Διασπορά Ψευδών Ειδήσεων Σε Κοινωνικά Δίκτυα Μικρο-Ιστολογίων

Τεχνικές Πρώιμης Ανίχνευσης και Αντιμετώπισης

Λέσχη των Μαθηματικών «Μεθόδιος Ανθρακίτης»

Σπύρος Κοντογιάννης

kontog@uoi.gr

Τμήμα Μηχανικών Η/Υ & Πληροφορικής, Πανεπιστήμιο Ιωαννίνων

Πέμπτη, 28 Νοεμβρίου 2019

Σκελετός Ομιλίας

1. Εισαγωγικά: Το Πρόβλημα των Ψευδών Ειδήσεων
   - Ορισμός, Ιδιότητες, Προκλήσεις, Εμπλεκόμενοι
   - Υπάρχουσες Τεχνικές Αντιμετώπισης Ψευδών Ειδήσεων
2. Μέθοδοι Ανίχνευσης Ψευδών Ειδήσεων
   - Μέθοδοι Αξιοποίησης Περιεχομένου
   - Μέθοδοι Αξιοποίησης Αναδράσεων
3. Μελέτη Περίπτωσης: Ακολουθιακό Μοντέλο Διάδοσης
   - Περιγραφή Μοντέλου
   - Ρόλος των Χρηστών
   - Ρόλος της Πλατφόρμας

Ψευδείς Ειδήσεις και Διασπορά τους

- Η διάδοση των ειδήσεων έχει αλλάξει ριζικά.
  - Στο παρελθόν τον πρώτο λόγο είχαν οι **επαγγελματίες** στο χώρο των ειδήσεων.
  - Σήμερα ο καθένας από εμάς είναι ένας **επίδοξος δημοσιογράφος**.

Ψευδείς Ειδήσεις και Διασπορά τους

- Η διάδοση των ειδήσεων έχει αλλάξει ριζικά.
  - Στο παρελθόν τον πρώτο λόγο είχαν οι **επαγγελματίες** στο χώρο των ειδήσεων.
  - Σήμερα ο καθένας από εμάς είναι ένας **επίδοξος δημοσιογράφος**.
- Η διασπορά – και «ψευδών» – ειδήσεων επιταχύνεται σημαντικά λόγω της **τεχνολογίας** (30-40%).

Ψευδείς Ειδήσεις και Διασπορά τους

- Η διάδοση των ειδήσεων έχει αλλάξει ριζικά.
  - Στο παρελθόν τον πρώτο λόγο είχαν οι **επαγγελματίες** στο χώρο των ειδήσεων.
  - Σήμερα ο καθένας από εμάς είναι ένας **επίδοξος δημοσιογράφος**.
- Η διασπορά – και «ψευδών» – ειδήσεων επιταχύνεται σημαντικά λόγω της **τεχνολογίας** (30-40%).
- Ανάγκη για επαναπροσανατολισμό των τεχνολογικών κολοσσών, για τον **εντοπισμό** και κυρίως τον **περιορισμό** της επίδρασης της διασποράς «ψευδών» ειδήσεων.

Ψευδείς Ειδήσεις και Διασπορά τους

- Η διάδοση των ειδήσεων έχει αλλάξει ριζικά.
  - Στο παρελθόν τον πρώτο λόγο είχαν οι **επαγγελματίες** στο χώρο των ειδήσεων.
  - Σήμερα ο καθένας από εμάς είναι ένας **επίδοξος δημοσιογράφος**.
- Η διασπορά – και «ψευδών» – ειδήσεων επιταχύνεται σημαντικά λόγω της **τεχνολογίας** (30-40%).
- Ανάγκη για επαναπροσανατολισμό των τεχνολογικών κολοσσών, για τον **εντοπισμό** και κυρίως τον **περιορισμό** της επίδρασης της διασποράς «ψευδών» ειδήσεων.
- Σίγουρα γίνεται. Πρέπει όμως;

ΟΡΙΣΜΟΙ: Ψευδείς Πληροφορίες

hoax | Μια **ψευδής** ιστορία που χρησιμοποιείται για να αποκρύψει την αλήθεια.

ΟΡΙΣΜΟΙ: Ψευδείς Πληροφορίες

hoax	Μια **ψευδής** ιστορία που χρησιμοποιείται για να αποκρύψει την αλήθεια.
rumor	Μια **ανεπιβεβαίωτη** πληροφορία (στο μέλλον θα επιβεβαιωθεί, θα διαψευστεί, ή θα παραμείνει ανεπιβεβαίωτη). Διαδίδεται σε ΜΜΕ/κοινωνικά δίκτυα και επηρεάζει απόψεις τρίτων.

ΟΡΙΣΜΟΙ: Ψευδείς Πληροφορίες

hoax	Μια **ψευδής** ιστορία που χρησιμοποιείται για να αποκρύψει την αλήθεια.
rumor	Μια **ανεπιβεβαίωτη** πληροφορία (στο μέλλον θα επιβεβαιωθεί, θα διαψευστεί, ή θα παραμείνει ανεπιβεβαίωτη). Διαδίδεται σε ΜΜΕ/κοινωνικά δίκτυα και επηρεάζει απόψεις τρίτων.
fake news	Μια είδηση που περιλαμβάνει **διαστρευλωμένη πληροφορία**, η οποία μπορεί να διαψευστεί μέσω αποδεικτικών στοιχείων. Μεταδίδεται σε ΜΜΕ/κοινωνικά δίκτυα.

ΟΡΙΣΜΟΙ: Ψευδείς Πληροφορίες

hoax	Μια **ψευδής** ιστορία που χρησιμοποιείται για να αποκρύψει την αλήθεια.
rumor	Μια **ανεπιβεβαίωτη** πληροφορία (στο μέλλον θα επιβεβαιωθεί, θα διαψευστεί, ή θα παραμείνει ανεπιβεβαίωτη). Διαδίδεται σε ΜΜΕ/κοινωνικά δίκτυα και επηρεάζει απόψεις τρίτων.
fake news	Μια είδηση που περιλαμβάνει **διαστρευλωμένη πληροφορία**, η οποία μπορεί να διαψευστεί μέσω αποδεικτικών στοιχείων. Μεταδίδεται σε ΜΜΕ/κοινωνικά δίκτυα.
fake review	Μια **ανειλικρινής αξιολόγηση** αγαθού/υπηρεσίας. Βασίζεται σε υποκειμενική άποψη (δεν υπάρχει αντικειμενική αλήθεια παρά μόνο υποκειμενικές κρίσεις). Δε διασπείρεται σε κοινωνικά δίκτυα. Συνοδεύεται με **αξιολόγηση** (rating).

ΟΡΙΣΜΟΙ: Ψευδείς Πληροφορίες

fake news

Μια είδηση που περιλαμβάνει **διαστρευλωμένη πληροφορία**, η οποία μπορεί να διαψευστεί μέσω αποδεικτικών στοιχείων. Μεταδίδεται σε ΜΜΕ/κοινωνικά δίκτυα.

fake news = hoaxes + false rumors

Εξειδικεύσεις Ψευδών Ειδήσεων και Κίνητρα

- Wardle (2017)
  - **misinformation:** **Ακούσια** διάδοση ψευδούς πληροφορίας.
  - **disinformation:** **Εκούσια** δημιουργία και διάδοση πληροφορίας, η οποία είναι γνωστό ότι είναι ψευδής.

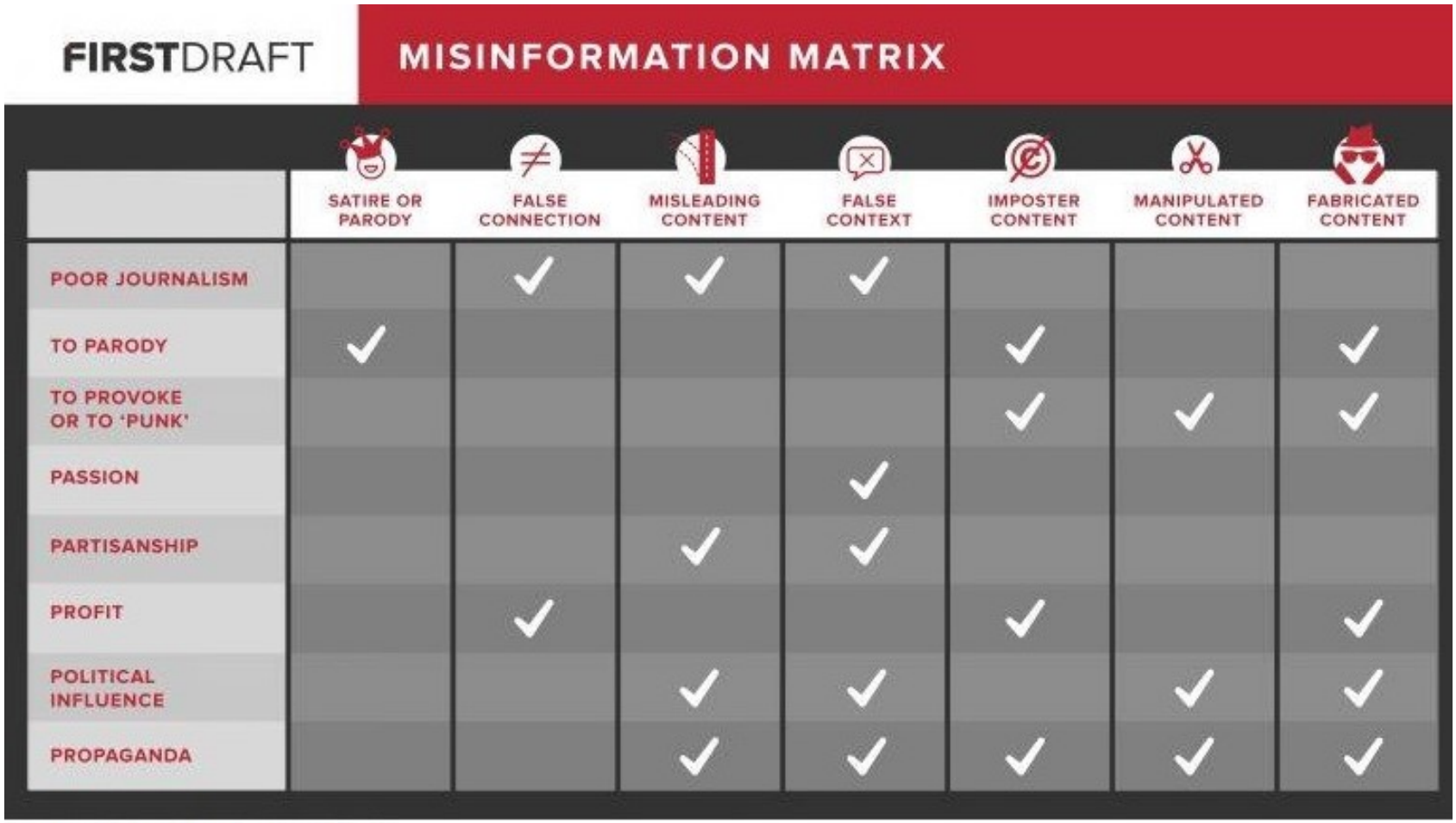
Εξειδικεύσεις Ψευδών Ειδήσεων και Κίνητρα

- Wardle (2017)
  - **misinformation:** **Ακούσια** διάδοση ψευδούς πληροφορίας.
  - **disinformation:** **Εκούσια** δημιουργία και διάδοση πληροφορίας, η οποία είναι γνωστό ότι είναι ψευδής.

