

Ιστορίες τρόμου από το *Θαυμαστό Καινούργιο Κόσμο* ~~του Aldus Huxley~~ των Social Media



Καθηγητής Δημήτρης Γκρίτζαλης, Υ.Δ. Μίλτος Κάνδιας
Τμήμα Πληροφορικής, Οικονομικό Πανεπιστήμιο Αθηνών

Περιεχόμενα



- Δεδομένα στα Social Media
- Πώς μπορούν να χρησιμοποιηθούν αυτά τα δεδομένα
- Δύο ιστορίες τρόμου (;)

Twitter: Πρόβλεψη *εκ των έσω* απειλών

You Tube: Αποκάλυψη *ευαίσθητων* προσωπικών δεδομένων

- Συμπεράσματα και αναδυόμενοι κίνδυνοι
- Σύννομες (;) εφαρμογές

Δεδομένα στα Social Media



- Η δομή των Social Media επιτρέπει προσωποποιημένη χρήση
- Η πλειονότητα των δεδομένων στο διαδίκτυο είναι *user generated content*
- Αποκαλύπτονται στοιχεία της προσωπικότητας των χρηστών
- Υπάρχουν ποικίλα κίνητρα χρήσης (Επαγγελματικά, Διασκέδαση, Επικοινωνία κλπ.)
- Οι χρήστες αναπαράγουν τη συμπεριφορά τους (και) στο Διαδίκτυο^{1,2}
- Συγκεντρώνεται τεράστιος όγκος δημόσια διαθέσιμων δεδομένων τύπου Open

Source Intelligence (OSINT)

- Πληροφορίες προερχόμενες από δημόσια διαθέσιμα δεδομένα, που αξιοποιούνται μεταδιδόμενες έγκαιρα για την αντιμετώπιση ειδικών αναγκών πληροφόρησης (DoD, USA).

[1]. Ross C., Orr E, Sisic M., Arseneault J., Simmering M., Orr R., “Personality and mo-ti-va-ti--ons associated with Facebook use”, *Computers in Human Behavior*, Vol. 25, pp. 578-586, 2009.

[2]. Amichai-Hamburger Y., Vinitzky G., “Social network use and personality”, *Computers in Human Behavior*, Vol. 26, pp. 1289-1295, 2010.

Αξιοποίηση δεδομένων Open Source Intelligence (OSINT)



- Δημιουργία μορφότυπων (profiles) χρήσης και χρηστών
- Δυνατότητα (αυτόματων) ψυχομετρικών αξιολογήσεων
- Αποκάλυψη ευαίσθητων προσωπικών δεδομένων³
 - Πολιτικές πεποιθήσεις
 - Θρησκευτικές επιλογές
 - Σεξουαλικές προτιμήσεις
- Ζητούμενο για τη διαχείριση των δεδομένων αυτών είναι η θέσμιση σχετικών ψηφιακών δικαιωμάτων

[3]. Kosinski M., Stillwell D., Graepel T., “Private traits and attributes are predictable from digital records of human behavior”, in *Proc. of the National Academy of Sciences*, 2013.



1^η ιστορία (τρόμου;)

Πρόβλεψη εκ των έσω απειλών
μέσω **Twitter**

Εντοπίζοντας εκ των έσω απειλές μέσω ...ναρκισσισμού και ...Twitter!



- **Social Medium:**

Twitter

- **Δεδομένα:**

1.075.859 χρήστες, 7.125.561 συνδέσεις

- **Περιεχόμενο:**

Συνολικά παραχθέν από 41.818 χρήστες

- **Γραφο-θεωρητική ανάλυση⁴:**

Strongly Connected Components (153.121 χρήστες σχηματίζουν γράφο, όπου υπάρχει μονοπάτι από όλους προς όλους)

Node Loneliness (1.075.815 χρήστες έχουν σύνδεση με ≥ 1 χρήστη)

Small World Phenomenon (όλοι οι χρήστες γνωρίζονται μεταξύ τους, με απόσταση ≤ 6 βημάτων)

Indegree Distribution (κατανομή εισερχόμενων ακμών σε κάθε κόμβο)

Outdegree Distribution (κατανομή εξερχόμενων ακμών σε κάθε κόμβο)

Ελεύθερη πρόσβαση σε δεδομένα χρηστών. Ύπαρξη μηχανισμών διευκόλυνσης της συλλογής τους.

[4]. Kandias M., Galbogini L., Mitrou L., Gritzalis D., "Insiders trapped in the mirror reveal themselves in social media", in *Proc. of the 7th International Conference on Network and System Security*, Spain, June 2013.

Εντοπίζοντας εκ των έσω απειλές μέσω ...ναρκισσισμού και ...Twitter!



