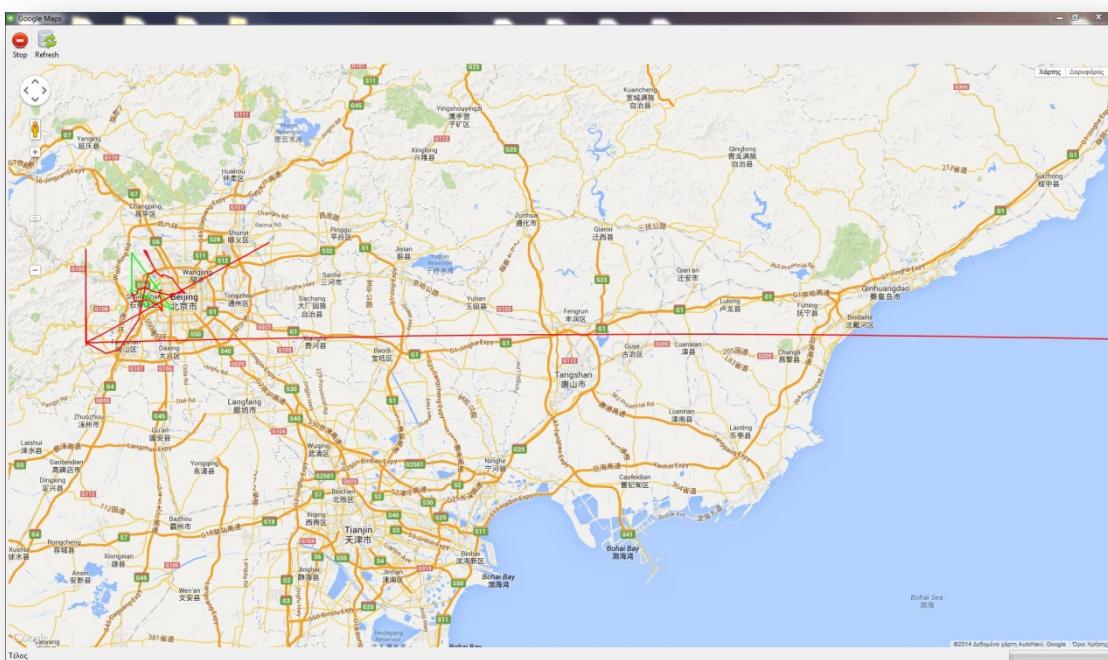
**366.txt - 6665.txt**

Similarity: 29.9 %



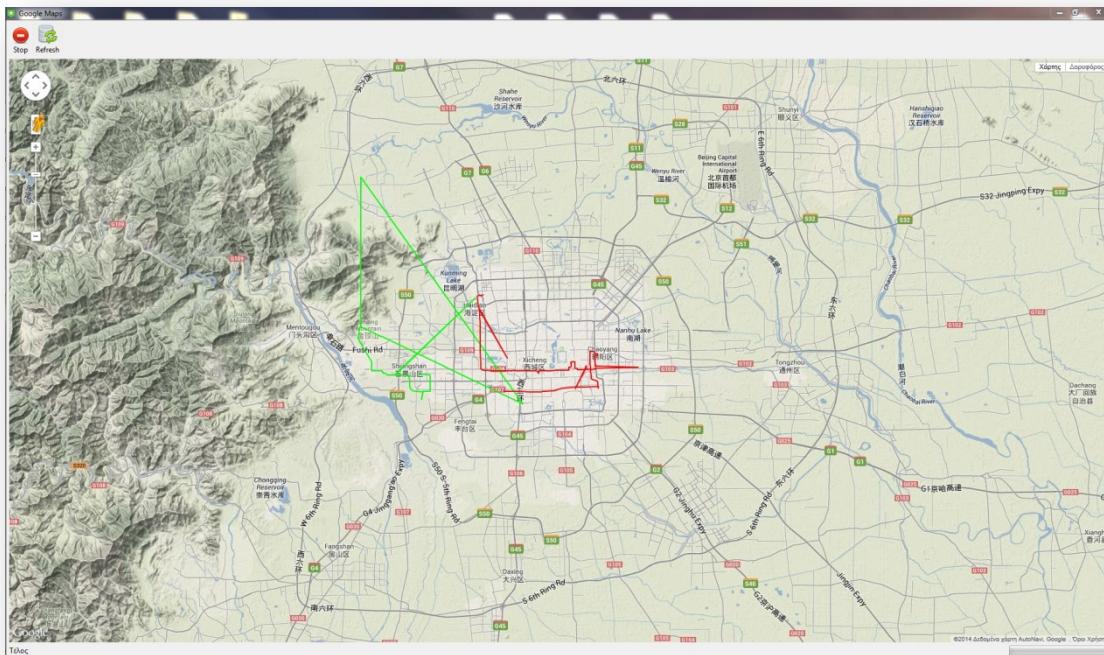
## 3557.txt - 5099.txt

Similarity: 14.89 %



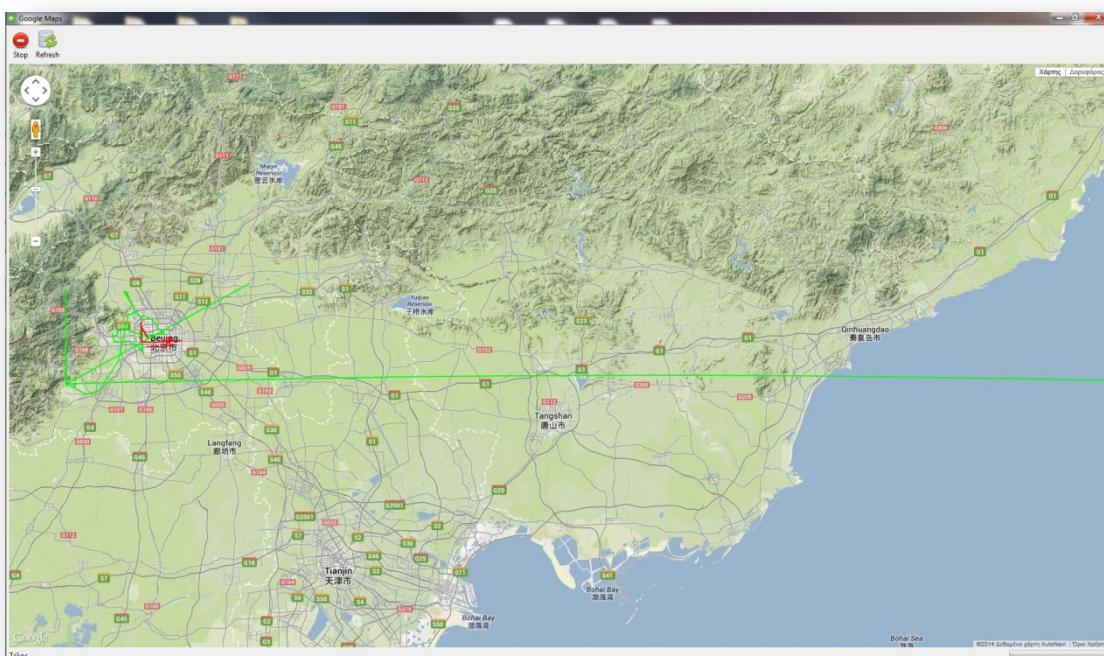
### 3557.txt - 6665.txt

Similarity: 0.2 %



### 5099.txt - 6665.txt

Similarity: 17.5 %

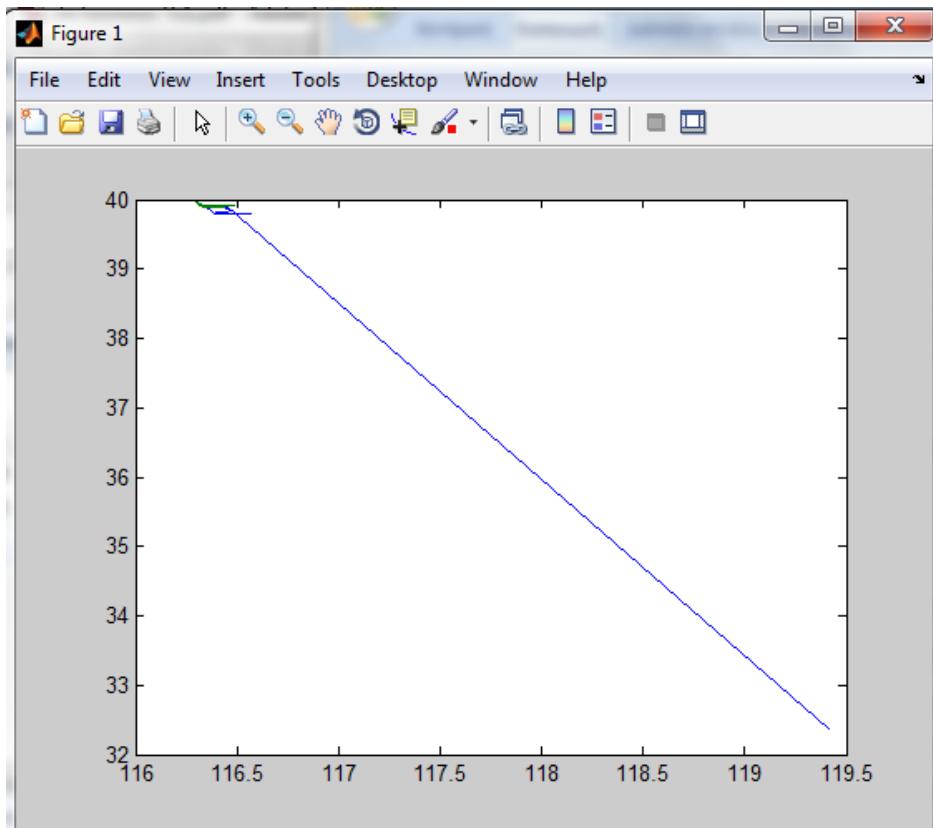


Παρατηρώντας τα παραπάνω αποτελέσματα προκύπτουν τα εξής:

- ❖ Το αρχείο 366.txt από το φάκελο 1000Points παρουσιάζει μεγαλύτερη ομοιότητα με το αρχείο 6665.txt από τον ίδιο φάκελο, με ποσοστό ομοιότητας 29.9%. Ο χρόνος που απαιτήθηκε ήταν 0.0153 sec. Η εμφάνιση σε χάρτη της τροχιάς σύγκρισης που περιγράφεται αρχείο 366.txt, σε συνδυασμό με την ομοιότερη τροχιά που βρέθηκε μέσω του LCSS ( 6665.txt ) έχει παρουσιαστεί παραπάνω. Η τροχιά σύγκρισης ( 366.txt ), σε συνδυασμό με την ομοιότερη τροχιά ( 6665.txt ) πάνω στους άξονες x, y παρουσιάζεται παρακάτω.

### 366.txt - 6665.txt

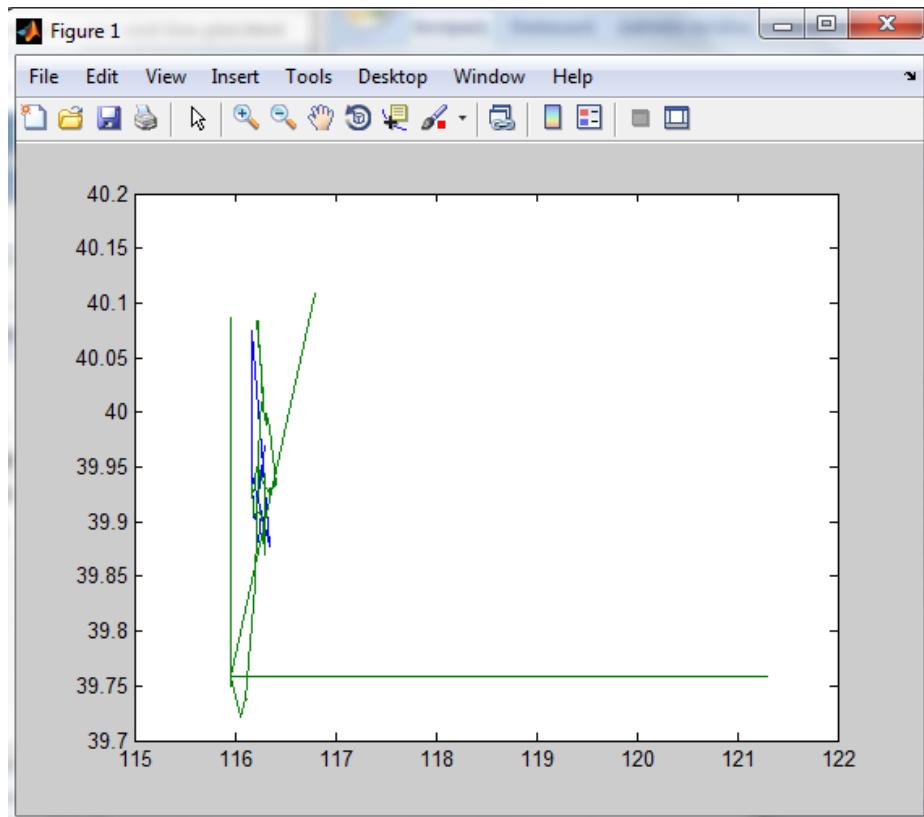
Similarity: 29.9 %



- ❖ Το αρχείο 3557.txt από το φάκελο 1000Points παρουσιάζει μεγαλύτερη ομοιότητα με το αρχείο 5099.txt από τον ίδιο φάκελο, με ποσοστό ομοιότητας 14.89%. Ο χρόνος που απαιτήθηκε ήταν 0.0086 sec. Η εμφάνιση σε χάρτη της τροχιάς σύγκρισης που περιγράφεται αρχείο 3557.txt, σε συνδυασμό με την ομοιότερη τροχιά που βρέθηκε μέσω του LCSS ( 5099.txt ) έχει παρουσιαστεί παραπάνω. Η τροχιά σύγκρισης ( 3557.txt ), σε συνδυασμό με την ομοιότερη τροχιά ( 5099.txt ) πάνω στους άξονες x, y παρουσιάζεται παρακάτω.