Credit: Claire Wardle, First Draft

Εξειδικεύσεις Ψευδών Ειδήσεων και Κίνητρα

- Wardle (2017)
  - **misinformation:** **Ακούσια** διάδοση ψευδούς πληροφορίας.
  - **disinformation:** **Εκούσια** δημιουργία και διάδοση πληροφορίας, η οποία είναι γνωστό ότι είναι ψευδής.



Credit: Claire Wardle. First Draft

Εξειδικεύσεις Ψευδών Ειδήσεων και Κίνητρα

- Wardle (2017)
  - **misinformation:** **Ακούσια** διάδοση ψευδούς πληροφορίας.
  - **disinformation:** **Εκούσια** δημιουργία και διάδοση πληροφορίας, η οποία είναι γνωστό ότι είναι ψευδής.

- Collins English Dictionary
  - **fake news:** Ψευδής πληροφορία που διαδίδεται με τη μορφή έγκυρης είδησης.

Εξειδικεύσεις Ψευδών Ειδήσεων και Κίνητρα

- Wardle (2017)
  - **misinformation:** **Ακούσια** διάδοση ψευδούς πληροφορίας.
  - **disinformation:** **Εκούσια** δημιουργία και διάδοση πληροφορίας, η οποία είναι *γνωστό* ότι είναι ψευδής.

- Collins English Dictionary
  - **fake news:** Ψευδής πληροφορία που διαδίδεται με τη μορφή έγκυρης είδησης.

- Allcott & Gentzkow (2017)
  - **fake news:** Αποδεδειγμένα *εσφαλμένη* πληροφορία που παρουσιάζεται ως είδηση, *με σκοπό την εξαπάτηση*.

Εξειδικεύσεις Ψευδών Ειδήσεων και Κίνητρα

- Wardle (2017)
  - **misinformation:** **Ακούσια** διάδοση ψευδούς πληροφορίας.
  - **disinformation:** **Εκούσια** δημιουργία και διάδοση πληροφορίας, η οποία είναι γνωστό ότι είναι ψευδής.

- Collins English Dictionary
  - **fake news:** Ψευδής πληροφορία που διαδίδεται με τη μορφή έγκυρης είδησης.

- Allcott & Gentzkow (2017)
  - **fake news:** Αποδεδειγμένα *εσφαλμένη* πληροφορία που παρουσιάζεται ως είδηση, *με σκοπό την εξαπάτηση*.

- Sharma et al (2019)
  - **fake news:** Ειδησεογραφικό άρθρο ή μήνυμα που δημοσιεύεται και μεταδίδεται σε ΜΜΕ/κοινωνικά δίκτυα, το οποίο περιλαμβάνει *εσφαλμένη* πληροφορία, *ανεξάρτητα από τους σκοπούς δημιουργίας του*.

Παραγωγή και Διάδοση Ειδήσεων

...παραδοσιακή γραμμή...

- Σύνθεση είδησης (συγγραφή άρθρου)

Παραγωγή και Διάδοση Ειδήσεων

...παραδοσιακή γραμμή...

- Διασταύρωση πληροφοριών και ποιοτικός έλεγχος

Παραγωγή και Διάδοση Ειδήσεων

...παραδοσιακή γραμμή...

- Οριστικοποίηση είδησης (τύπωση άρθρου)

Παραγωγή και Διάδοση Ειδήσεων

...παραδοσιακή γραμμή...

- Ανάδειξη σημαντικότητας

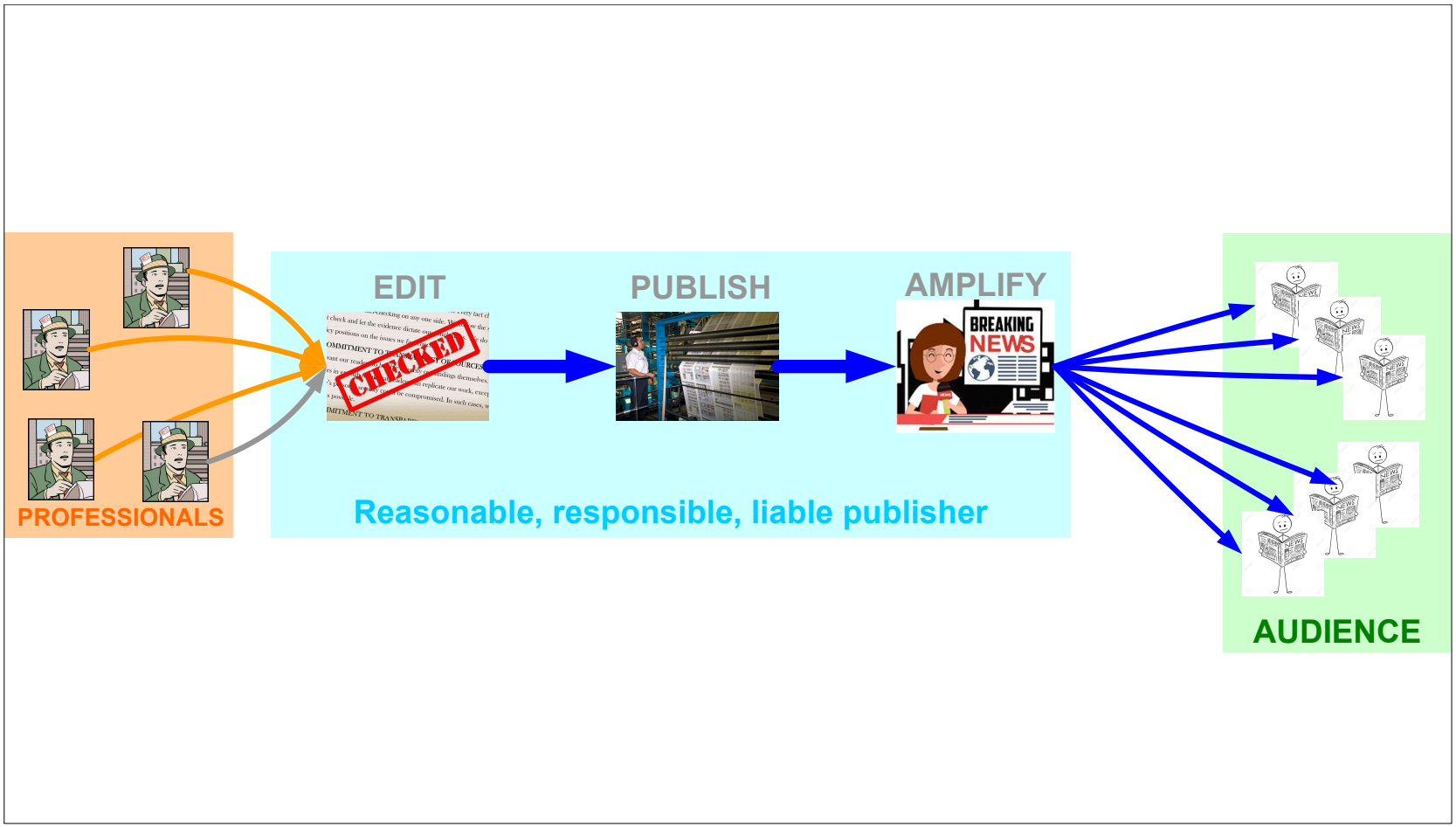
Παραγωγή και Διάδοση Ειδήσεων

...παραδοσιακή γραμμή...

- «Κατανάλωση» είδησης

Παραγωγή και Διάδοση Ειδήσεων

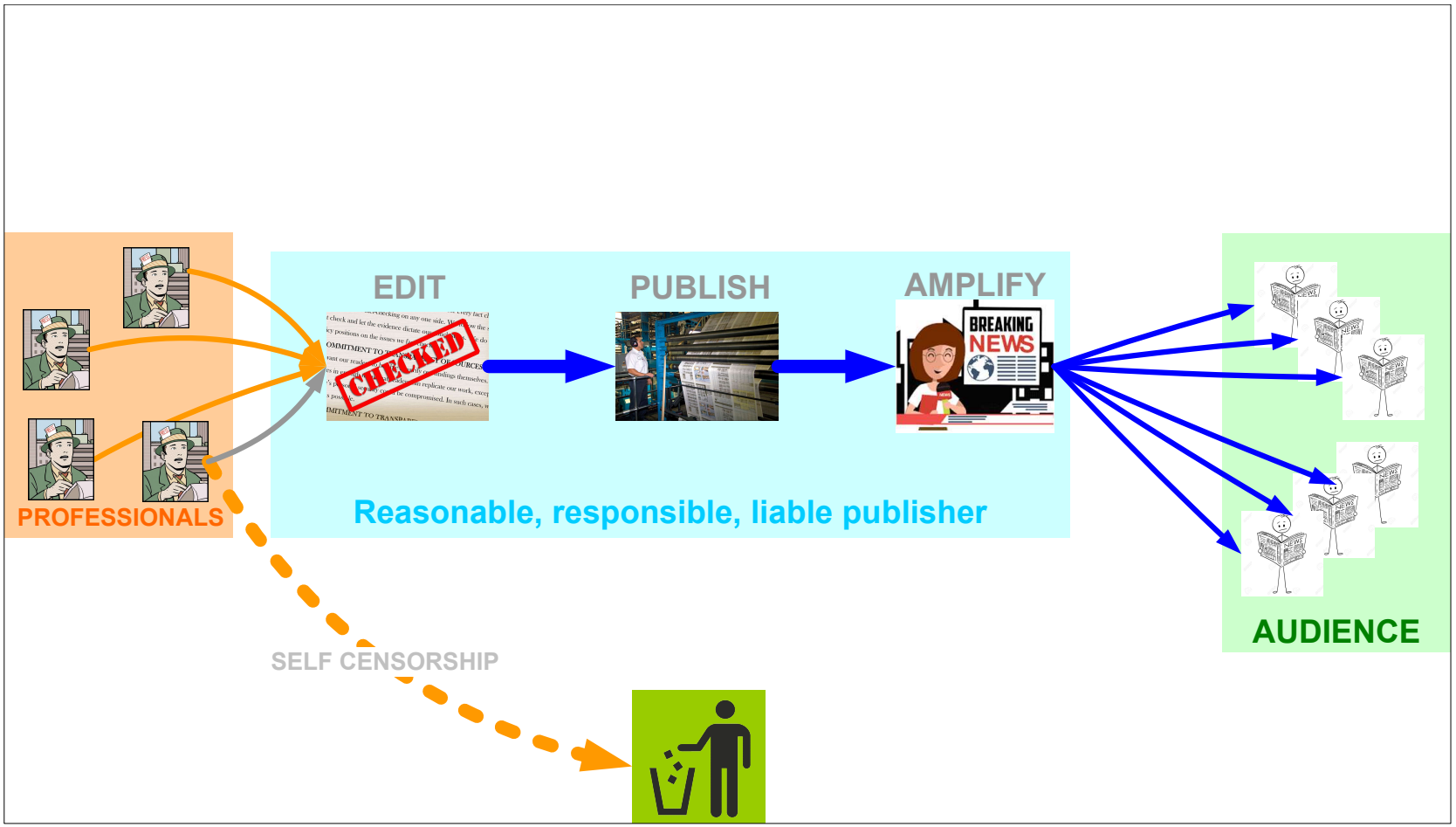
...η παραδοσιακή γραμμή (σχηματικά)...