Στατιστική κατανομή-ανάλυση χρηστών

Ένταση χρήσης του Social Medium (usage intensity)

Αξιολογείται με βάση το σύνολο του υλικού που έχει δημιουργήσει ο χρήστης

Επιρροή στο Social Medium (influence valuation)

Αξιολογείται με βάση τον αριθμό των χρηστών που εκτίθενται στα tweets ενός χρήστη (followers, retweets, mentions)

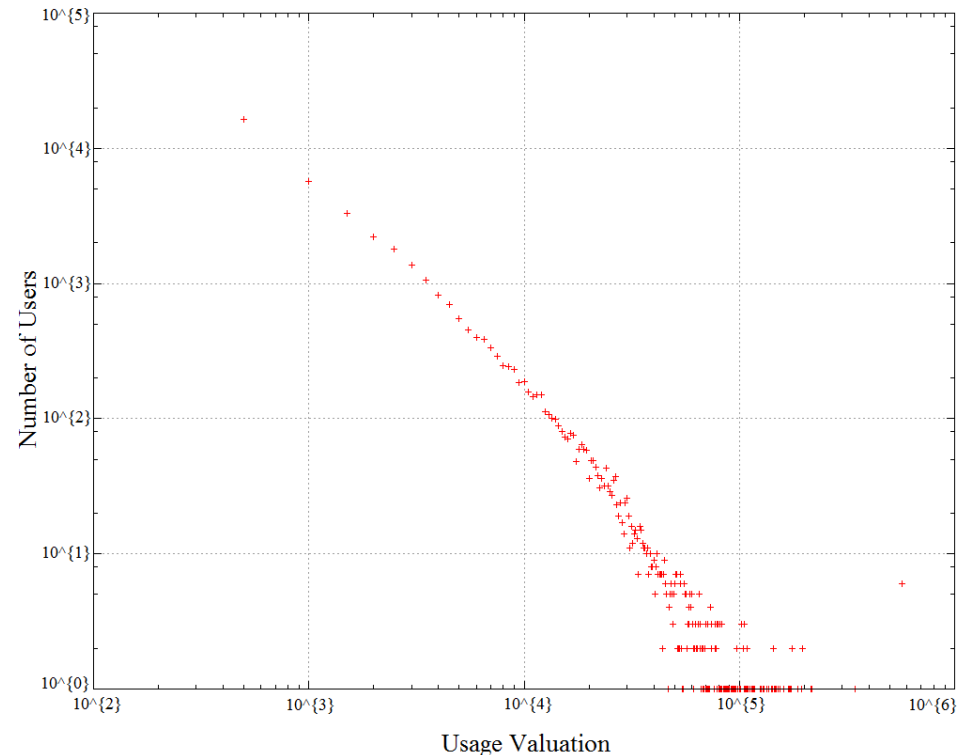
Klout score

Αξιολογείται ο βαθμός συμμετοχής ενός χρήστη στο Social Medium.

Εντοπίζοντας εκ των έσω απειλές μέσω ...ναρκισσισμού και ...Twitter!



- **Small World Phenomenon**
 - 99% των χρηστών βρίσκεται σε απόσταση ≤ 6 βημάτων από κάθε άλλο χρήστη μέσα στο γράφο.
- **Indegree Distribution**
 - Κατανομή εισερχόμενων ακμών σε κάθε κόμβο. Μέσος όρος 13.2 followers/χρήστη.
- **Outdegree Distribution**
 - Κατανομή εξερχόμενων ακμών σε κάθε κόμβο. Μέσος όρος 11 followings/χρήστη.
- **Usage Intensity Distribution**
 - Κατανομή έντασης χρήσης.



Εντοπίζοντας εκ των έσω απειλές μέσω ...ναρκισσισμού και ...Twitter!

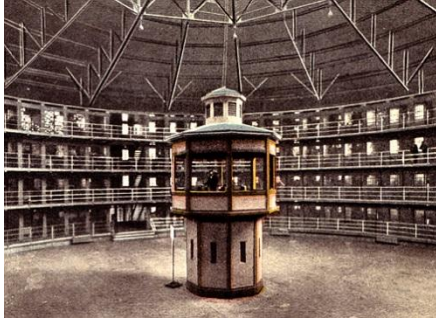


Κατηγορία	Αξιολόγηση επιρροής	Klout score	Αξιολόγηση έντασης χρήσης
Απομονωμένοι χρήστες	0 - 90	3.6 - 11.1	0 - 500
Συνήθεις χρήστες	90 - 283	11.1 - 26.0	500 - 4500
Γνωστοί χρήστες	283 - 1011	26.0 - 50.0	4500 - 21000
ΜΜΕ, νάρκισσοι, διασημότητες	1011 - 3604	50.0 - 81.99	21000 - 569000

Βασικά συμπεράσματα



- ✓ Οι εντασσόμενοι στην 4η κατηγορία επιδεικνύουν ναρκισσιστική συμπεριφορά που συνδέεται με προδιάθεση παραβατικής συμπεριφοράς.
- ✓ Μπορούν να μελετηθούν χρήστες υπό το πρίσμα διαφόρων χαρακτηριστικών (επάγγελμα, γεωγραφική περιοχή κλπ.).
- ✓ Νέα κατηγορία δεδομένων: Όσα προκύπτουν μέσω των δεδομένων που οικειοθελώς αποκαλύπτουν οι χρήστες.