### 3557.txt - 5099.txt

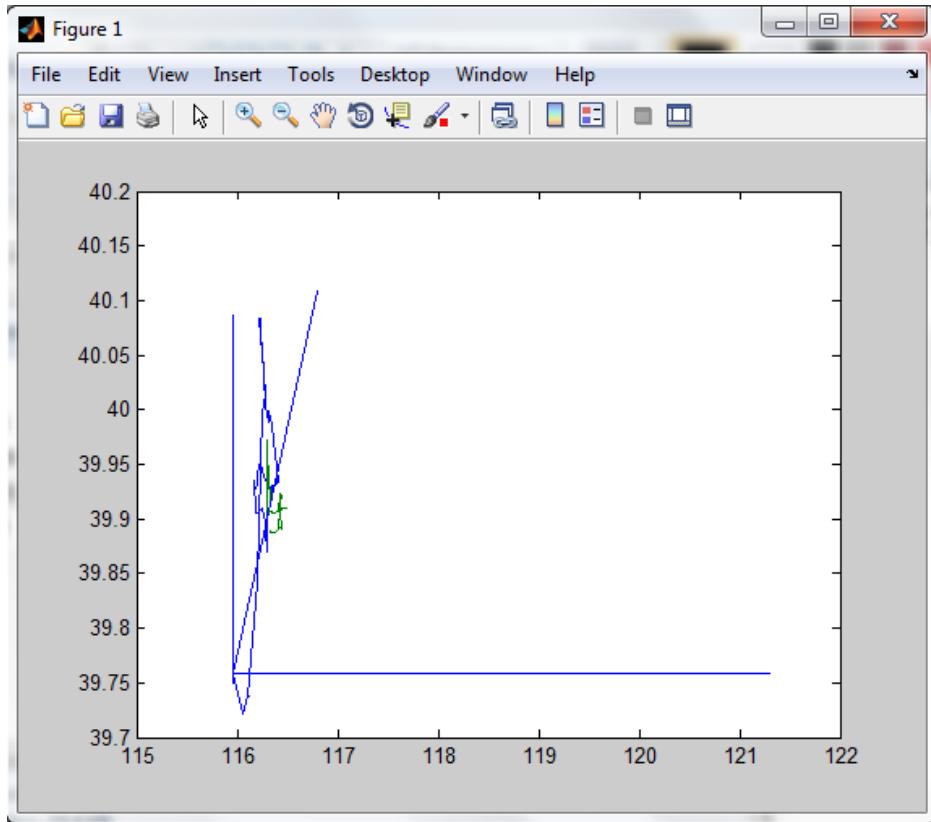
Similarity: 14.89 %



- ❖ Το αρχείο 5099.txt από το φάκελο 1000Points παρουσιάζει μεγαλύτερη ομοιότητα με το αρχείο 6665.txt από τον ίδιο φάκελο, με ποσοστό ομοιότητας 17.5%. Ο χρόνος που απαιτήθηκε ήταν 0.0133 sec. Η εμφάνιση σε χάρτη της τροχιάς σύγκρισης που περιγράφεται αρχείο 5099.txt, σε συνδυασμό με την ομοιότερη τροχιά που βρέθηκε μέσω του LCSS ( 6665.txt ) έχει παρουσιαστεί παραπάνω. Η τροχιά σύγκρισης ( 5099.txt ), σε συνδυασμό με την ομοιότερη τροχιά ( 6665.txt ) πάνω στους άξονες x, y παρουσιάζεται παρακάτω.

## 5099.txt - 6665.txt

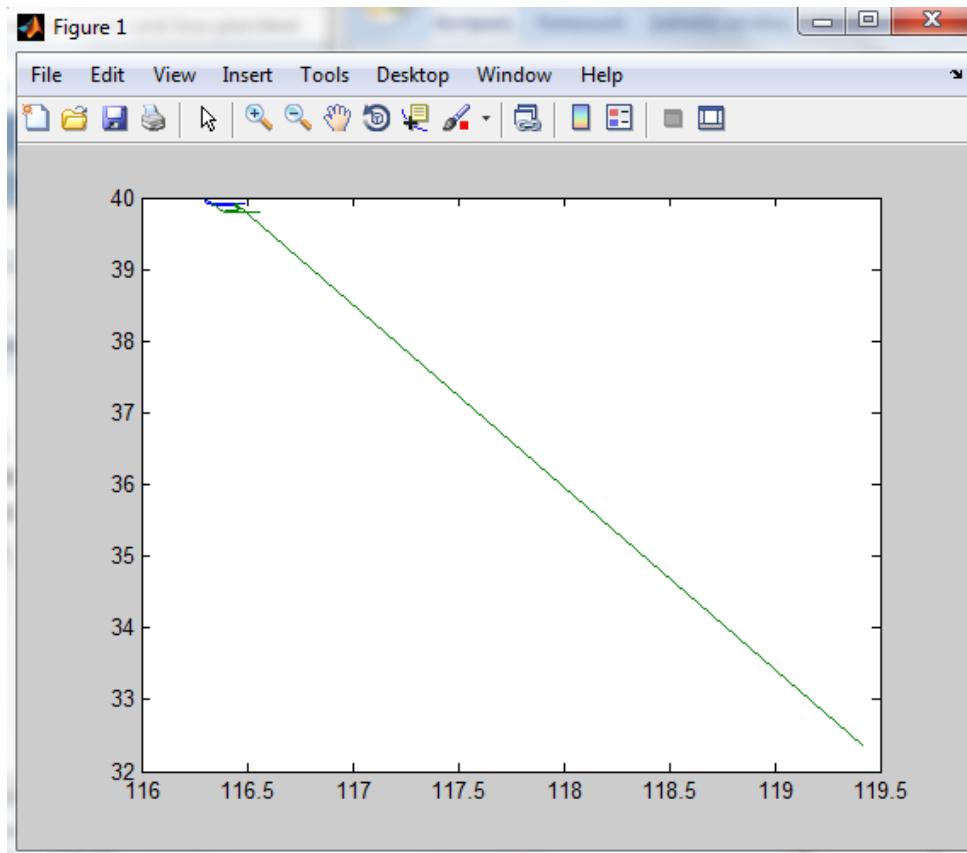
Similarity: 17.5 %



- ❖ Το αρχείο 6665.txt από το φάκελο 1000Points παρουσιάζει μεγαλύτερη ομοιότητα με το αρχείο 366.txt από τον ίδιο φάκελο, με ποσοστό ομοιότητας 29.9%. Ο χρόνος που απαιτήθηκε ήταν 0.0153 sec. Η εμφάνιση σε χάρτη της τροχιάς σύγκρισης που περιγράφεται αρχείο 6665.txt, σε συνδυασμό με την ομοιότερη τροχιά που βρέθηκε μέσω του LCSS ( 366.txt ) έχει παρουσιαστεί παραπάνω. Η τροχιά σύγκρισης ( 6665.txt ), σε συνδυασμό με την ομοιότερη τροχιά ( 366.txt ) πάνω στους άξονες x, γ παρουσιάζεται παρακάτω.

### 6665.txt - 366.txt

Similarity: 29.9 %



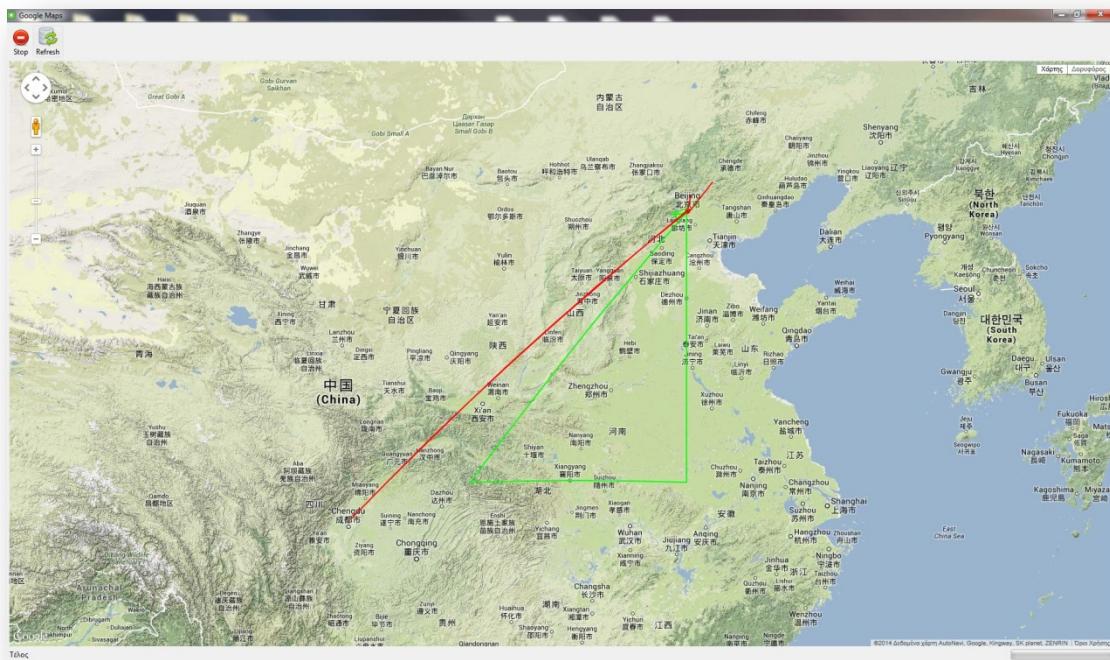
Προκύπτει επίσης, ότι ο χρόνος που απαιτείται για την εκτέλεση του αλγορίθμου με δεδομένα αρχεία των 1000 Σημείων είναι πολύ μικρός. Αυτό είναι προφανές, αφού γνωρίζουμε πως η χρονική πολυπλοκότητα του αλγορίθμου LCSS είναι  $O(n*m)$ , όπου η είναι το μήκος της 1<sup>ης</sup> τροχιάς σύγκρισης (εδώ 1000) και  $m$ , το μήκος της 2<sup>ης</sup> τροχιάς σύγκρισης (εδώ 1000).

- **Για τα 5000 Points:**

Αρχείο A	Αρχείο B	ε	Similarity
3579.txt	6275.txt	0.05	20.54 %
3579.txt	7146.txt	0.05	4.76 %
3579.txt	9109.txt	0.05	33.04 %
6275.txt	7146.txt	0.05	29.78 %
6275.txt	9109.txt	0.05	18.96 %
7146.txt	9109.txt	0.05	28.66 %