Producer – Consumer Model

Παραγωγή και Διάδοση Ειδήσεων

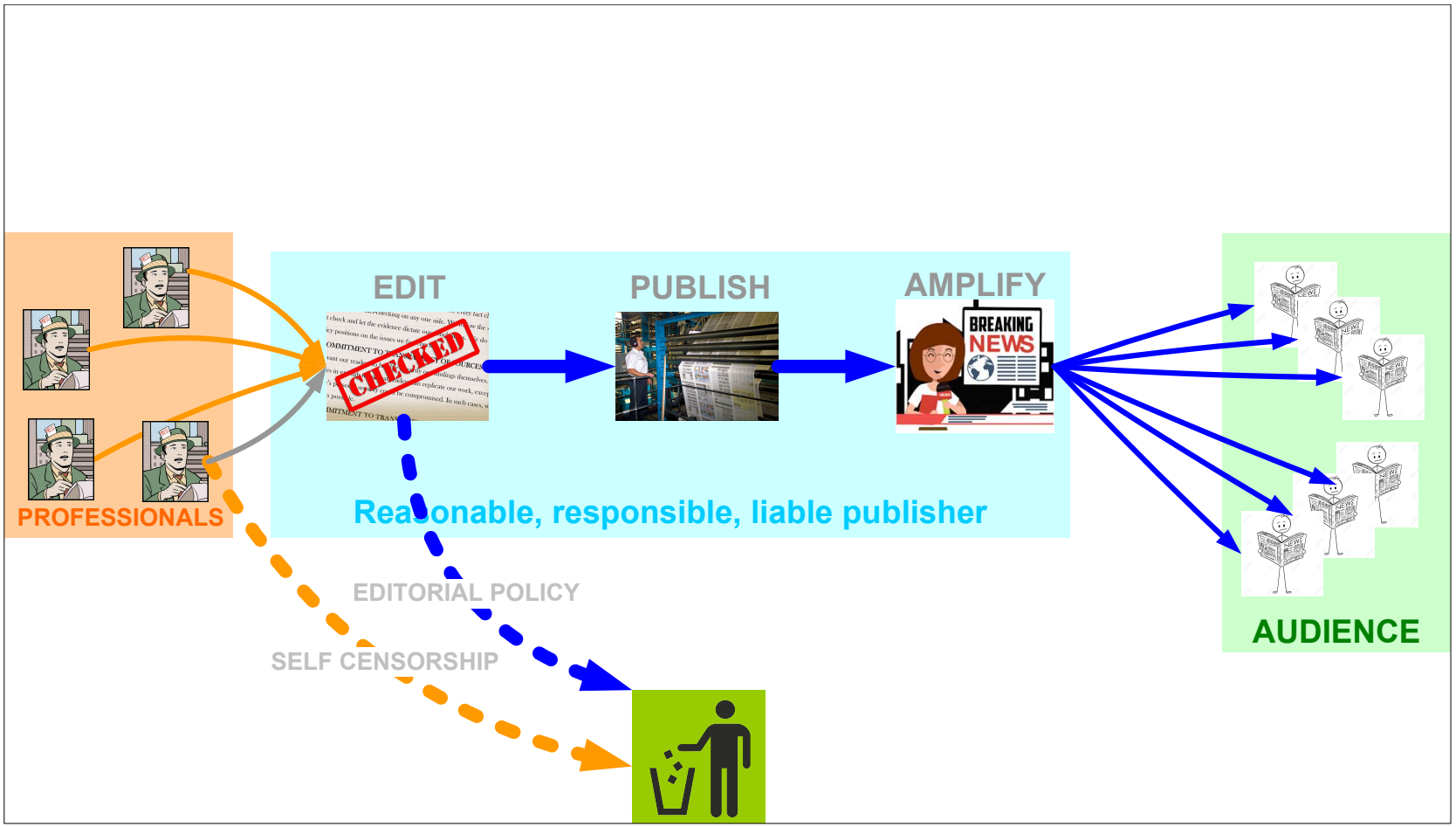
...η παραδοσιακή γραμμή (σχηματικά)...



Producer – Consumer Model

Παραγωγή και Διάδοση Ειδήσεων

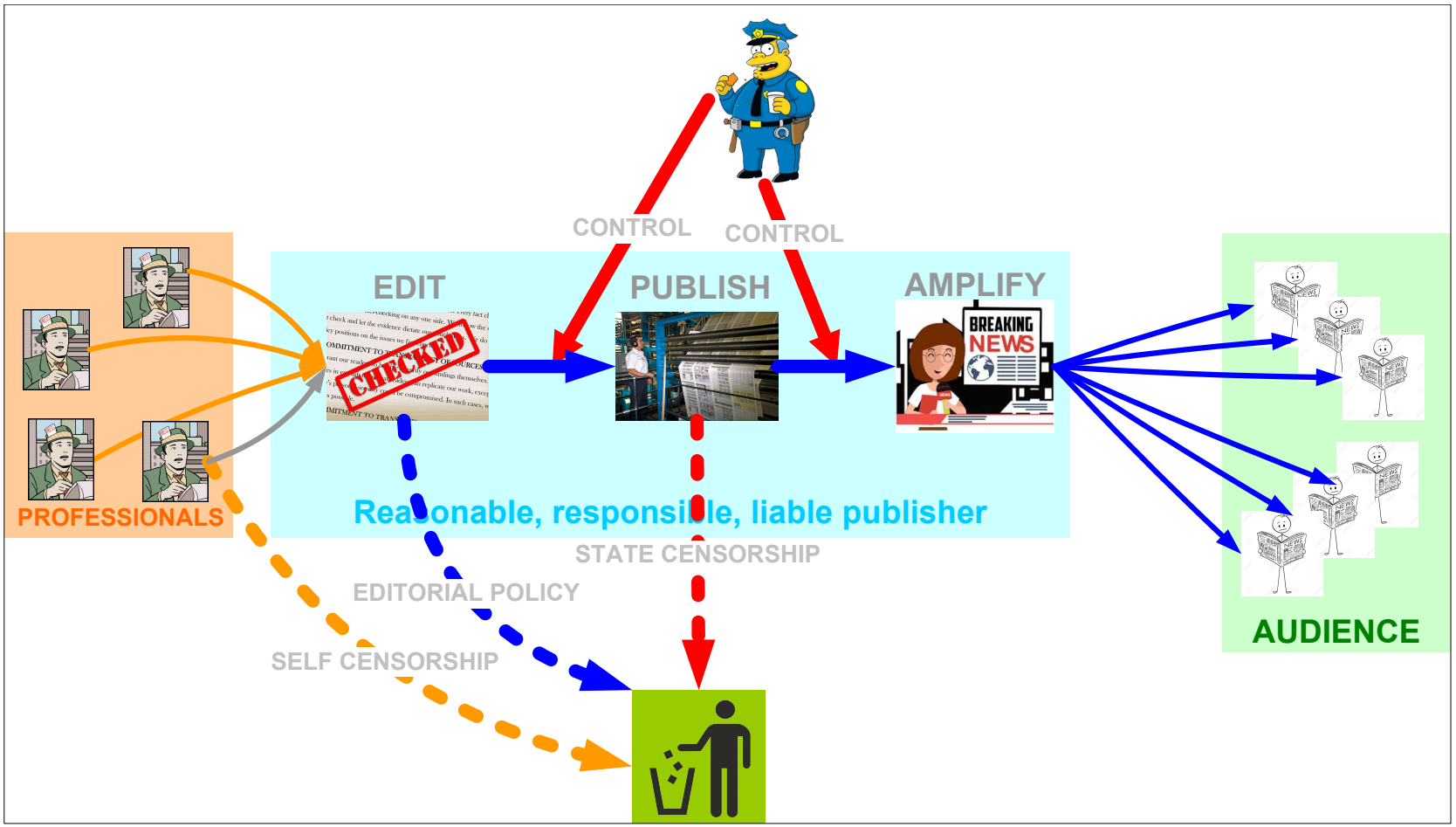
...η παραδοσιακή γραμμή (σχηματικά)...



Producer – Consumer Model

Παραγωγή και Διάδοση Ειδήσεων

...η παραδοσιακή γραμμή (σχηματικά)...