2^η ιστορία (τρόμου;)

Πανοπτικόν: Αποκάλυψη ευαίσθητων προσωπικών δεδομένων μέσω **You Tube**

Πανοπτικόν: Η «ιδανική φυλακή» του J. Bentham. Ένα οικοδόμημα σχεδιασμένο με τέτοιο τρόπο, ώστε οι κατάδικοι να παρακολουθούνται διαρκώς από ένα κεντρικό σημείο.

Αποκαλύπτοντας πολιτικές πεποιθήσεις μέσω ...data mining και ...You Tube!



- **Social Medium:**

YouTube

- **Δεδομένα (Νοε 2005 - Δεκ 2012):**

12.964 χρήστες, 207.377 βίντεο, 2.034.362 σχόλια

Ανωνυμοποίηση όλων των δεδομένων, για να μην υπάρχει συσχέτιση ονόματος χρήστη – πεποιθήσεων (MD5 hashes αντί ονομάτων χρηστών), με βάση το N. 2472/1997.

- **Πρόσφορο πεδίο εφαρμογής:**

Πολιτικό περιεχόμενο, οπτικοακουστικά ερεθίσματα, συναισθηματική φόρτιση, ευρεία συμμετοχή χρηστών

- **Μέθοδοι ανάλυσης:**

Γραφοθεωρητική ανάλυση (Small World Phenomenon, Indegree/Outdegree Distribution, Node Loneliness)

Ανάλυση περιεχομένου (εξαγωγή συμπερασμάτων από ανάλυση σχολίων χρηστών, λεκτική ανάλυση – Opinion Mining, Machine Learning)

Ανάλυση νέφους ετικετών (συγκέντρωση ετικετών βίντεο και απεικόνιση σε νέφος ώστε να μελετηθεί το περιεχόμενο – Tag Cloud)

Εξελιγμένοι μηχανισμοί διευκόλυνσης της συλλογής δεδομένων χρηστών.



Αποκαλύπτοντας πολιτικές πεποιθήσεις μέσω ...data mining και ...You Tube!



Κατηγορίες Χρηστών (ενδεικτικές):

Κεντρώοι-Κεντροαριστεροί, Ουδέτεροι, Κεντρώοι-Κεντροδεξιοί

Συνδρομή Ειδικών (εντοπισμός patterns):

Κοινωνιολόγος και Πολιτικός Επιστήμονας

Αλγόριθμος: Multinomial Logistic Regression (MLR)

ΚΑΤΗΓΟΡΙΕΣ ΜΕΤΡΙΚΕΣ	Κεντρώοι - Κεντροαριστεροί	Ουδέτεροι - Απαθείς	Κεντρώοι - Κεντροδεξιοί
Precision	83%	91%	77%
Recall	77%	93%	78%
F-Score	80%	92%	77%
Accuracy	87%		

Precision: Ο αριθμός χρηστών που έχουν κατηγοριοποιηθεί ορθά, διαιρούμενος με το πλήθος των χρηστών που εντάχθηκαν σε αυτή την κατηγορία.

Recall: Ο αριθμός χρηστών που έχουν κατηγοριοποιηθεί ορθά, διαιρούμενος με το συνολικό αριθμό των χρηστών της κατηγορίας.

F-Score: Σταθμισμένος αρμονικός μέσος των **Precision** και **Recall**.

Accuracy: Ο αριθμός των ορθών κατηγοριοποιήσεων. Ισούται με το πηλίκο των ορθών κατηγοριοποιήσεων δια του συνόλου των δεδομένων.



Βασικά συμπεράσματα



- ✓ Δυνατότητα αυτοματοποιημένης ένταξης χρηστών σε προκαθορισμένες κατηγορίες
- ✓ Δυνατότητα αποκάλυψης ευαίσθητων προσωπικών δεδομένων με αυτοματοποιημένο τρόπο

Αναδυόμενοι κίνδυνοι



- ✓ Διεύρυνση κοινωνικών ανισοτήτων και διάχυση προκαταλήψεων
- ✓ Δημιουργία αρνητικού περιβάλλοντος (εργασίας, ζωής)
- ✓ Εκμετάλλευση μειονοτήτων
- ✓ Διακρίσεις κοινωνικού και εργασιακού χαρακτήρα
- ✓ Διεύρυνση ψηφιακού χάσματος
- ✓ Νέα ζητήματα (διαχείριση πληροφορίας, ιστορικότητα πληροφορίας, επεξεργασία πληροφορίας)



- ✓ Αξιολόγηση προδιάθεσης παραβατικής συμπεριφοράς⁵ (πχ. εκ των έσω απειλές σε κρίσιμες υποδομές).
- ✓ Μορφότυπος χρήστη με στόχο προσωποποιημένη προβολή (πχ. μορφότυπος καταναλωτή).
- ✓ Προληπτική προστασία ανηλίκων (πχ. παρεμπόδιση αυτοκτονιών εφήβων).

[5]. U.S. Dept. of Justice, “The insider threat: An in-tro-duc-tion to detecting and deterring insider spy”, FBI, USA, 2012, <http://www.fbi.gov/about-us/investigate/counterintelligence/the-insider-threat>