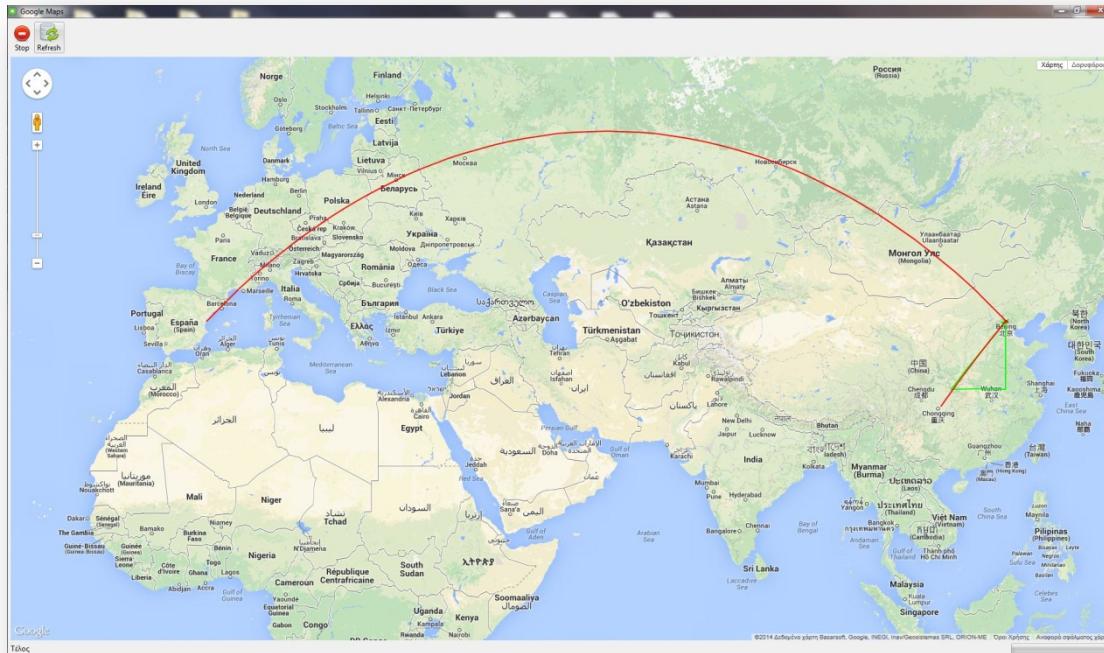
### 3579.txt - 6275.txt

Similarity: 20.54 %



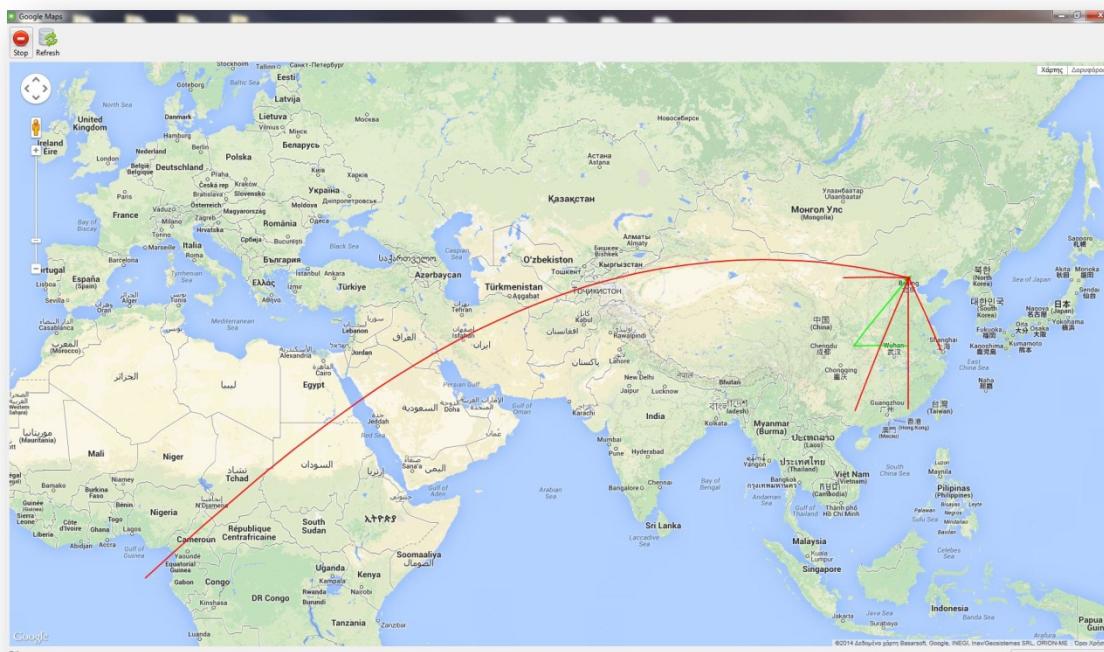
## 3579.txt - 7146.txt

Similarity: 4.76 %



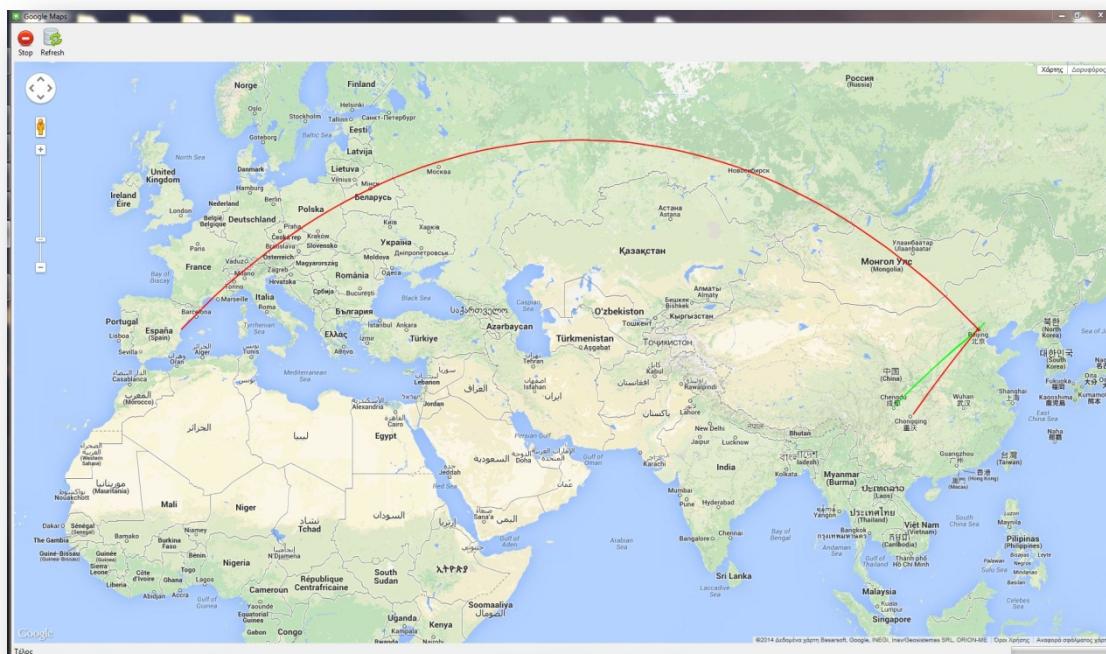
## 3579.txt - 9109.txt

Similarity: 33.04 %



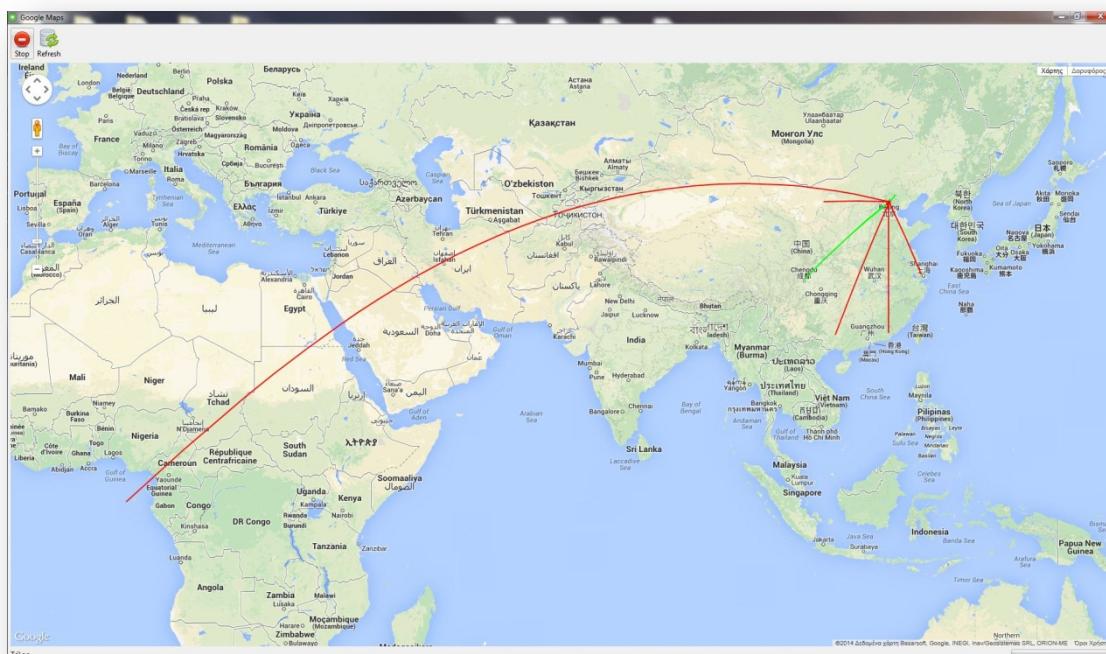
## 6275.txt - 7146.txt

Similarity: 29.78 %



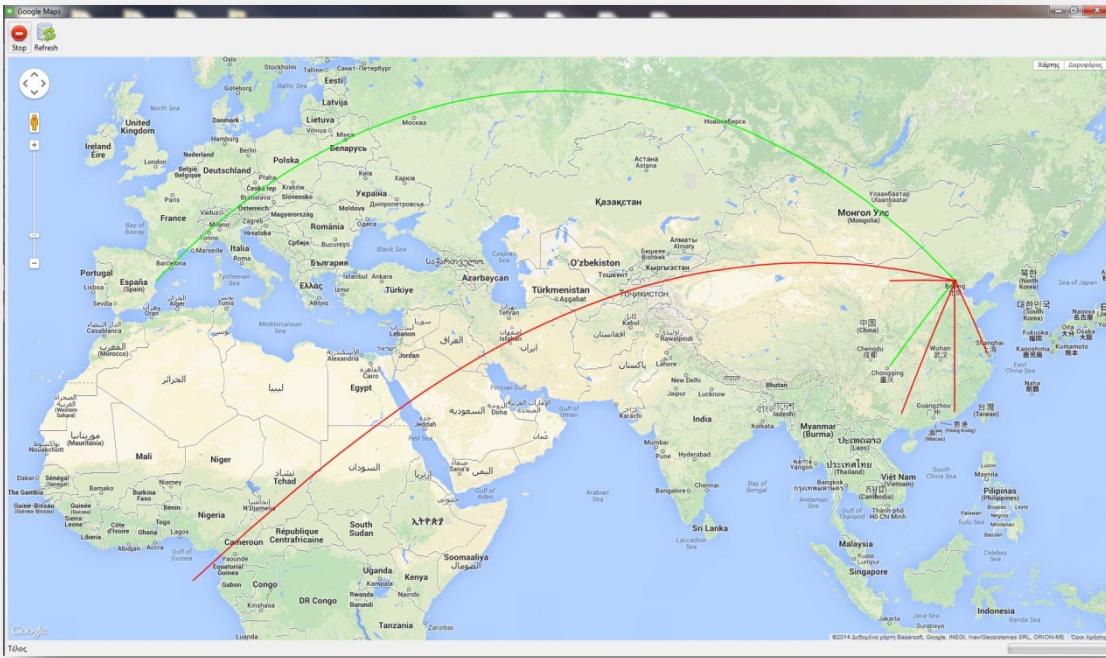
## 6275.txt - 9109.txt

Similarity: 18.96 %



## 7146.txt - 9109.txt

Similarity: 28.66 %

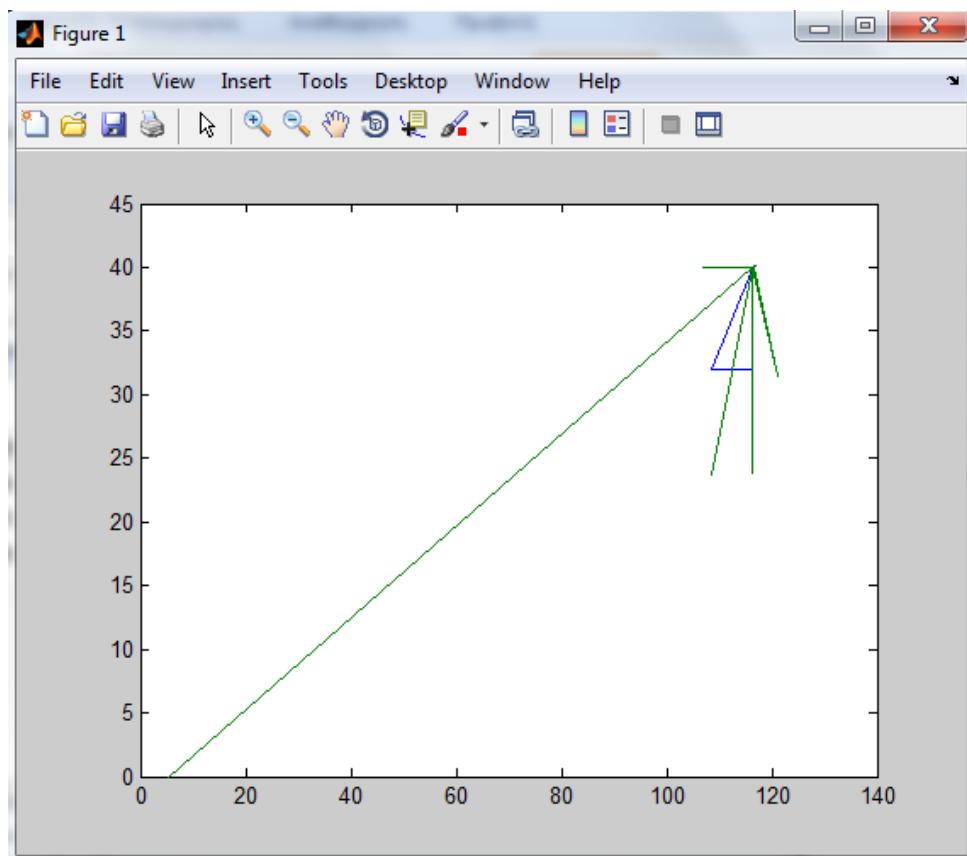


Παρατηρώντας τα παραπάνω αποτελέσματα προκύπτουν τα εξής:

- ❖ Το αρχείο 3579.txt από το φάκελο 5000Points παρουσιάζει μεγαλύτερη ομοιότητα με το αρχείο 9109.txt από τον ίδιο φάκελο, με ποσοστό ομοιότητας 33.04%. Ο χρόνος που απαιτήθηκε ήταν 0.1679 sec. Η εμφάνιση σε χάρτη της τροχιάς σύγκρισης που περιγράφεται αρχείο 3579.txt, σε συνδυασμό με την ομοιότερη τροχιά που βρέθηκε μέσω του LCSS ( 9109.txt ) έχει παρουσιαστεί παραπάνω. Η τροχιά σύγκρισης ( 3579.txt ), σε συνδυασμό με την ομοιότερη τροχιά ( 9109.txt ) πάνω στους άξονες x, y παρουσιάζεται παρακάτω.

### 3579.txt - 9109.txt

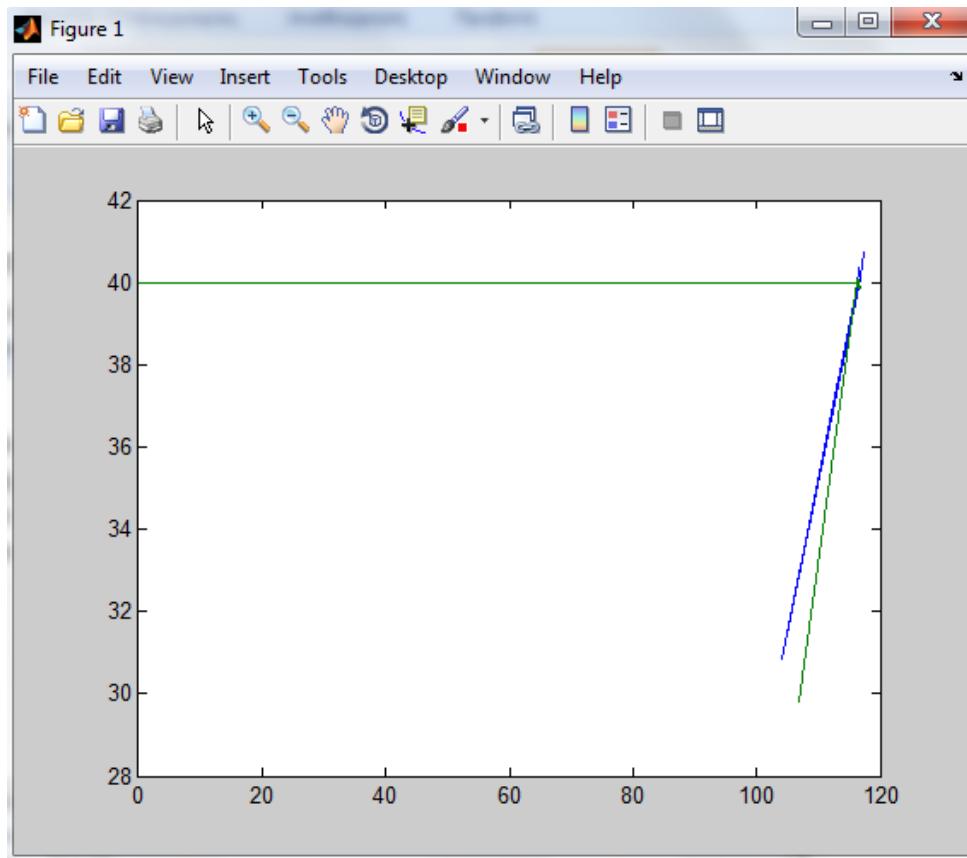
Similarity: 33.04 %



- ❖ Το αρχείο 6275.txt από το φάκελο 5000Points παρουσιάζει μεγαλύτερη ομοιότητα με το αρχείο 7146.txt από τον ίδιο φάκελο, με ποσοστό ομοιότητας 29.78%. Ο χρόνος που απαιτήθηκε ήταν 0.1653 sec. Η εμφάνιση σε χάρτη της τροχιάς σύγκρισης που περιγράφεται αρχείο 6275.txt, σε συνδυασμό με την ομοιότερη τροχιά που βρέθηκε μέσω του LCSS ( 7146.txt ) έχει παρουσιαστεί παραπάνω. Η τροχιά σύγκρισης ( 6275.txt ), σε συνδυασμό με την ομοιότερη τροχιά ( 7146.txt ) πάνω στους άξονες x, y παρουσιάζεται παρακάτω.