Producer – Consumer Model

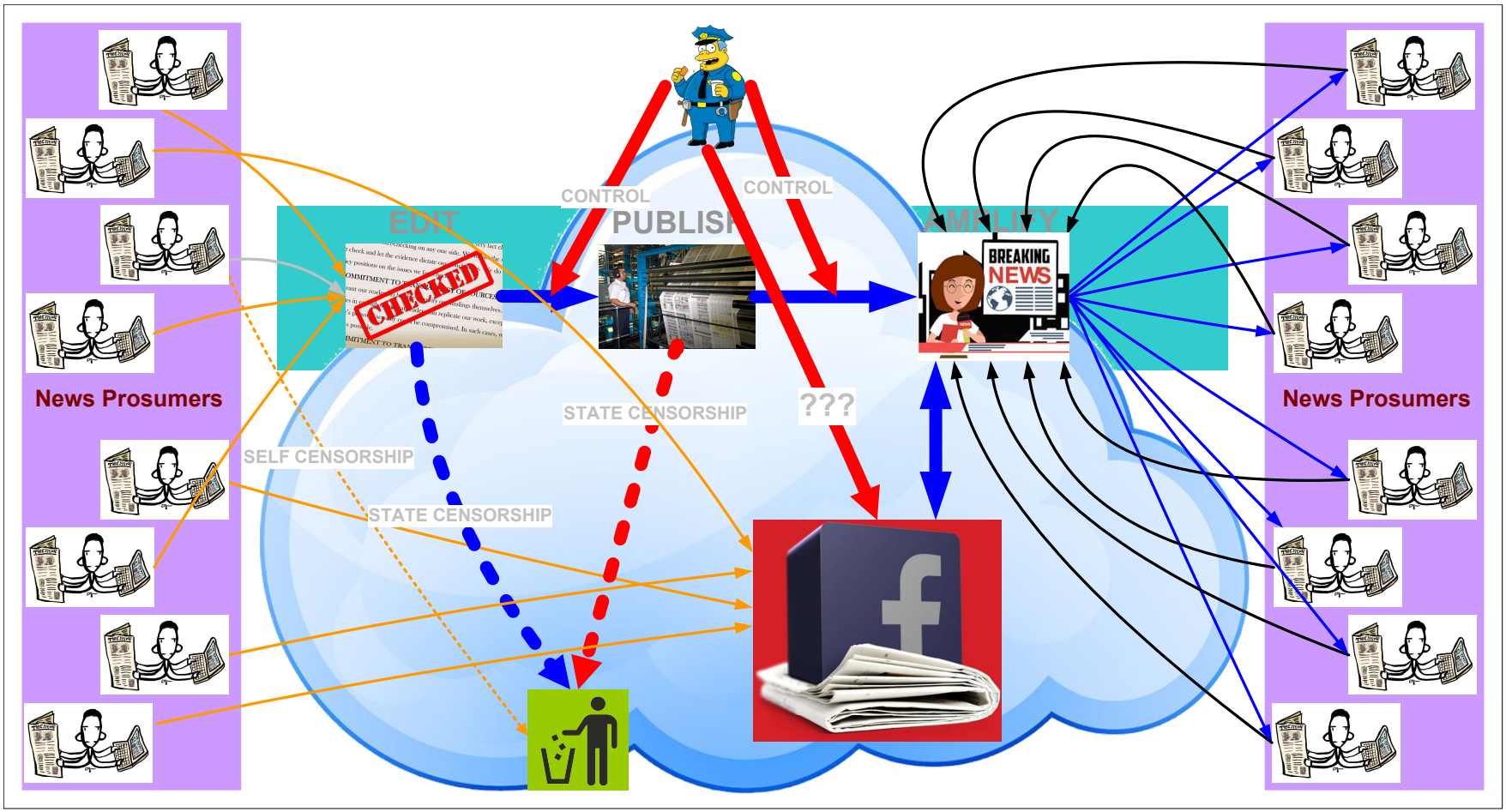
Η Ειδησεογραφία στο Διαδίκτυο

...η σύγχρονη γραμμή...

- Οποιοσδήποτε μπορεί να «δημοσιεύσει» τη γνώμη του.
- Οποιοσδήποτε μπορεί να υπερτονίσει αυτά που θεωρεί σημαντικά.
- Ο ρόλος των συντακτών ακυρώνεται στην πράξη.
- Ο «κάδος απορριμάτων» δεν προφταίνει να «ανακυκλώνει».

Παραγωγή και Διάδοση Ειδήσεων

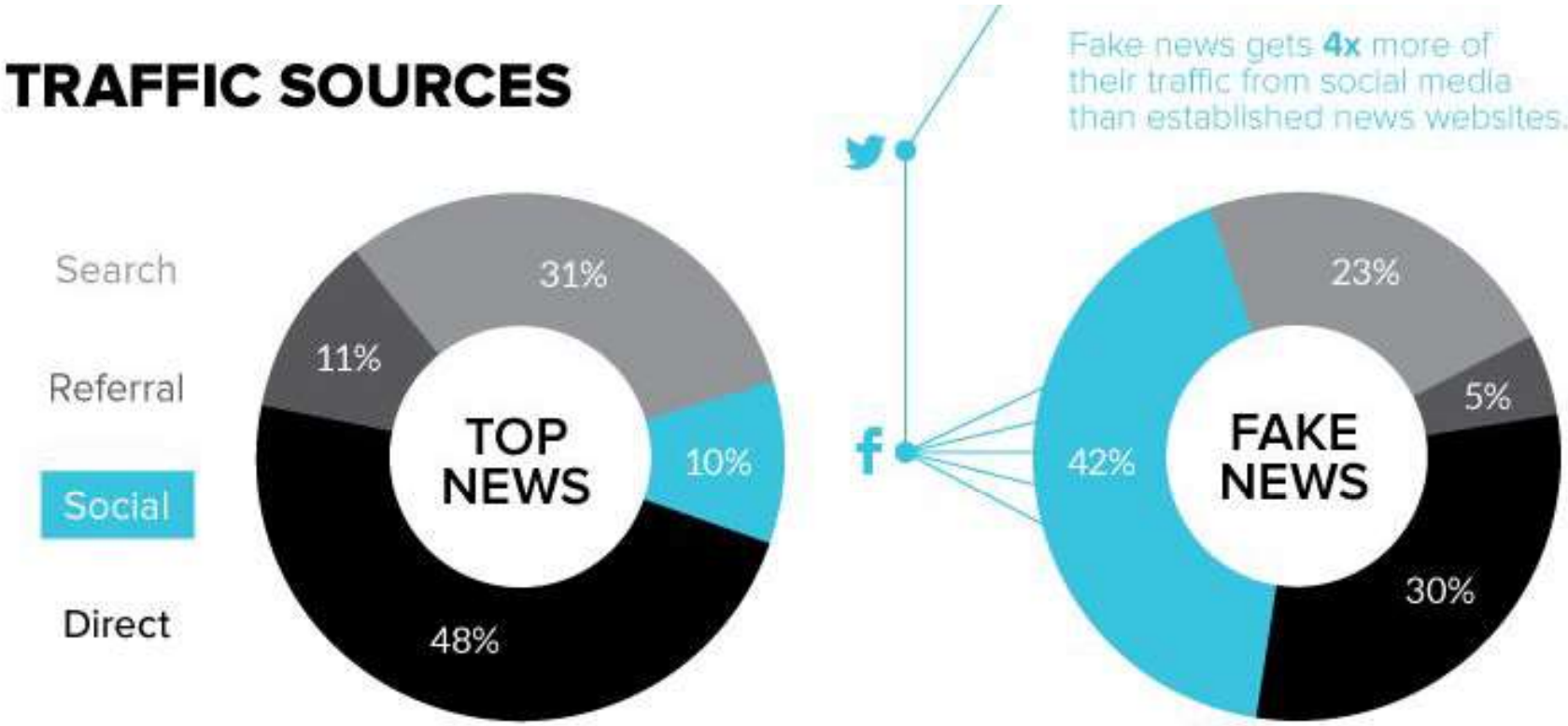
...η σύγχρονη γραμμή (σχηματικά)...



Prosumer Model

Διασπορά Ψευδών Ειδήσεων

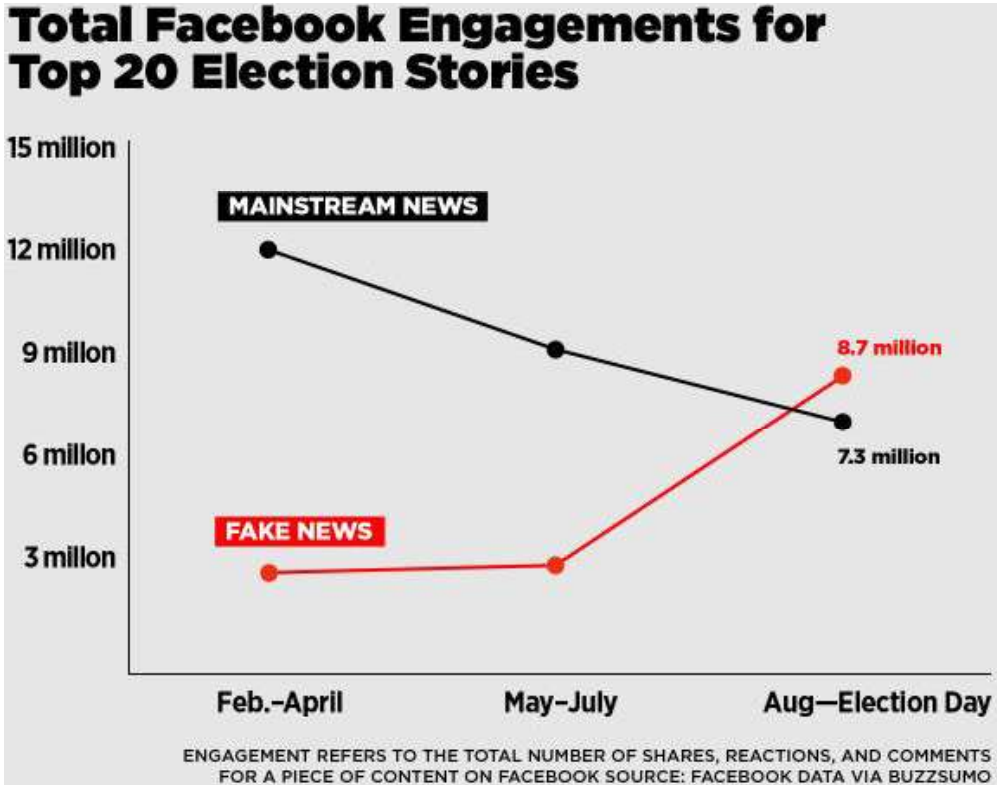
...ο ρόλος των κοινωνικών δικτύων...



- **10%** των αναγνωστών των **κορυφαίων ειδήσεων** προέρχονται από τα κοινωνικά δίκτυα.
- **40%** των αναγνωστών **«ψευδών» ειδήσεων** προέρχονται από τα κοινωνικά δίκτυα.