### 6275.txt - 7146.txt

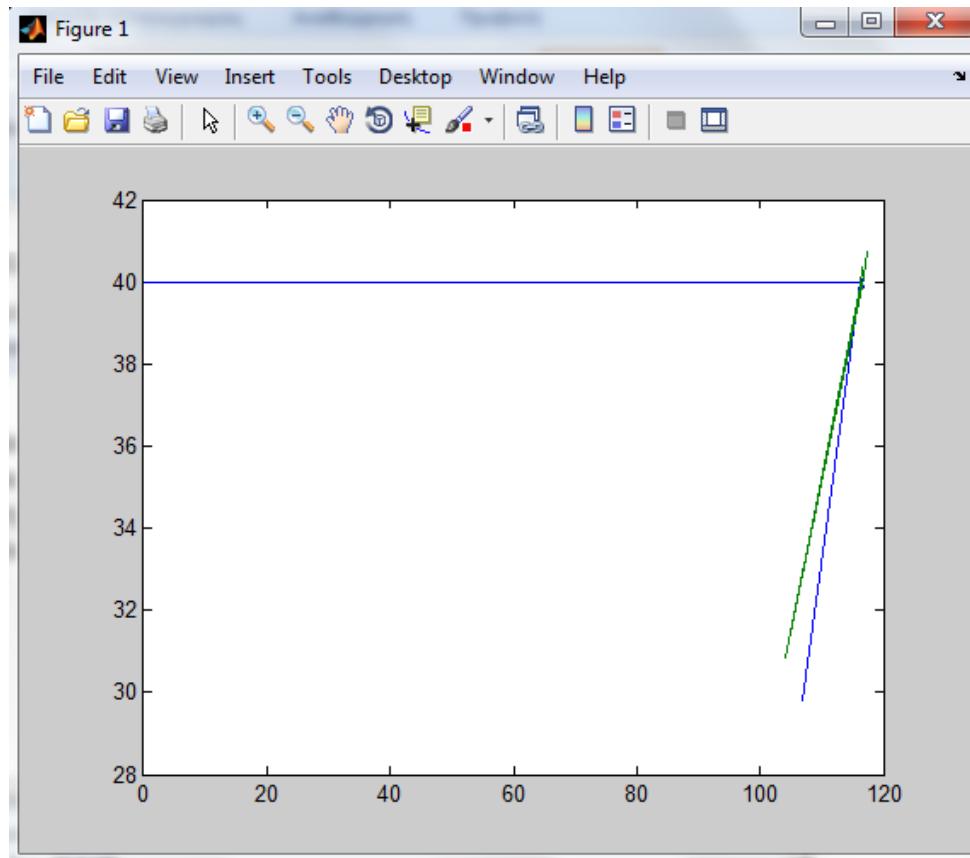
Similarity: 29.78 %



- ❖ Το αρχείο 7146.txt από το φάκελο 5000Points παρουσιάζει μεγαλύτερη ομοιότητα με το αρχείο 6275.txt από τον ίδιο φάκελο, με ποσοστό ομοιότητας 29.78%. Ο χρόνος που απαιτήθηκε ήταν 0.1653 sec. Η εμφάνιση σε χάρτη της τροχιάς σύγκρισης που περιγράφεται αρχείο 7146.txt, σε συνδυασμό με την ομοιότερη τροχιά που βρέθηκε μέσω του LCSS ( 6275.txt ) έχει παρουσιαστεί παραπάνω. Η τροχιά σύγκρισης ( 7146.txt ), σε συνδυασμό με την ομοιότερη τροχιά ( 6275.txt ) πάνω στους άξονες x, y παρουσιάζεται παρακάτω.

### 7146.txt - 6275.txt

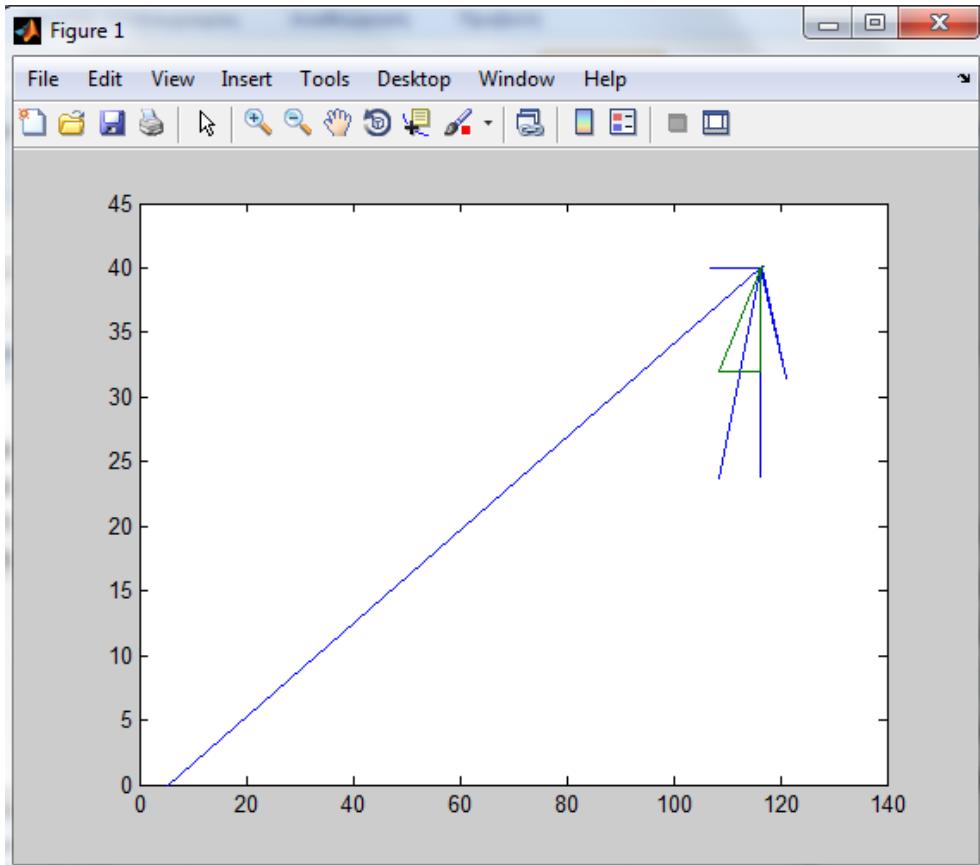
Similarity: 29.78 %



- ❖ Το αρχείο 9109.txt από το φάκελο 5000Points παρουσιάζει μεγαλύτερη ομοιότητα με το αρχείο 3579.txt από τον ίδιο φάκελο, με ποσοστό ομοιότητας 33.04%. Ο χρόνος που απαιτήθηκε ήταν 0.1718 sec. Η εμφάνιση σε χάρτη της τροχιάς σύγκρισης που περιγράφεται αρχείο 9109.txt, σε συνδυασμό με την ομοιότερη τροχιά που βρέθηκε μέσω του LCSS ( 3579.txt ) έχει παρουσιαστεί παραπάνω. Η τροχιά σύγκρισης ( 9109.txt ), σε συνδυασμό με την ομοιότερη τροχιά ( 3579.txt ) πάνω στους άξονες x, y παρουσιάζεται παρακάτω.

### 9109.txt - 3579.txt

Similarity: 33.04 %



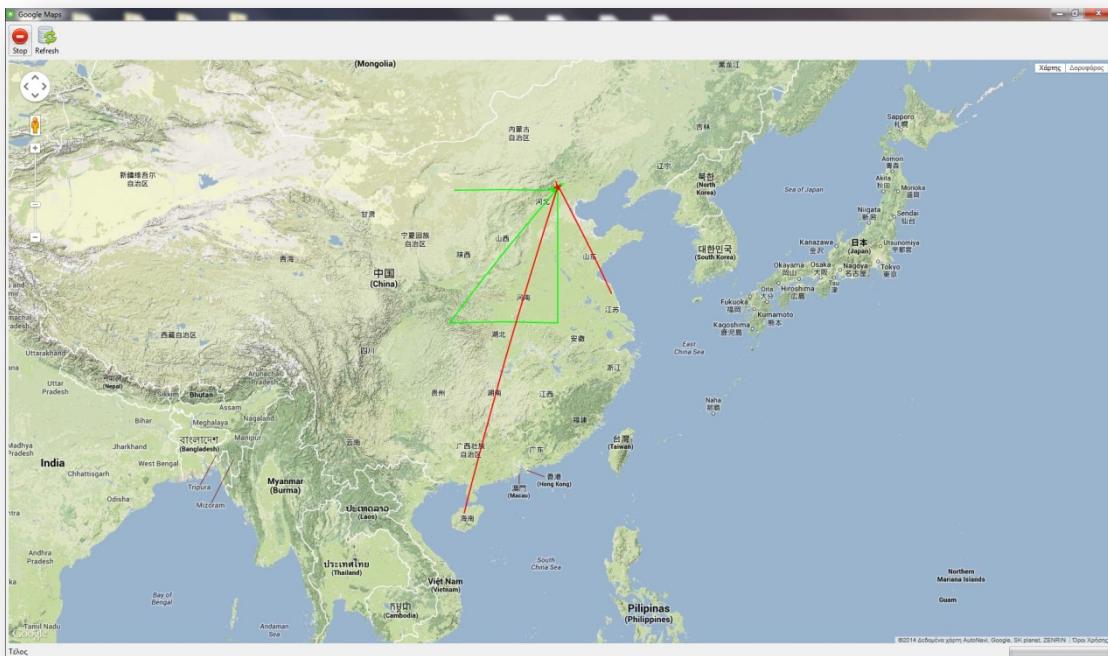
Προκύπτει επίσης, ότι ο χρόνος που απαιτείται για την εκτέλεση του αλγορίθμου με δεδομένα αρχεία των 5000 Σημείων είναι μεγαλύτερος από το χρόνο για την εκτέλεση του αλγορίθμου κατά τη σύγκριση αρχείων με 1000 Σημεία. Αυτό είναι προφανές, αφού γνωρίζουμε πως η χρονική πολυπλοκότητα του αλγορίθμου LCSS είναι  $O(n^m)$ , όπου η είναι το μήκος της 1<sup>ης</sup> τροχιάς σύγκρισης (εδώ 5000) και  $m$ , το μήκος της 2<sup>ης</sup> τροχιάς σύγκρισης (εδώ 5000).

- **Για τα 10000 Points:**

Αρχείο A	Αρχείο B	$\epsilon$	Similarity
3579.txt	3781.txt	0.05	8.87 %
3579.txt	4798.txt	0.05	13.489 %
3579.txt	5075.txt	0.05	8.24 %
3781.txt	4798.txt	0.05	67.22 %
3781.txt	5075.txt	0.05	50.83 %
4798.txt	5075.txt	0.05	58.91 %

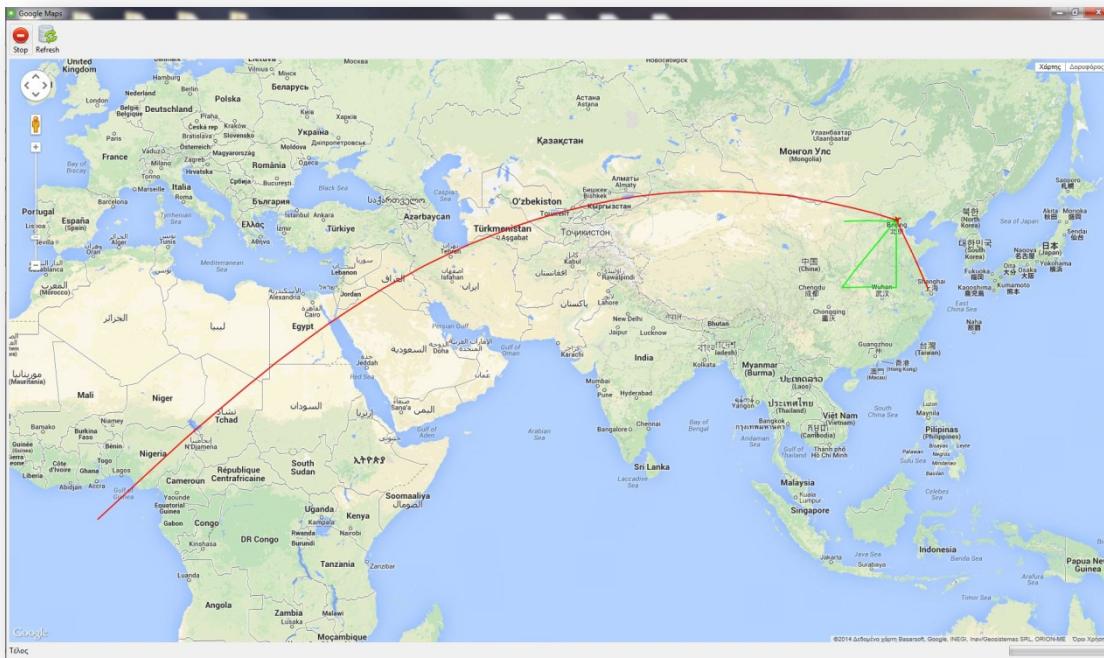
### 3579.txt - 3781.txt

Similarity: 8.87 %



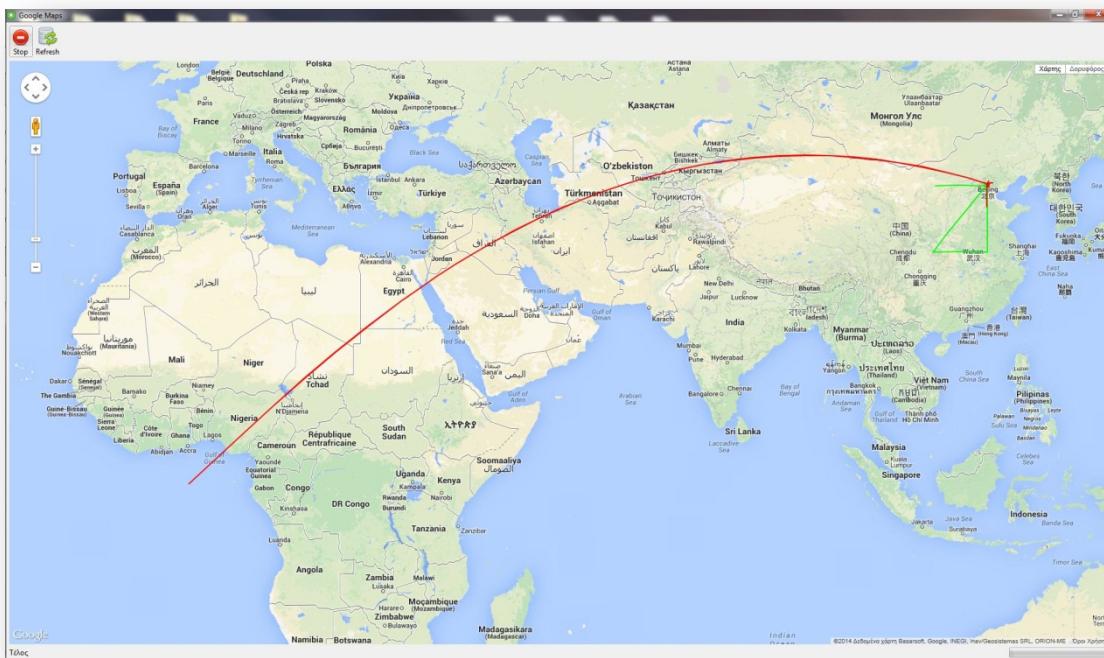
### 3579.txt - 4798.txt

Similarity: 13.489 %



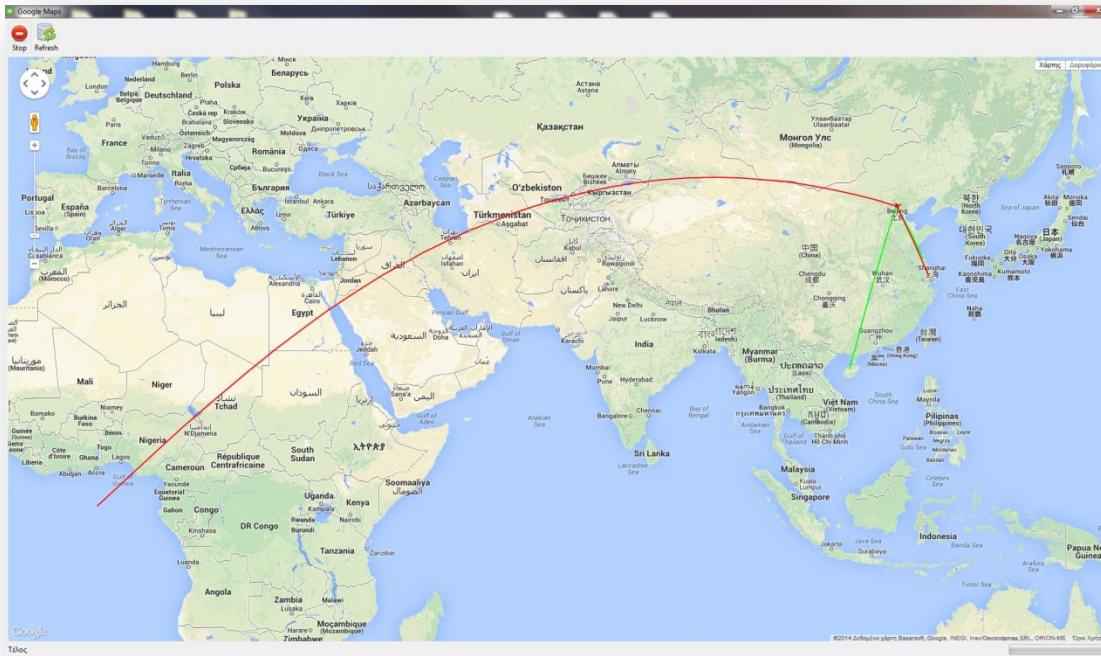
### 3579.txt - 5075.txt

Similarity: 8.24 %



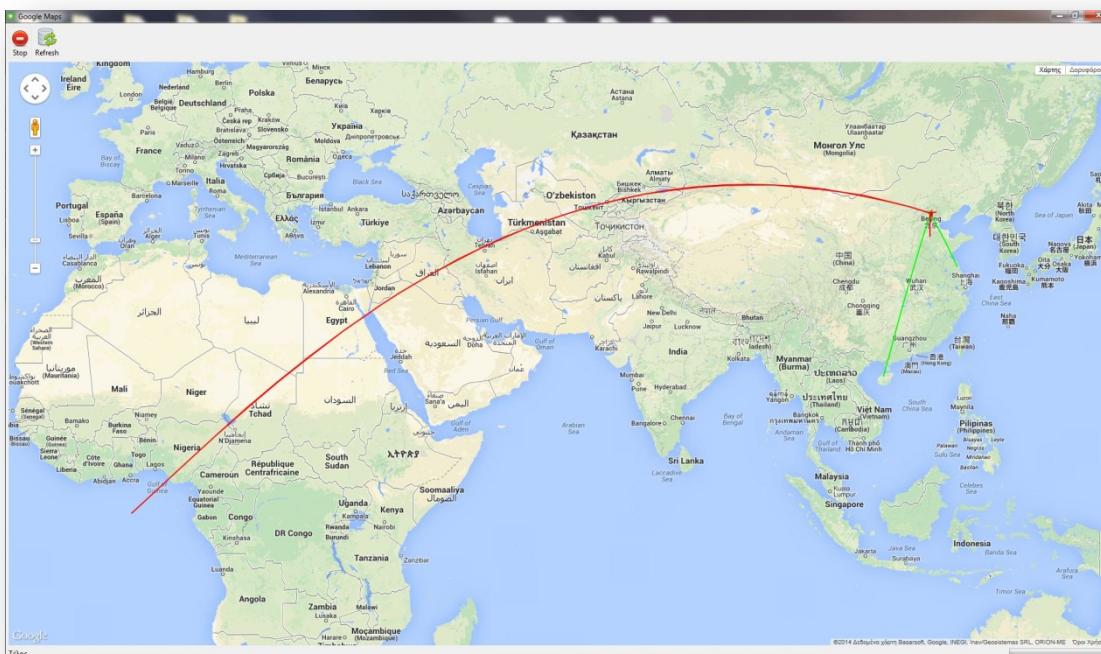
## 3781.txt - 4798.txt

Similarity: 67.22 %



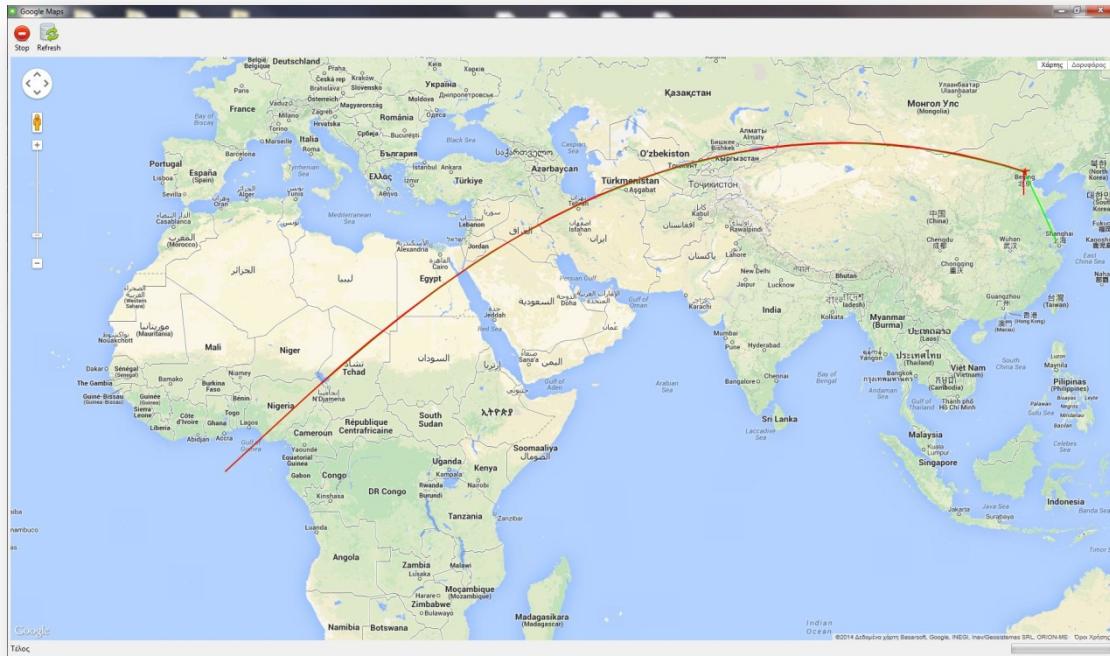
## 3781.txt - 5057.txt

Similarity: 50.83 %



## 4798.txt - 5075.txt

Similarity: 58.91 %

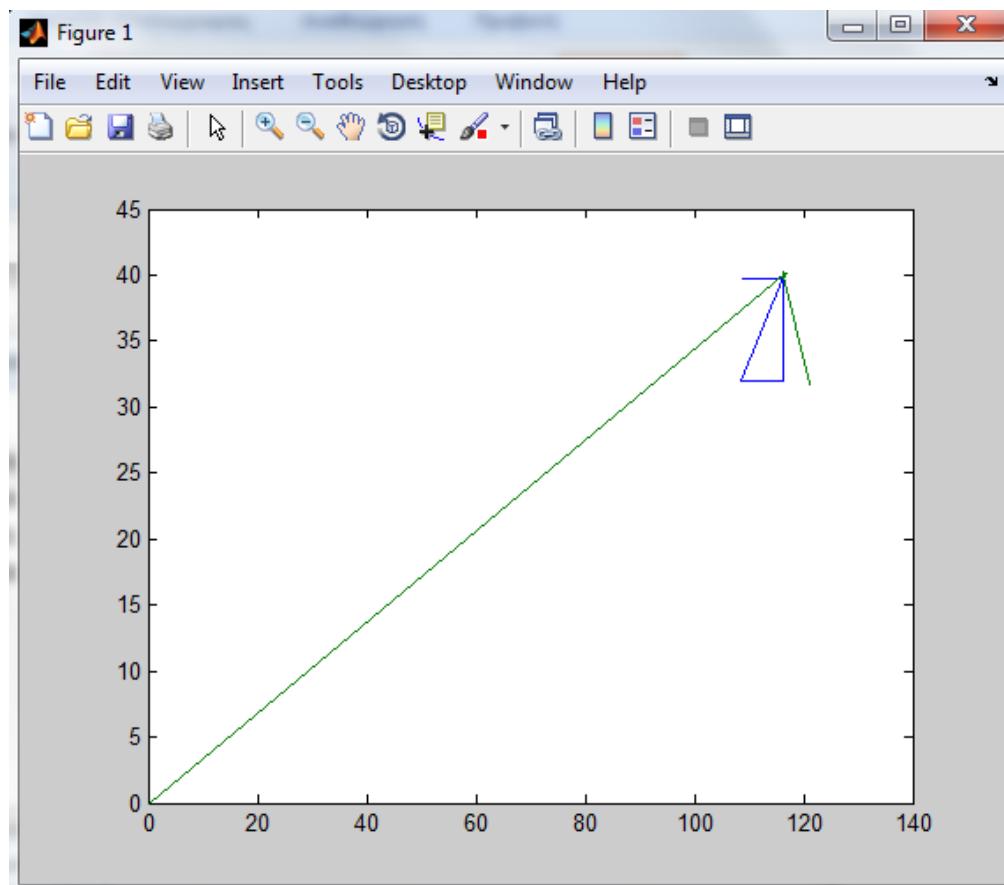


Παρατηρώντας τα παραπάνω αποτελέσματα προκύπτουν τα εξής:

- ❖ Το αρχείο 3579.txt από το φάκελο 10000Points παρουσιάζει μεγαλύτερη ομοιότητα με το αρχείο 4798.txt από τον ίδιο φάκελο, με ποσοστό ομοιότητας 13.489%. Ο χρόνος που απαιτήθηκε ήταν 0.6967 sec. Η εμφάνιση σε χάρτη της τροχιάς σύγκρισης που περιγράφεται αρχείο 3579.txt, σε συνδυασμό με την ομοιότερη τροχιά που βρέθηκε μέσω του LCSS ( 4798.txt ) έχει παρουσιαστεί παραπάνω. Η τροχιά σύγκρισης ( 3579.txt ), σε συνδυασμό με την ομοιότερη τροχιά ( 4798.txt ) πάνω στους άξονες x, γ παρουσιάζεται παρακάτω.

### 3579.txt - 4798.txt

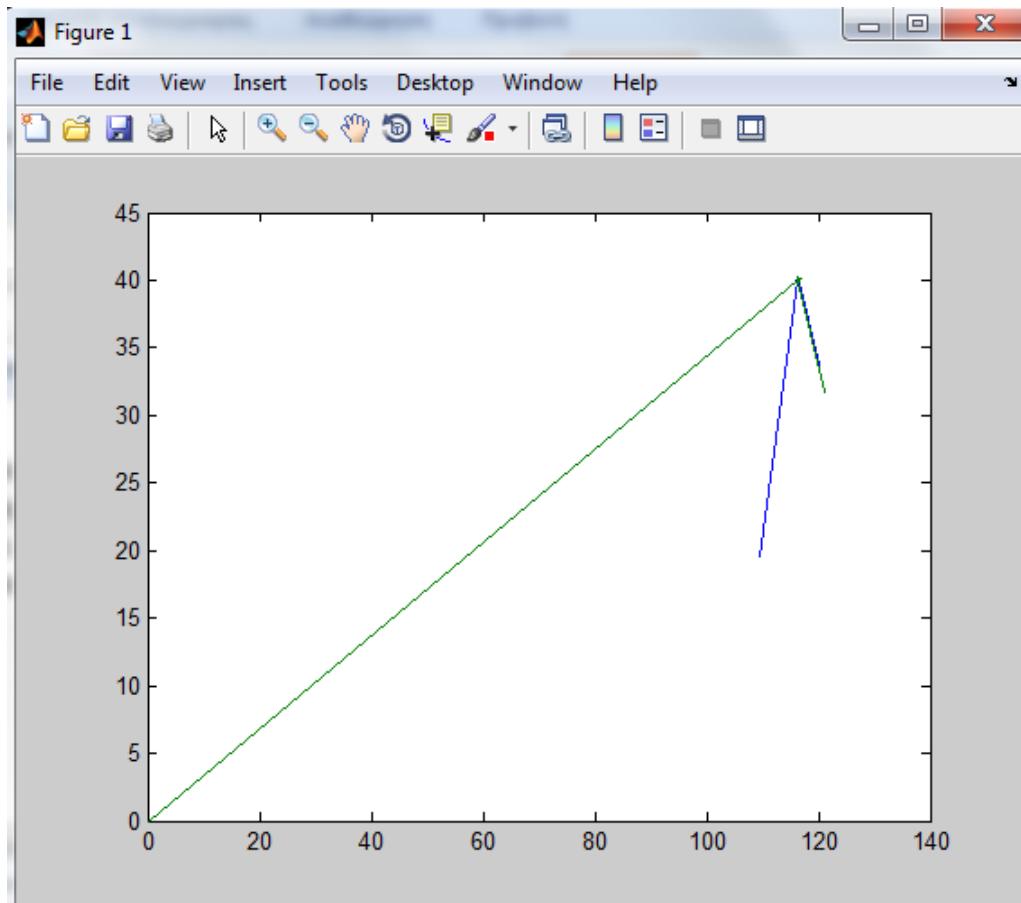
Similarity: 13.489 %



- ❖ Το αρχείο 3781.txt από το φάκελο 10000Points παρουσιάζει μεγαλύτερη ομοιότητα με το αρχείο 4798.txt από τον ίδιο φάκελο, με ποσοστό ομοιότητας 67.22%. Ο χρόνος που απαιτήθηκε ήταν 0.6417 sec. Η εμφάνιση σε χάρτη της τροχιάς σύγκρισης που περιγράφεται αρχείο 3781.txt, σε συνδυασμό με την ομοιότερη τροχιά που βρέθηκε μέσω του LCSS ( 4798.txt ) έχει παρουσιαστεί παραπάνω. Η τροχιά σύγκρισης ( 3781.txt ), σε συνδυασμό με την ομοιότερη τροχιά ( 4798.txt ) πάνω στους άξονες x, y παρουσιάζεται παρακάτω.