Διασπορά Ψευδών Ειδήσεων

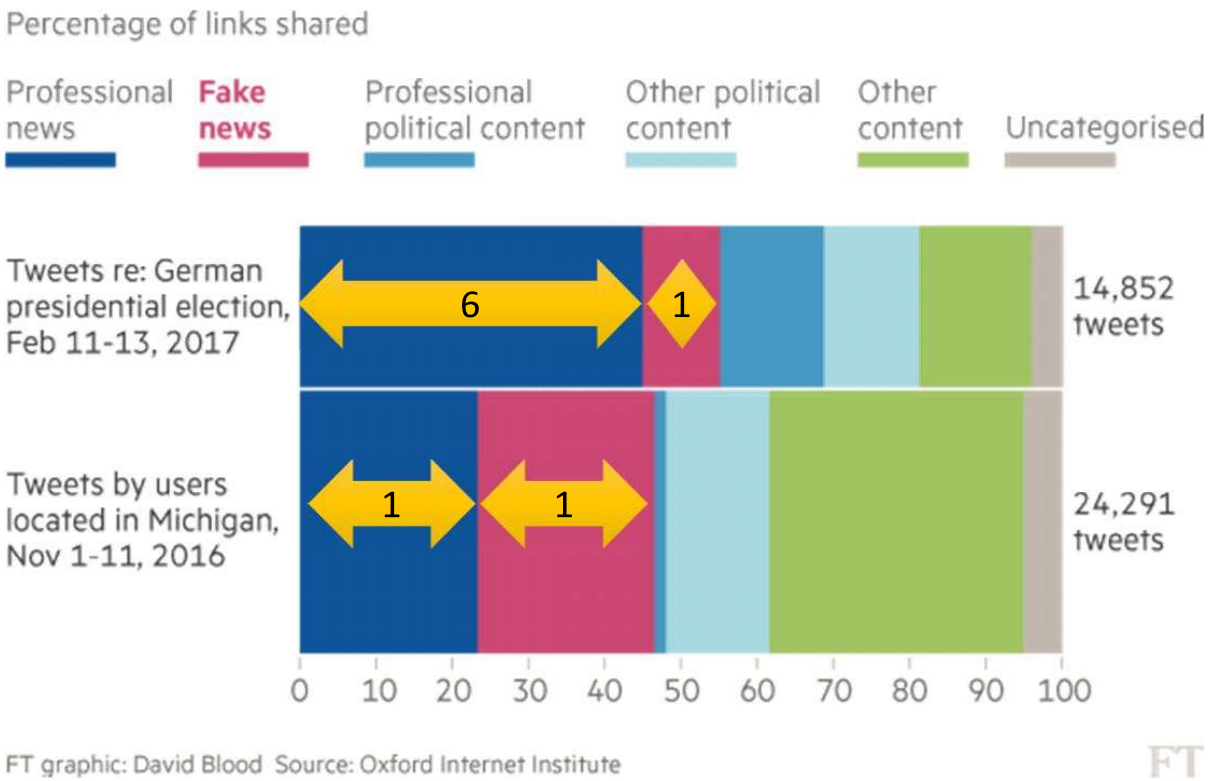
...πραγματικές και ψευδείς ειδήσεις στα κοινωνικά δίκτυα...



- Στις **αμερικανικές εκλογές του 2016** ο όγκος των σχετικών ειδήσεων στο FaceBook ήταν ισομοιρασμένος μεταξύ πραγματικών και ψευδών ειδήσεων, αλλά:
  - **10%** των **πραγματικών ειδήσεων** διακινήθηκαν μέσω FB.
  - **40%** των **ψευδών ειδήσεων** διακινήθηκαν μέσω FB.

Διασπορά Ψευδών Ειδήσεων

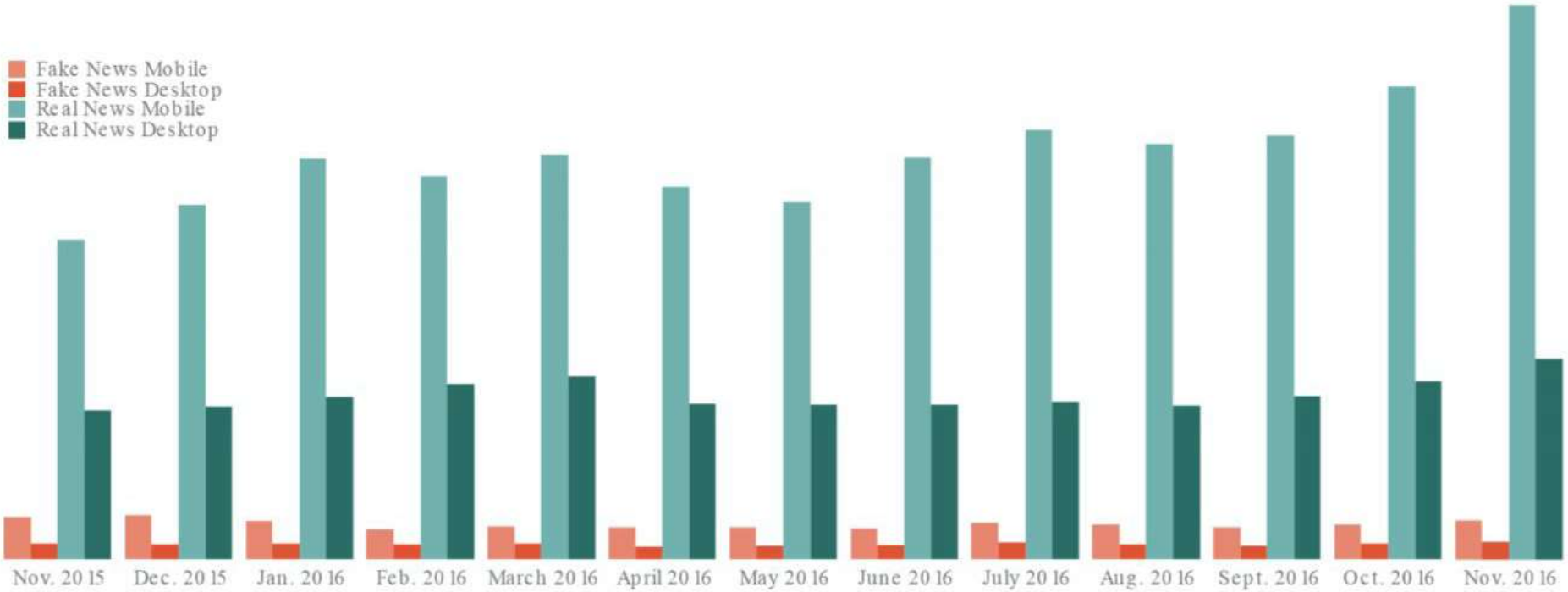
...η σημασία των διαφορετικών κουλτούρων...



- Έρευνα Πανεπιστημίου της Οξφόρδης: «Διαφορετικές αναλογίες ψευδών ειδήσεων προς πραγματικές ειδήσεις στο Twitter, ανάλογα με το κοινωνικό περιβάλλον».
  - Ευρώπη: 6:1
  - Αμερική: 1:1

Διασπορά Ψευδών Ειδήσεων

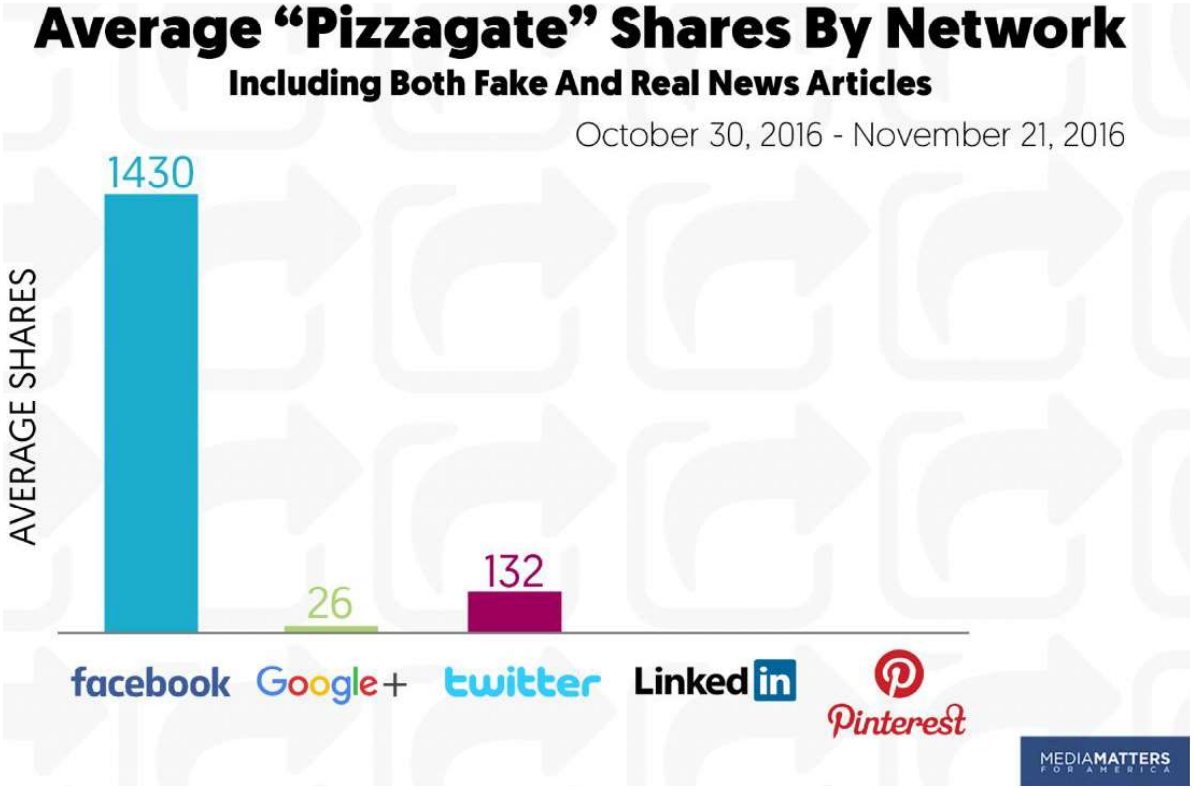
...ο ρόλος των συσκευών ανάγνωσης ειδήσεων...



- Columbia Journalism Review: Το ακροατήριο «ψευδών» ειδήσεων είναι το ίδιο (10%), ανεξάρτητα από το είδος της συσκευής.
- Κυριαρχία των κινητών τηλεφώνων.

Διασπορά Ψευδών Ειδήσεων

...κυριαρχία του Facebook στις ψευδείς ειδήσεις...