### 3781.txt - 4798.txt

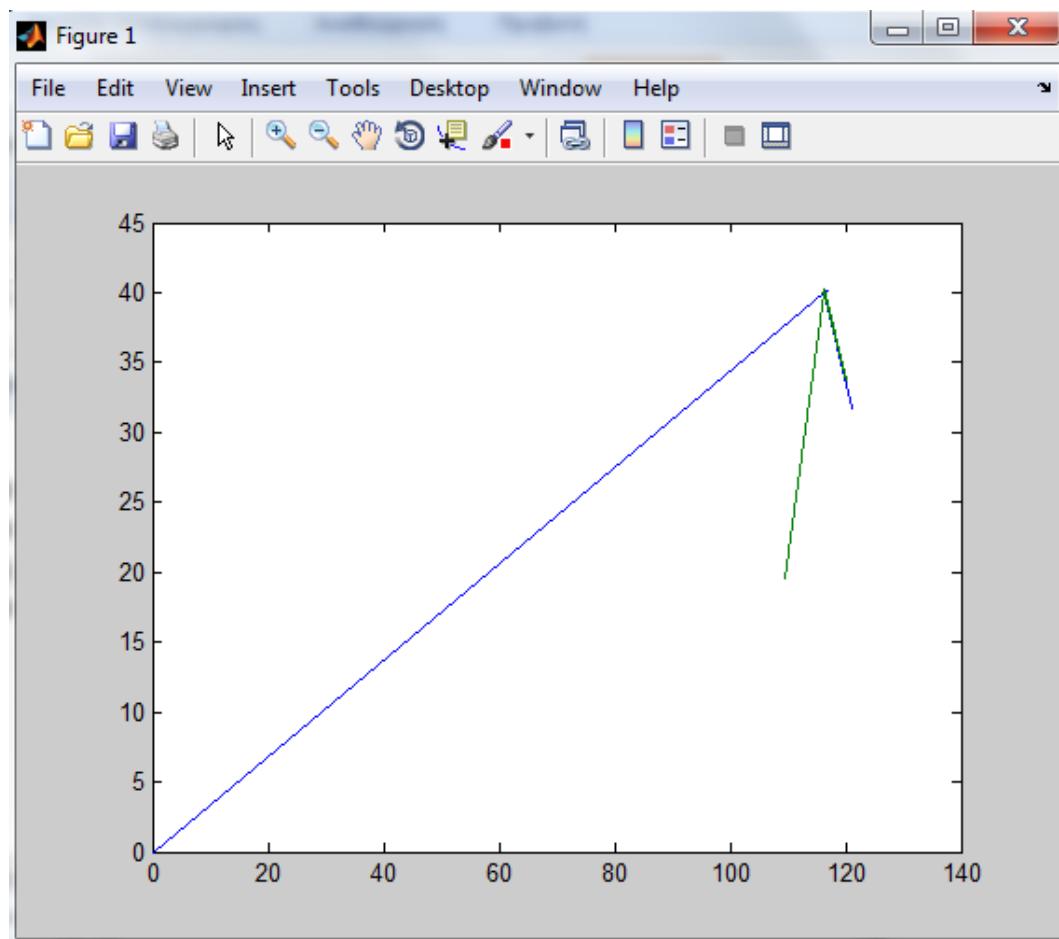
Similarity: 67.22 %



- ❖ Το αρχείο 4798.txt από το φάκελο 10000Points παρουσιάζει μεγαλύτερη ομοιότητα με το αρχείο 3781.txt από τον ίδιο φάκελο, με ποσοστό ομοιότητας 67.22%. Ο χρόνος που απαιτήθηκε ήταν 0.6417 sec. Η εμφάνιση σε χάρτη της τροχιάς σύγκρισης που περιγράφεται αρχείο 4798.txt, σε συνδυασμό με την ομοιότερη τροχιά που βρέθηκε μέσω του LCSS ( 3781.txt ) έχει παρουσιαστεί παραπάνω. Η τροχιά σύγκρισης ( 4798.txt ), σε συνδυασμό με την ομοιότερη τροχιά ( 3781.txt ) πάνω στους άξονες x, y παρουσιάζεται παρακάτω.

### 4798.txt - 3781.txt

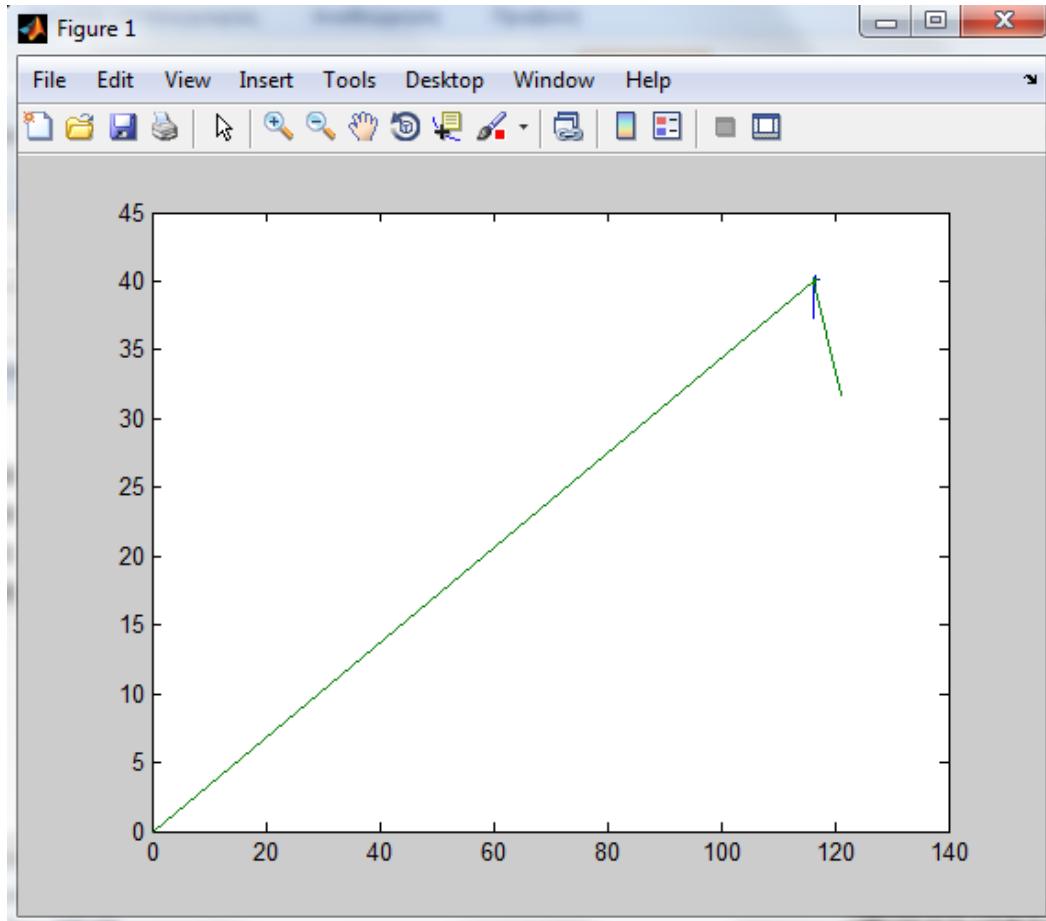
Similarity: 67.22 %



- ❖ Το αρχείο 5075.txt από το φάκελο 10000Points παρουσιάζει μεγαλύτερη ομοιότητα με το αρχείο 4798.txt από τον ίδιο φάκελο, με ποσοστό ομοιότητας 58.91%. Ο χρόνος που απαιτήθηκε ήταν 0.6795 sec. Η εμφάνιση σε χάρτη της τροχιάς σύγκρισης που περιγράφεται αρχείο 5075.txt, σε συνδυασμό με την ομοιότερη τροχιά που βρέθηκε μέσω του LCSS ( 4798.txt ) έχει παρουσιαστεί παραπάνω. Η τροχιά σύγκρισης ( 5075.txt ), σε συνδυασμό με την ομοιότερη τροχιά ( 4798.txt ) πάνω στους άξονες x, y παρουσιάζεται παρακάτω.

## 5075.txt - 4798.txt

Similarity: 58.91 %



Προκύπτει επίσης, ότι ο χρόνος που απαιτείται για την εκτέλεση του αλγορίθμου με δεδομένα αρχεία των 10000 Σημείων είναι μεγαλύτερος από το χρόνο για την εκτέλεση του αλγορίθμου κατά τη σύγκριση αρχείων με 5000 Σημεία (και ακόμα μεγαλύτερος για σύγκριση αρχείων με 1000 Σημεία). Αυτό είναι προφανές, αφού γνωρίζουμε πως η χρονική πολυπλοκότητα του αλγορίθμου LCSS είναι  $O(n*m)$ , όπου  $n$  είναι το μήκος της 1<sup>ης</sup> τροχιάς σύγκρισης (εδώ 10000) και  $m$ , το μήκος της 2<sup>ης</sup> τροχιάς σύγκρισης (εδώ 10000).

Τέλος, παρατηρούμε ότι στον πίνακα με τα 10000 σημεία έχουμε τις περισσότερες ομοιότητας. Αυτό οφείλεται στην πυκνότητα των σημείων που είναι κοινά στις δύο ακολουθίες (τροχιές σύγκρισης) και μπορούμε εύκολα να το διακρίνουμε παρατηρώντας τόσο τους χάρτες όσο και τα ίδια τα δεδομένα των αρχείων.

Η επιλογή του «ε» και του «δ» έγινε μετά από δοκιμές σε συγκεκριμένα δεδομένα. Πιο ειδικά τα αποτελέσματα βρίσκονται παρακάτω:

Aρχείο A	Aρχείο B	$\epsilon$	$\delta$	Similarity	Elapsed Time (sec)
366.txt	3579.txt	0.1	300	72.7 %	126.4056
366.txt	3579.txt	0.1	800	72.7 %	191.3650
366.txt	3579.txt	0.05	300	22.6 %	122.3338
366.txt	3579.txt	0.05	800	22.7 %	159.0269
366.txt	3579.txt	0.005	300	0.1 %	90.1613
366.txt	3579.txt	0.005	800	0.1 %	117.3501

Παρατηρούμε ότι καθώς το δ αυξάνεται, ο χρόνος αυξάνεται, ενώ δεν επηρεάζει το ποσοστό ομοιότητας. Το ε συμπεριφέρεται όπως και πριν. Αυτό είναι αναμενόμενο αφού αυξάνοντας το  $\epsilon$ , αυξάνουμε και την ανεκτικότητα σε σφάλμα, ενώ μειώνοντας το δ, μειώνεται εμφανώς το μέγεθος του πίνακα που δημιουργεί ο LCSS για τις ανάγκες του δυναμικού προγραμματισμού (Ο LCSS παίρνει ως είσοδο τροχιά μήκους 1000 και τροχιά μήκους 1000+δ, άρα όσο το δ μειώνεται, τόσο μικρότερο είναι το μήκος της  $2^{\text{nd}}$  τροχιάς σύγκρισης). Επιλέξαμε το 0.05 ως ένα ικανό σφάλμα για την ανάλυση των δεδομένων και το 500 μία μέση τιμή για το δ.

Έτσι τα αποτελέσματα των αναλύσεων διαμορφώνονται ως εξής:

- **Για το αρχείο 366.txt από το φάκελο 1000 Points:**

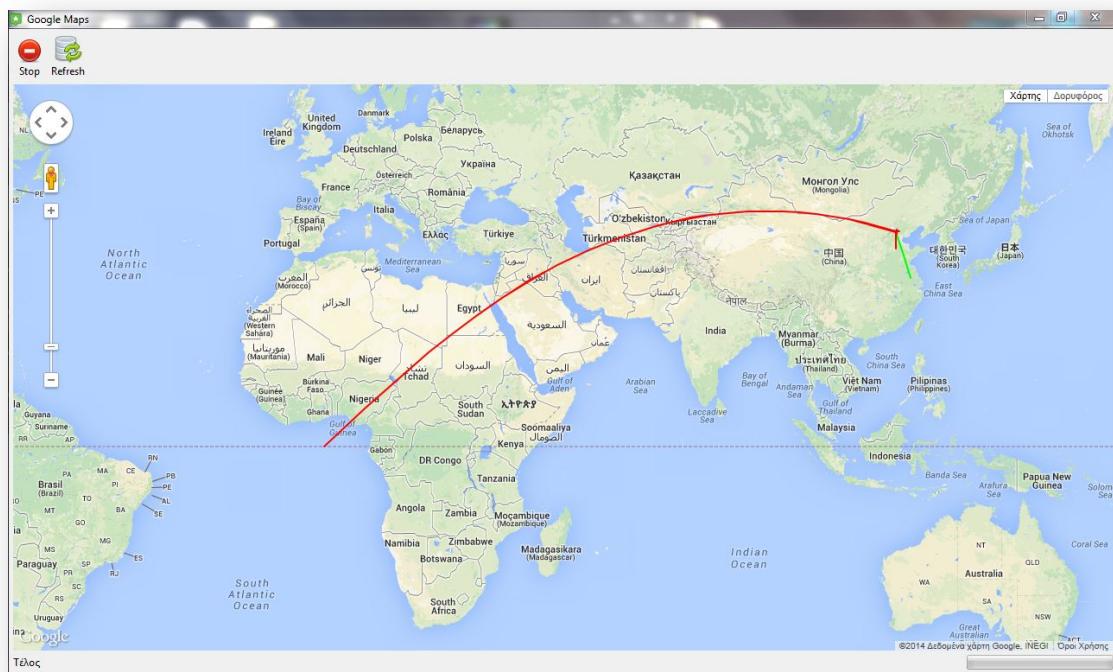
Aρχείο A	Aρχείο B	$\epsilon$	$\delta$	Similarity
366.txt	3579.txt	0.05	500	22.7 %
366.txt	3781.txt	0.05	500	33.5 %
366.txt	4798.txt	0.05	500	32.0 %
366.txt	5075.txt	0.05	500	37.2 %