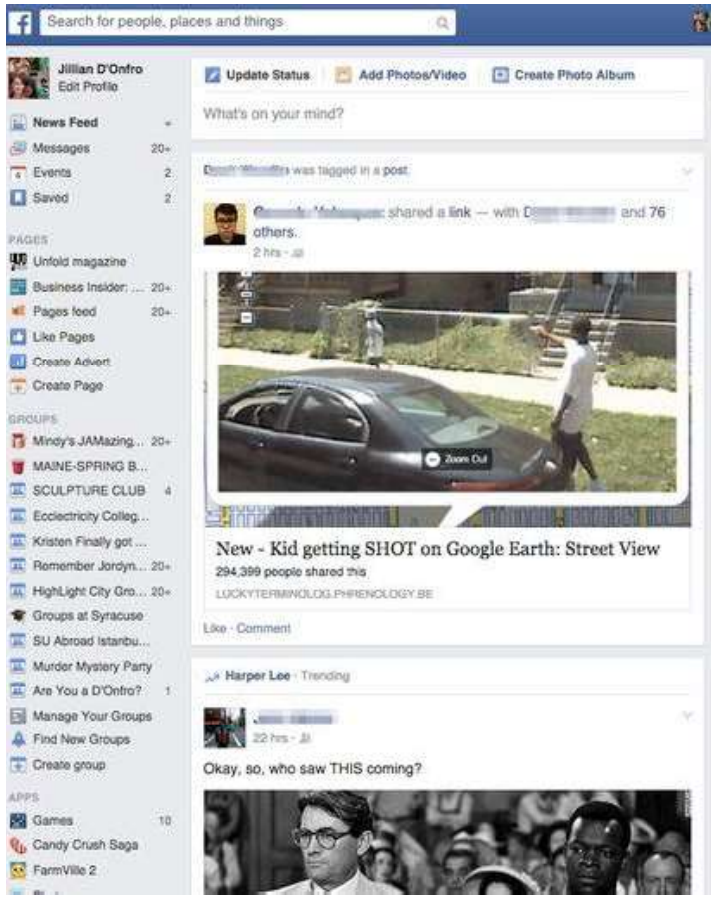


- Κύρια πηγή για την παραγωγή και διάδοση ψευδών ειδήσεων είναι (μακράν) το Facebook!!!
- Ακολουθεί το Twitter.

Διασπορά Ψευδών Ειδήσεων

...η αντίδραση του Facebook στις ψευδείς ειδήσεις...

- Στο Facebook οι χρήστες βλέπουν **μόλις το 10%** του περιεχομένου στο οποίο «εγγράφονται»!!!
- Το Facebook αποφασίζει **ΠΟΙΟ 10%** είναι αυτό.
- Το Facebook επέλεξε να **αφαιρέσει εντελώς** τις ειδήσεις από τη ροή περιεχομένου (feeds), βλάπτοντας:
  - και το 10% των αναγνωστών πραγματικών ειδήσεων!
  - και το 40% των αναγνωστών ψευδών ειδήσεων!



- Και οι δυο κοινότητες δυσαρεστημένες!!!

Διασπορά Ψευδών Ειδήσεων

...ο αλγόριθμος του Facebook για τις ψευδείς ειδήσεις...

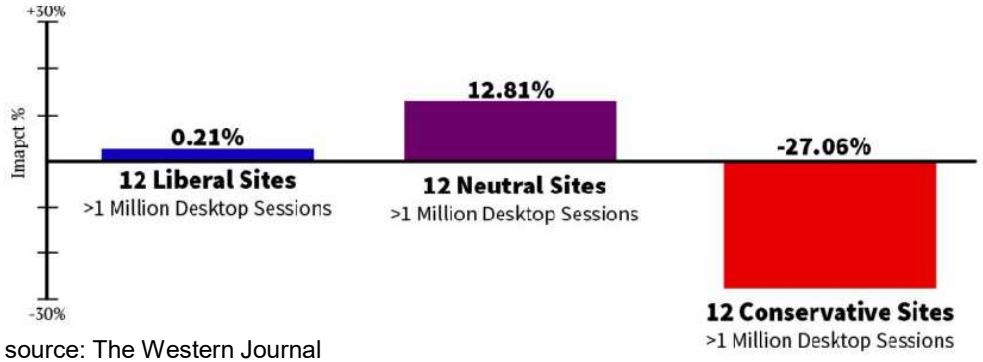
- ΣΤΟΧΟΣ: «Διατήρηση ακροατηρίου σε επαφή με την πλατφόρμα και ενεργό».

- Ο αλγόριθμος δε δημοσιοποιείται και μεταβάλλεται συνεχώς.

- Είναι «δίκαιος» αλγόριθμος;

Facebook's News Feed Algorithm:

Average Impact of FB's Algorithm for News Feeds Selection:



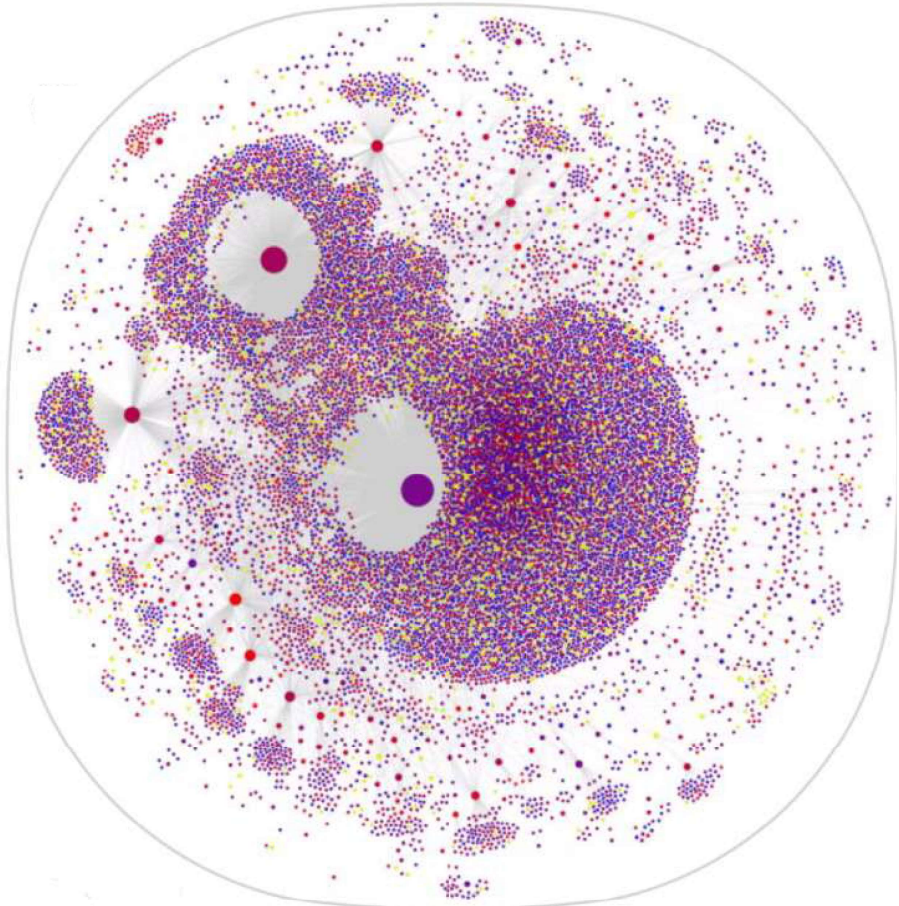
source: The Western Journal

Διασπορά Ψευδών Ειδήσεων

...ο ρόλος των bots στο Twitter...

Ασαφής Εικόνα: Τα bots επιταχύνουν την εξάπλωση τόσο πραγματικών όσο και ψευδών ειδήσεων.

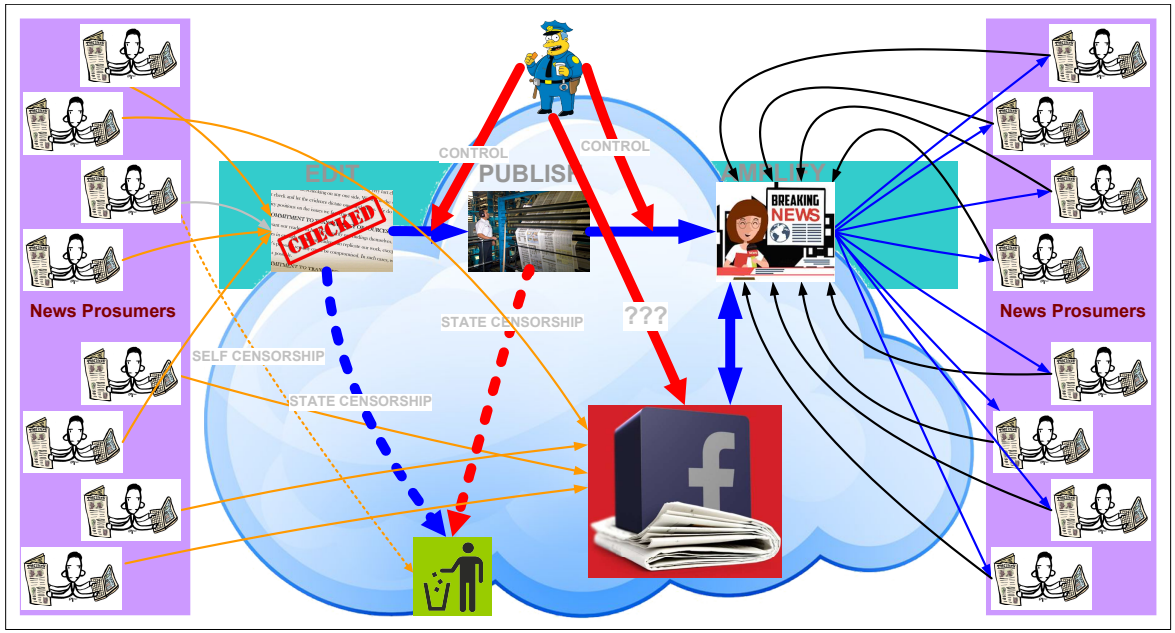
- «...social bots play a **disproportionate role** in spreading and repeating misinformation...»
- «...robots accelerated the spread of true and false news **at the same rate**, implying that false news spreads more than the truth because humans, not robots, are more likely to spread it...»



Κόμβοι	Λογαριασμοί στο Twitter.
Ακμές	Αντιδράσεις (retweets) μιας είδησης.
Μέγεθος Κόμβου	Επιρροή (πλήθος αναμετάδοσεων ιστοριών) του κόμβου.
Χρώμα Κόμβου	Βαθμός αξιοπιστίας (**μπλε: άνθρωπος**, **κόκκινο: bot**).

Παρεμπόδιση Ψευδών Ειδήσεων

...μπορεί να γίνει κάτι?...



- NAI, με τη βοήθεια:
  - του συντάκτη
  - του εκδότη
  - της πλατφόρμας κοινωνικής δικτύωσης
  - του αναμεταδότη
  - του αναγνώστη

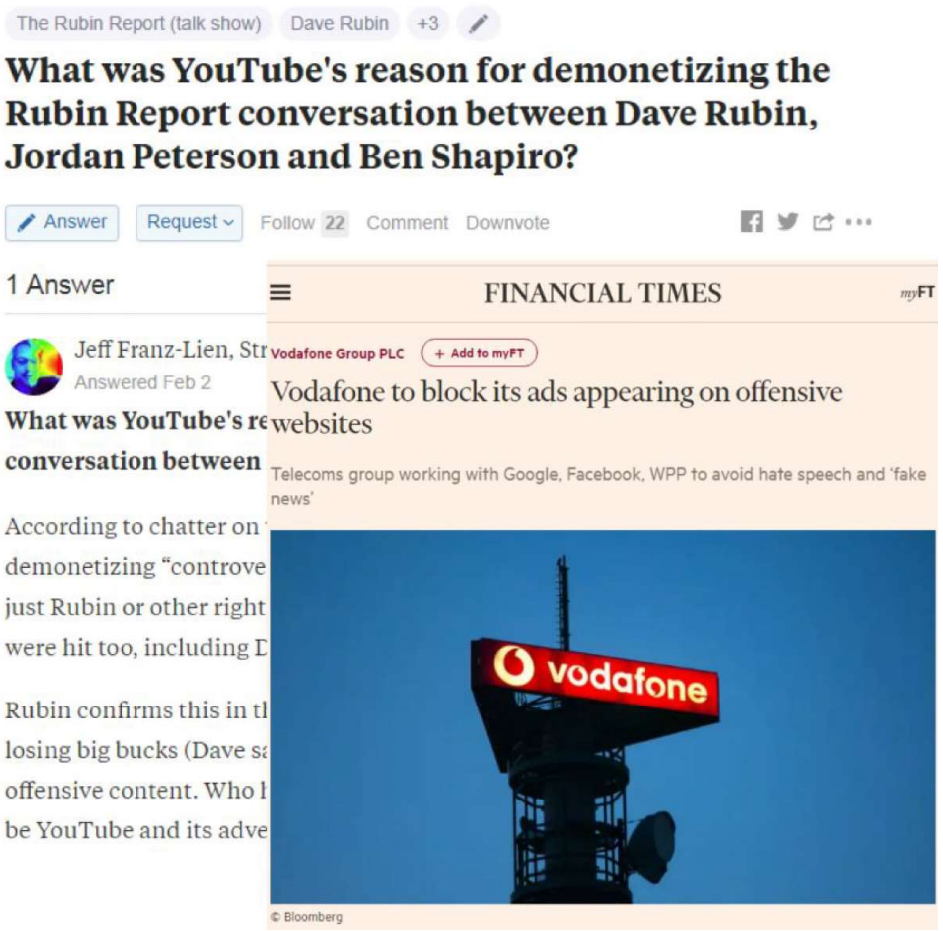
Παρεμπόδιση Ψευδών Ειδήσεων

...αποθάρρυνση συντακτών ψευδών ειδήσεων...

- Ανάκληση οικονομικών κινήτρων
  Άρνηση παροχής διαφημίσεων σε ιστοθεσίες διασποράς ψευδών ειδήσεων

- Ποινικές επιπτώσεις
  Πολωνική νομοθεσία για τοθετήσεις ενάντια στο Ολοκαύτωμα

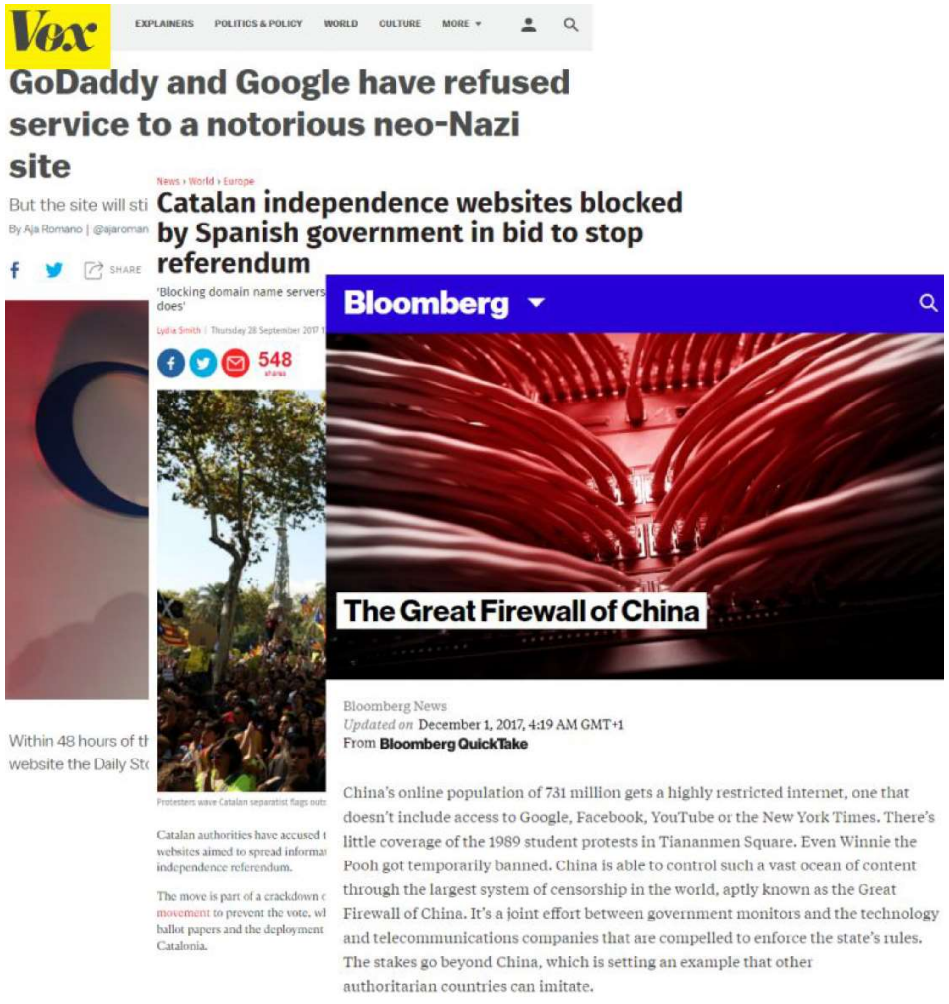
- Φυλάκιση
  Πχ, Τουρκία



Παρεμπόδιση Ψευδών Ειδήσεων

...αποθάρρυνση εκδοτών ψευδών ειδήσεων...

- Άρνηση φιλοξενίας ιστοθεσιών.
- Άρνηση καταχώρισης σε καταλόγους διευθυνσιοδότησης (DNS).
- Μπλοκάρισμα δικτυακού φόρτου προς ιστοθεσίες «ψευδών» ειδήσεων.



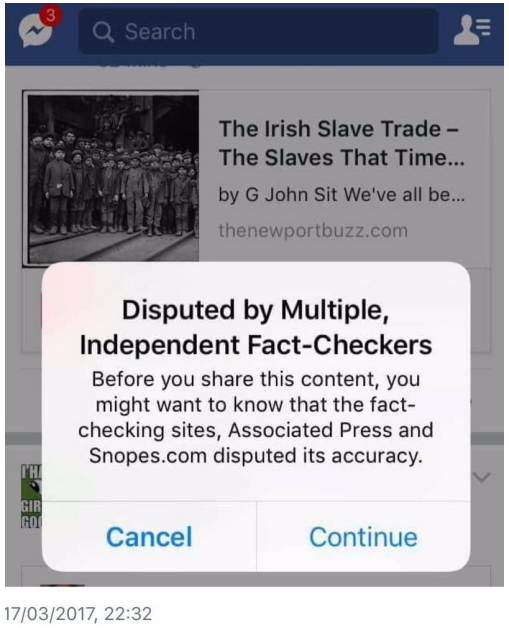
Παρεμπόδιση Ψευδών Ειδήσεων

...ο ρόλος των πλατφορμών...

- **Δημόσιες πλατφόρμες κοινωνικής δικτύωσης** (Facebook, Google, Twitter, Baidu, ...)
  Για κάθε ψευδή είδηση:
  - Επισήμανση (πχ, παροχή προειδοποιήσεων)
  - Επαύξηση (πχ, παροχή αποδεικτικών διάψευσης)
  - Απόκρυψη (ψευδών ειδήσεων)
  - Διαγραφή (λογαριασμών κατασκευής ψευδών ειδήσεων)
- **Ιδιωτικές πλατφόρμες επικοινωνίας** (Messgnger, Viber, WeChat, Snapchat, ...)
  - Παρακολούθηση ιδιωτικών συζητήσεων (!?!)
- Ανάγκη συνεργασίας διαφορετικών πλατφορμών.

Παρεμπόδιση Ψευδών Ειδήσεων

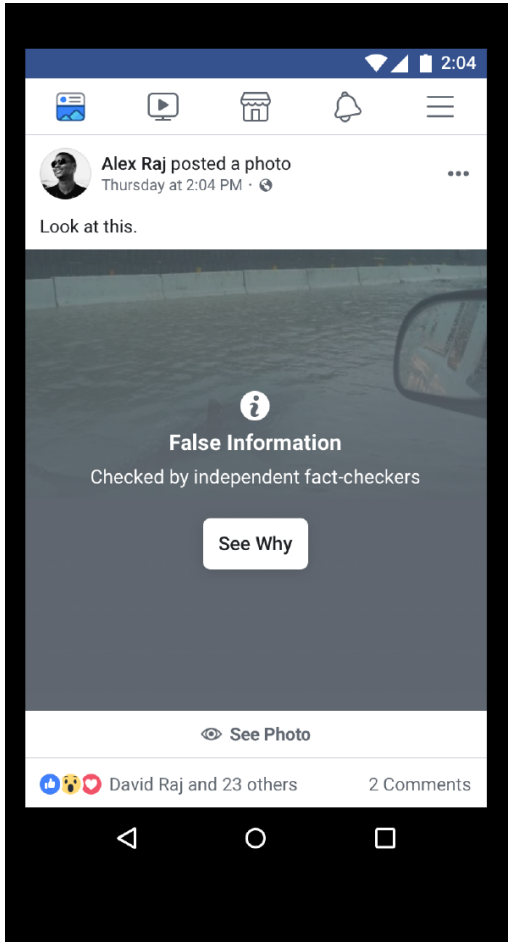
...ο ρόλος του Facebook...