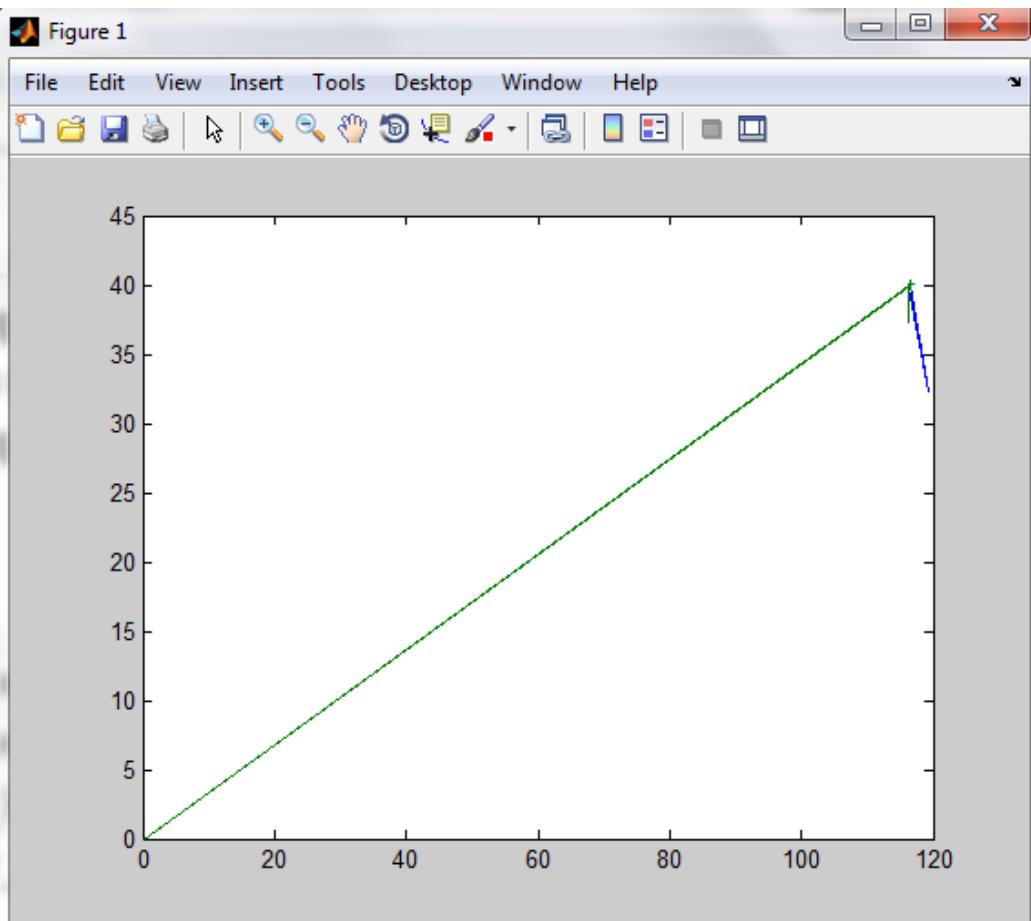
Παρατηρώντας τα παραπάνω αποτελέσματα προκύπτουν τα εξής:

Το αρχείο 366.txt από το φάκελο 1000Points παρουσιάζει μεγαλύτερη ομοιότητα με το αρχείο 5075.txt από το φάκελο 10000Points, με ποσοστό ομοιότητας 37.2%. Η εμφάνιση σε χάρτη της τροχιάς σύγκρισης που περιγράφεται αρχείο 366.txt, σε συνδυασμό με την ομοιότερη τροχιά που βρέθηκε μέσω του LCSS ( 5075.txt ) και η τροχιά σύγκρισης ( 366.txt ), σε συνδυασμό με την ομοιότερη τροχιά ( 5075.txt ) πάνω στους άξονες x, y παρουσιάζονται παρακάτω.

### **366.txt - 5075.txt**

Similarity: 37.2 %





- Για το αρχείο 3557.txt από το φάκελο 1000 Points:

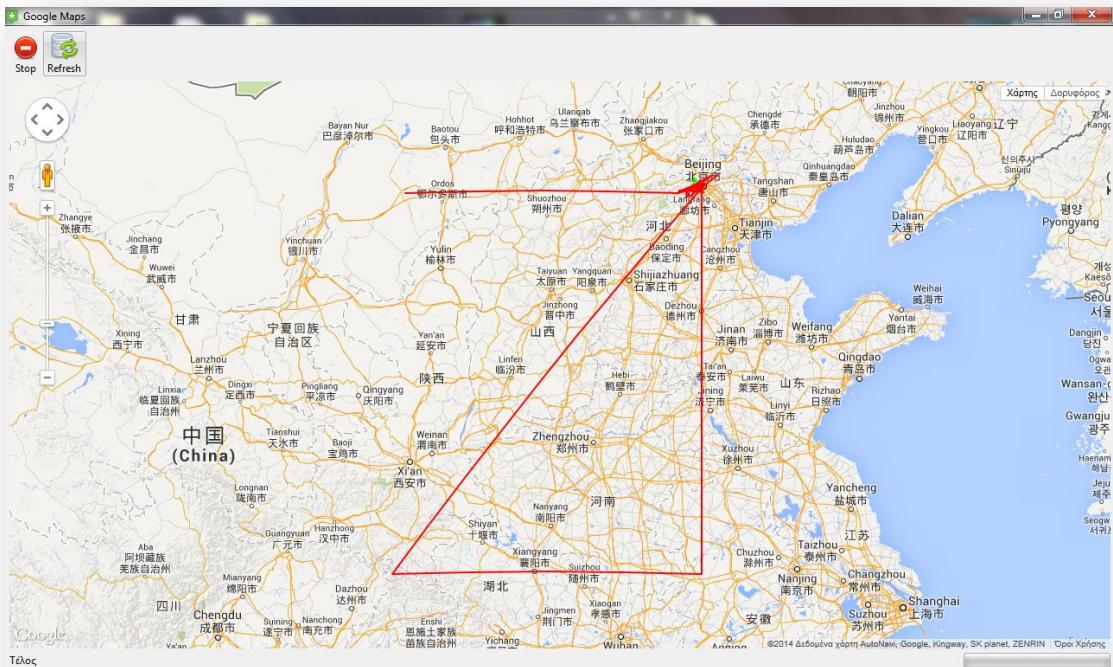
Αρχείο A	Αρχείο B	$\varepsilon$	$\delta$	Similarity
3557.txt	3579.txt	0.05	500	23.1 %
3557.txt	3781.txt	0.05	500	0.3 %
3557.txt	4798.txt	0.05	500	11.6 %
3557.txt	5075.txt	0.05	500	12.9 %

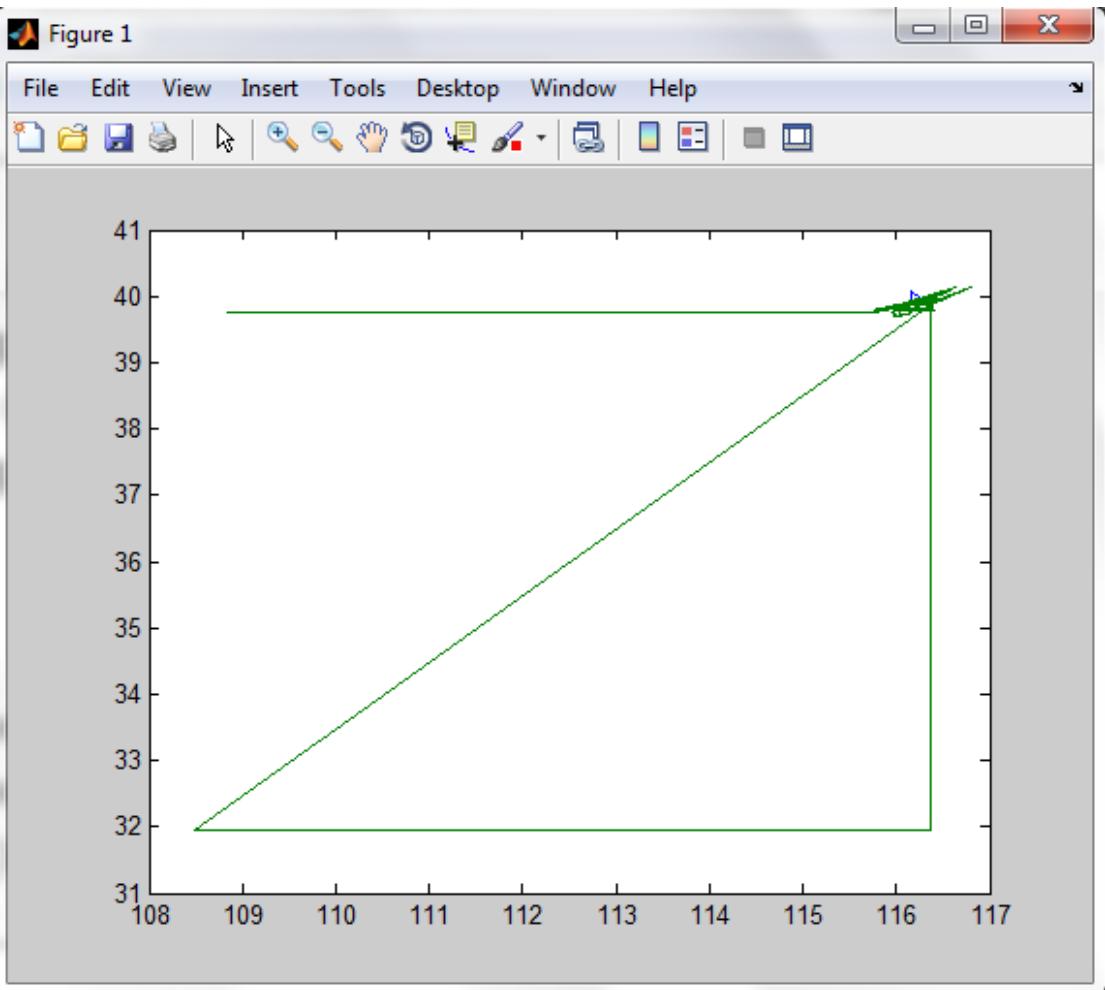
Παρατηρώντας τα παραπάνω αποτελέσματα προκύπτουν τα εξής:

Το αρχείο 3557.txt από το φάκελο 1000Points παρουσιάζει μεγαλύτερη ομοιότητα με το αρχείο 3579.txt από το φάκελο 10000Points, με ποσοστό ομοιότητας 23.1%. Η εμφάνιση σε χάρτη της τροχιάς σύγκρισης που περιγράφεται αρχείο 3557.txt, σε συνδυασμό με την ομοιότερη τροχιά που βρέθηκε μέσω του LCSS ( 3579.txt ) και η τροχιά σύγκρισης ( 3557.txt ), σε συνδυασμό με την ομοιότερη τροχιά ( 3579.txt ) πάνω στους άξονες x, y παρουσιάζονται παρακάτω.

### 3557.txt - 3579.txt

Similarity: 23.1 %





- Για το αρχείο 5099.txt από το φάκελο 1000 Points:

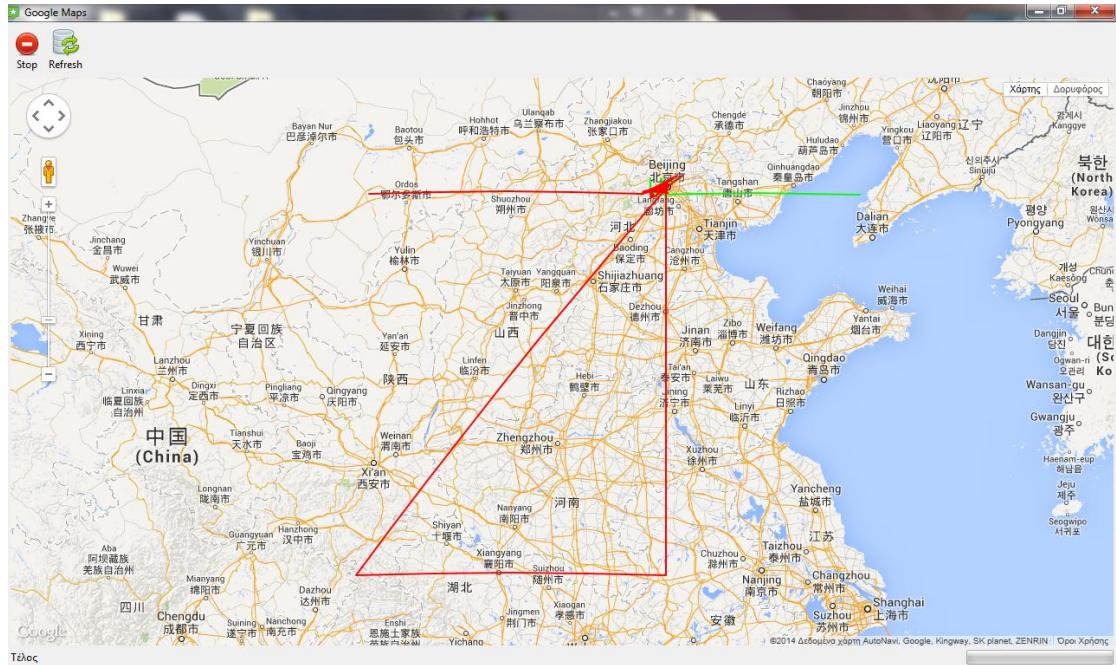
Αρχείο A	Αρχείο B	$\varepsilon$	$\delta$	Similarity
5099.txt	3579.txt	0.05	500	58.8 %
5099.txt	3781.txt	0.05	500	23.79 %
5099.txt	4798.txt	0.05	500	24.3 %
5099.txt	5075.txt	0.05	500	33.4 %

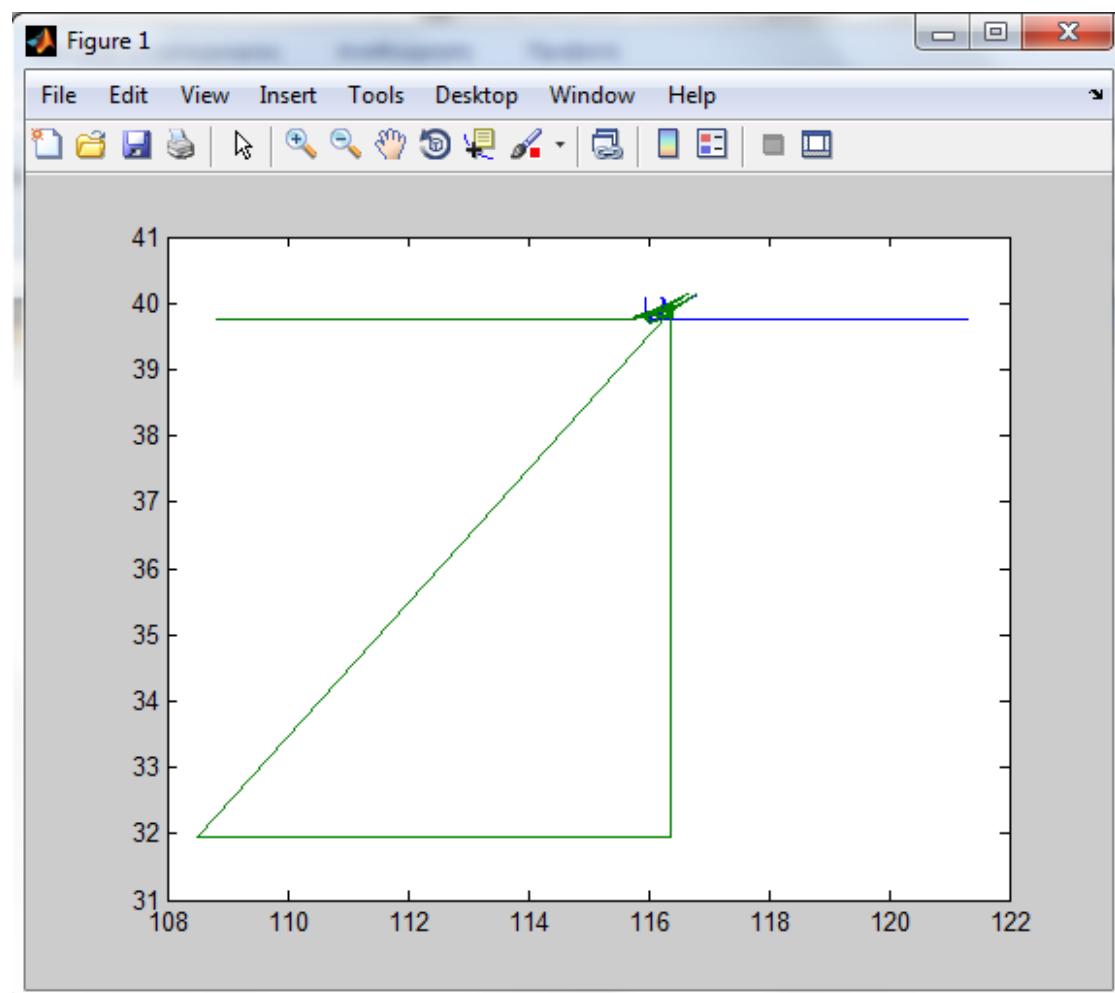
Παρατηρώντας τα παραπάνω αποτελέσματα προκύπτουν τα εξής:

Το αρχείο 5099.txt από το φάκελο 1000Points παρουσιάζει μεγαλύτερη ομοιότητα με το αρχείο 3579.txt από το φάκελο 10000Points, με ποσοστό ομοιότητας 37.2%. Η εμφάνιση σε χάρτη της τροχιάς σύγκρισης που περιγράφεται αρχείο 5099.txt, σε συνδυασμό με την ομοιότερη τροχιά που βρέθηκε μέσω του LCSS (3579.txt) και η τροχιά σύγκρισης ( 5099.txt ), σε συνδυασμό με την ομοιότερη τροχιά (3579.txt ) πάνω στους άξονες x, y παρουσιάζονται παρακάτω.

### 5099.txt - 3579.txt

Similarity: 58.8 %





- Για το αρχείο 6665.txt από το φάκελο 1000 Points:

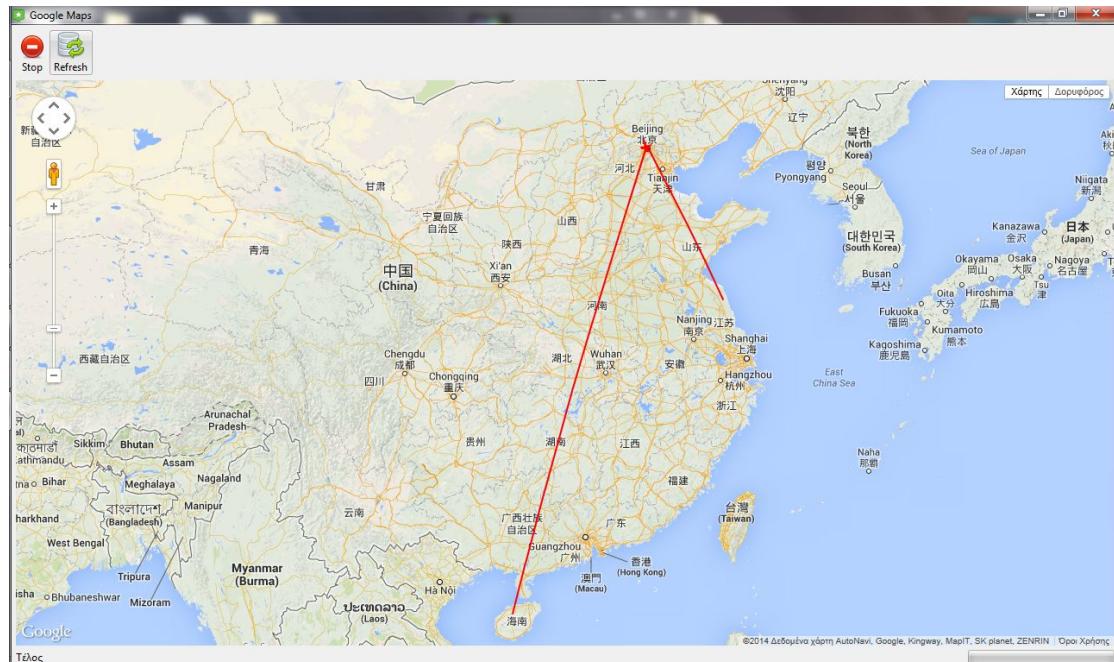
Αρχείο A	Αρχείο B	$\varepsilon$	$\delta$	Similarity
6665.txt	3579.txt	0.05	500	78.7 %
6665.txt	3781.txt	0.05	500	88.0 %
6665.txt	4798.txt	0.05	500	77.9 %
6665.txt	5075.txt	0.05	500	64.1 %

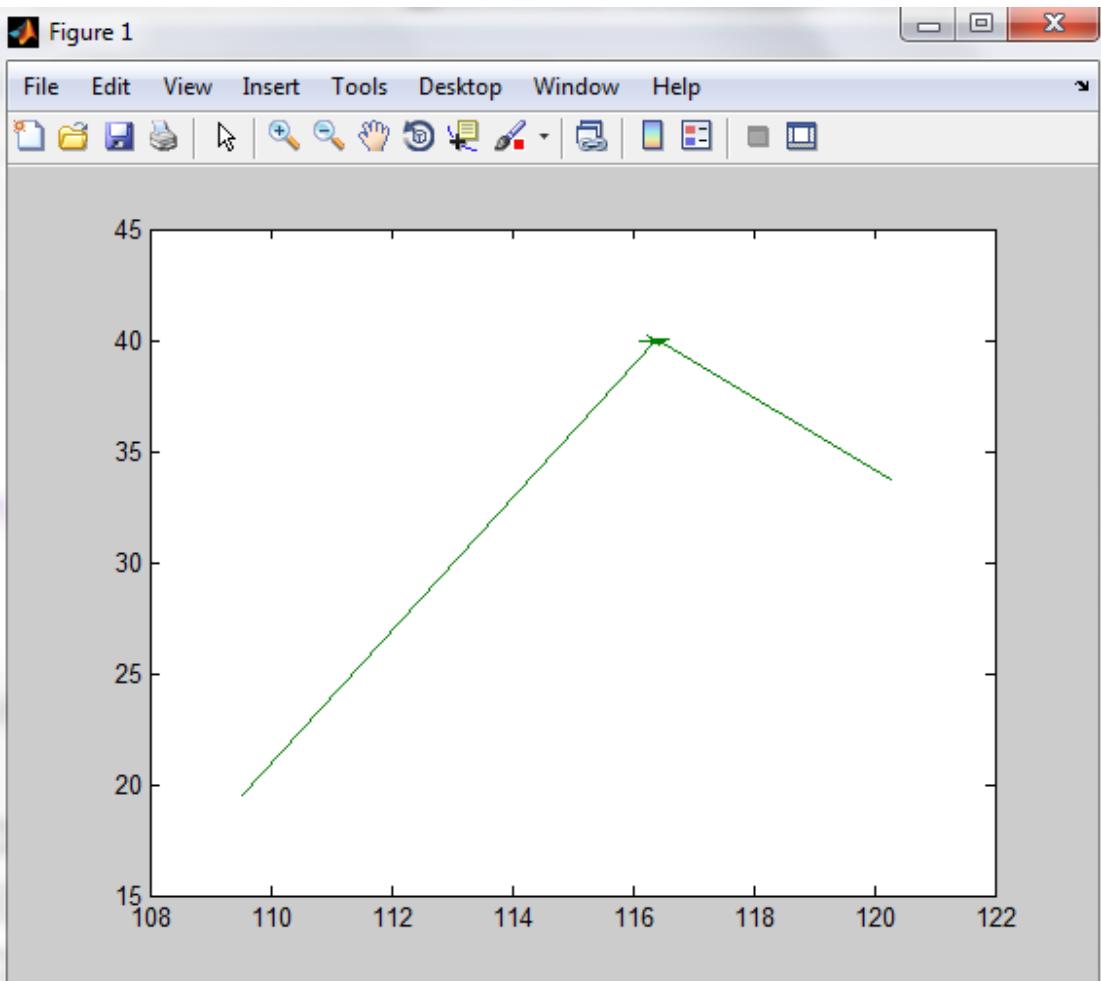
Παρατηρώντας τα παραπάνω αποτελέσματα προκύπτουν τα εξής:

Το αρχείο 6665.txt από το φάκελο 1000Points παρουσιάζει μεγαλύτερη ομοιότητα με το αρχείο 3781.txt από το φάκελο 10000Points, με ποσοστό ομοιότητας 37.2%. Η εμφάνιση σε χάρτη της τροχιάς σύγκρισης που περιγράφεται αρχείο 6665.txt, σε συνδυασμό με την ομοιότερη τροχιά που βρέθηκε μέσω του LCSS (3781.txt) και η τροχιά σύγκρισης ( 6665.txt ), σε συνδυασμό με την ομοιότερη τροχιά (3781.txt ) πάνω στους άξονες x, y παρουσιάζονται παρακάτω.

### 6665.txt - 3781.txt

Similarity: 88.0 %





Προκύπτει επίσης, ότι ο χρόνος που απαιτείται για την εκτέλεση του αλγορίθμου με δεδομένα αρχείο των 1000 Σημείων και αρχείο των 10000 Σημείων είναι προσεγγιστικά 115.0 seconds. Ο χρόνος αυτός είναι αρκετά μεγάλος. Αυτό είναι προφανές, αφού γνωρίζουμε πως η χρονική πολυπλοκότητα του αλγορίθμου LCSS είναι  $O(n^*m)$ , όπου η είναι το μήκος της 1<sup>ης</sup> τροχιάς σύγκρισης (εδώ 1000) και m, το μήκος της 2<sup>ης</sup> τροχιάς σύγκρισης (εδώ 10000) και στην περίπτωση αυτή για  $\delta=500$  έχουμε 8500 επαναλήψεις του αλγορίθμου LCSS.

## **Διαφορές μεταξύ του αλγόριθμου LCSS και του Dynamic Time Warping**

Στην ανάλυση χρονοσειρών, η Δυναμική Χρονική Στρέβλωση (Dynamic Time Warping-DTW) είναι ένας αλγόριθμος για τη μέτρηση της ομοιότητας μεταξύ δύο χρονικών σειρών που μπορεί να διαφέρουν στο χρόνο ή την ταχύτητα. Σε γενικές γραμμές, ο DTW είναι μία μέθοδος που υπολογίζει τη βέλτιστη αντιστοιχία μεταξύ δύο δεδομένων ακολουθιών (π.χ. χρονοσειρές) με ορισμένους περιορισμούς. Οι ακολουθίες «στρεβλώνονται» μη γραμμικά στη διάσταση του χρόνου για τη μέτρηση της ομοιότητάς τους, ανεξάρτητης από ορισμένες μη γραμμικές παραλλαγές στη χρονική διάσταση. Το πρόβλημα της Μέγιστης Κοινής Υπακολουθίας (Longest Common Subsequence-LCS) αφορά στην εύρεση της μέγιστης υπακολουθίας, κοινής σε όλες τις ακολουθίες ενός σετ ακολουθιών (συχνά 2).

Η βασική διαφορά των δύο αλγορίθμων έγκειται στο γεγονός ότι στο DTW όλα τα στοιχεία αντιστοιχίζονται στις δύο ακολουθίες και με αυτόν τον τρόπο επιτυγχάνεται μία ένα-προς-ένα χαρτογράφηση. Το αποτέλεσμα αυτής της συμπεριφοράς του αλγορίθμου είναι ότι υπάρχει κίνδυνος οι ακραίες τιμές να στρεβλώσουν τη μέτρηση της απόστασης, δηλαδή το τελικό αποτέλεσμα. Σε αντίθεση, σύμφωνα με τον αλγόριθμο LCSS, οι απομακρυσμένες τιμές δεν αντιστοιχίζονται στην άλλη ακολουθία (επιτρέπονται τα κενά) και με αυτόν τον τρόπο υπάρχει μικρότερη παραμόρφωση της απόστασης/ομοιότητας εξαιτίας του θορύβου.

Συμπεραίνουμε ότι ο αλγόριθμος LCSS είναι προτιμότερος και αποδοτικότερος σε δεδομένα που περιέχουν αρκετό θόρυβο, καθώς είναι σε θέση να τον αγνοήσει και να παράγει αποτέλεσμα σύμφωνα με τα πραγματικά ωφέλιμα δεδομένα. Αντιθέτως, στην περίπτωση που τα δεδομένα είναι στην πραγματικότητα αυτά που εξετάζει ο αλγόριθμος και δεν έχει εισχωρήσει θόρυβος, ο αλγόριθμος DTW είναι ο καταλληλότερος. Ως προς τη χρονική πολυπλοκότητα, στην περίπτωση σύγκρισης δύο ακολουθιών (τροχιές σύγκρισης), και οι δύο αλγόριθμοι χρειάζονται  $O(n*m)$ , όπου  $n$  και  $m$  τα μήκη των δύο ακολουθιών αντίστοιχα. Αξίζει να σημειωθεί πως υπάρχουν βελτιστοποιήσεις του LCSS τόσο για καλύτερους χρόνους, όσο και για αποδοτικότερη διαχείριση μνήμης.