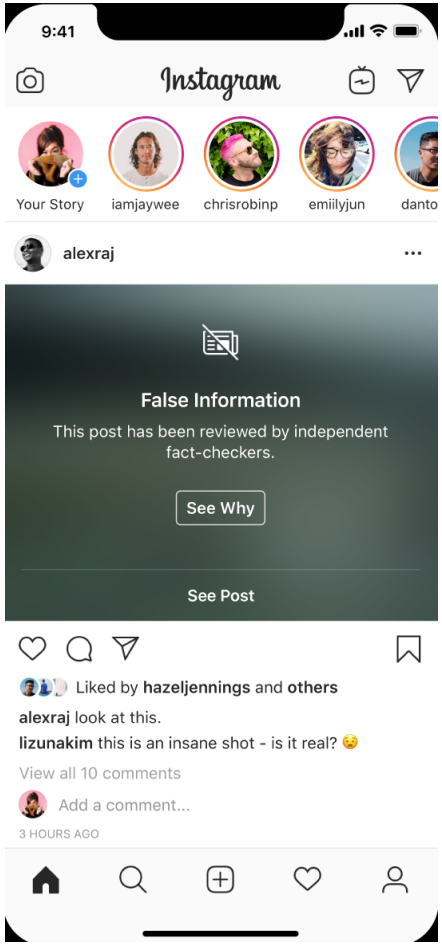


- Επισήμανση της ίδιας της πλατφόρμας, αξιοποιώντας γνωμοδοτήσεις ανεξάρτητων οργανισμών, ή έγκυρων ειδησεογραφικών πηγών.
  - https://www.snopes.com/
  - https://www.factcheck.org/
  - https://www.politifact.com/

Παρεμπόδιση Ψευδών Ειδήσεων

...ενόψει επικείμενων εκλογών στις ΗΠΑ...





- Αυστηροποίηση προειδοποιήσεων «ψευδούς είδησης» για Facebook και Instagram.

Παρεμπόδιση Ψευδών Ειδήσεων

...ο ρόλος του Twitter...

- **Επισήμανση** μέσω άμεσου διαλόγου μεταξύ των χρηστών (έμμεση παροχή προειδοποιήσεων)

Παρεμπόδιση Ψευδών Ειδήσεων

...ο ρόλος του Twitter...

- **Απόκρυψη** «ψευδών ειδήσεων», μέσω:
  - **downranking:** αμφισβήτηση σημαντικότητας «ειδήσεων» με συνδρομή εξωτερικών παρατηρητών.
  - **downscoring:** απόκρυψη ψευδών ειδήσεων από μηχανές αναζήτησης και προβολής.

Σκελετός Ομιλίας

1. Εισαγωγικά: Το Πρόβλημα των Ψευδών Ειδήσεων
   - Ορισμός, Ιδιότητες, Προκλήσεις, Εμπλεκόμενοι
   - Υπάρχουσες Τεχνικές Αντιμετώπισης Ψευδών Ειδήσεων
2. Μέθοδοι Ανίχνευσης Ψευδών Ειδήσεων
   - Μέθοδοι Αξιοποίησης Περιεχομένου
   - Μέθοδοι Αξιοποίησης Αναδράσεων
3. Μελέτη Περίπτωσης: Ακολουθιακό Μοντέλο Διάδοσης
   - Περιγραφή Μοντέλου
   - Ρόλος των Χρηστών
   - Ρόλος της Πλατφόρμας

Μέθοδοι Ανίχνευσης Ψευδών Ειδήσεων

- Με βάση το περιεχόμενο: Αξιοποίηση διαφοροποιήσεων στο στυλ γραφής, τη γλώσσα και το ύφος.
- Με βάση τα συμφραζόμενα: Αξιοποίηση ...
- Με βάση τα χαρακτηριστικά διάδοσης: Ρυθμός και μοτίβο διάδοσης, χρονική εξέλιξη.

Μέθοδοι Ανίχνευσης Ψευδών Ειδήσεων

...με βάση το περιεχόμενο: μη αυτόματη επιλογή χαρακτηριστικών...

*Το περιεχόμενο μιας τυπικής ψευδούς είδησης διαφέρει με **μετρήσιμο τρόπο** ως προς το περιεχόμενο μιας τυπικής αληθινής είδησης, με βάση σαφώς καθορισμένα χαρακτηριστικά του **λόγου**, της **γλώσσας** και του **συναισθήματος**.*

Μέθοδοι Ανίχνευσης Ψευδών Ειδήσεων

...με βάση το περιεχόμενο: μη αυτόματη επιλογή χαρακτηριστικών...

*Το περιεχόμενο μιας τυπικής ψευδούς είδησης διαφέρει με **μετρήσιμο τρόπο** ως προς το περιεχόμενο μιας τυπικής αληθινής είδησης, με βάση σαφώς καθορισμένα χαρακτηριστικά του **λόγου**, της **γλώσσας** και του **συναισθήματος**.*

- Χαρακτηριστικά: Χρήση πρώτου / τρίτου προσώπου, παρελθοντικού / ενεστώτα χρόνου, ενεργητικής / παθητικής φωνής. Συχνότητα ρημάτων, αντωνυμιών. Μήκος και σύνταξη προτάσεων. Λεκτική ποικιλότητα.
  - **S**cientific **C**ontent **AN**alysis, **L**inguistic-**B**ased **C**ue Set.

Μέθοδοι Ανίχνευσης Ψευδών Ειδήσεων

...με βάση το περιεχόμενο: μη αυτόματη επιλογή χαρακτηριστικών...

*Το περιεχόμενο μιας τυπικής ψευδούς είδησης διαφέρει με **μετρήσιμο τρόπο** ως προς το περιεχόμενο μιας τυπικής αληθινής είδησης, με βάση σαφώς καθορισμένα χαρακτηριστικά του **λόγου**, της **γλώσσας** και του **συναισθήματος**.*

- Χαρακτηριστικά: Χρήση πρώτου / τρίτου προσώπου, παρελθοντικού / ενεστώτα χρόνου, ενεργητικής / παθητικής φωνής. Συχνότητα ρημάτων, αντωνυμιών. Μήκος και σύνταξη προτάσεων. Λεκτική ποικιλότητα.
  - **S**cientific **C**ontent **AN**alysis, **L**inguistic-**B**ased **C**ue Set.
- Κατηγοριοποιητές: Πρώτα εκπαιδεύονται και μετά διακρίνουν ψευδείς από αληθινές ειδήσεις.
  - Νευρωνικά δίκτυα, δένδρα απόφασης, παλινδρόμηση.

Μέθοδοι Ανίχνευσης Ψευδών Ειδήσεων

...με βάση το περιεχόμενο: μη αυτόματη επιλογή χαρακτηριστικών...

*Το περιεχόμενο μιας τυπικής ψευδούς είδησης διαφέρει με **μετρήσιμο τρόπο** ως προς το περιεχόμενο μιας τυπικής αληθινής είδησης, με βάση σαφώς καθορισμένα χαρακτηριστικά του **λόγου**, της **γλώσσας** και του **συναισθήματος**.*

- Τα χαρακτηριστικά που καθορίζονται από ανθρώπους είναι ερμηνεύσιμα και κατανοητά.

Μέθοδοι Ανίχνευσης Ψευδών Ειδήσεων

...με βάση το περιεχόμενο: μη αυτόματη επιλογή χαρακτηριστικών...

*Το περιεχόμενο μιας τυπικής ψευδούς είδησης διαφέρει με **μετρήσιμο τρόπο** ως προς το περιεχόμενο μιας τυπικής αληθινής είδησης, με βάση σαφώς καθορισμένα χαρακτηριστικά του **λόγου**, της **γλώσσας** και του **συναισθήματος**.*

- Τα χαρακτηριστικά που καθορίζονται από ανθρώπους είναι ερμηνεύσιμα και κατανοητά.
- Ελάχιστη ευελιξία χαρακτηριστικών ως προς διαφορετικά είδη περιεχομένου ή/και γλώσσες.

Μέθοδοι Ανίχνευσης Ψευδών Ειδήσεων

...με βάση το περιεχόμενο: μη αυτόματη επιλογή χαρακτηριστικών...

*Το περιεχόμενο μιας τυπικής ψευδούς είδησης διαφέρει με **μετρήσιμο τρόπο** ως προς το περιεχόμενο μιας τυπικής αληθινής είδησης, με βάση σαφώς καθορισμένα χαρακτηριστικά του **λόγου**, της **γλώσσας** και του **συναισθήματος**.*

- Τα χαρακτηριστικά που καθορίζονται από ανθρώπους είναι ερμηνεύσιμα και κατανοητά.
- Ελάχιστη ευελιξία χαρακτηριστικών ως προς διαφορετικά είδη περιεχομένου ή/και γλώσσες.
- Αναπόφευκτη η εμπλοκή του ανθρώπινου παράγοντα.

Μέθοδοι Ανίχνευσης Ψευδών Ειδήσεων

...με βάση το περιεχόμενο: γλωσσολογική ανάλυση...

*Το περιεχόμενο μιας τυπικής ψευδούς είδησης διαφέρει με **μετρήσιμο τρόπο** ως προς το περιεχόμενο μιας τυπικής αληθινής είδησης, με βάση σαφώς καθορισμένα χαρακτηριστικά του **λόγου**, της **γλώσσας** και του **συναισθήματος**.*

- **N-grams** (ή Shingles): Αναζήτηση συχνά εμφανιζόμενων **φράσεων** σε ψευδείς ειδήσεις.

Μέθοδοι Ανίχνευσης Ψευδών Ειδήσεων

...με βάση το περιεχόμενο: γλωσσολογική ανάλυση...

*Το περιεχόμενο μιας τυπικής ψευδούς είδησης διαφέρει με **μετρήσιμο τρόπο** ως προς το περιεχόμενο μιας τυπικής αληθινής είδησης, με βάση σαφώς καθορισμένα χαρακτηριστικά του **λόγου**, της **γλώσσας** και του **συναισθήματος**.*

- **N-grams** (ή Shingles): Αναζήτηση συχνά εμφανιζόμενων **φράσεων** σε ψευδείς ειδήσεις.
- **P**art-**O**f-**S**peech Tags: Μελέτη συχνότητας διαφορετικών **μερών του λόγου** στο κείμενο.

Μέθοδοι Ανίχνευσης Ψευδών Ειδήσεων

...με βάση το περιεχόμενο: γλωσσολογική ανάλυση...

*Το περιεχόμενο μιας τυπικής ψευδούς είδησης διαφέρει με **μετρήσιμο τρόπο** ως προς το περιεχόμενο μιας τυπικής αληθινής είδησης, με βάση σαφώς καθορισμένα χαρακτηριστικά του **λόγου**, της **γλώσσας** και του **συναισθήματος**.*

- **N-grams** (ή Shingles): Αναζήτηση συχνά εμφανιζόμενων **φράσεων** σε ψευδείς ειδήσεις.
- **P**art-**O**f-**S**peech Tags: Μελέτη συχνότητας διαφορετικών **μερών του λόγου** στο κείμενο.
- **P**robabilistic **C**ontext-**F**ree **G**rammars: Μελέτη συχνότητας των **συντακτικών κανόνων** παραγωγής φράσεων.

Μέθοδοι Ανίχνευσης Ψευδών Ειδήσεων

...με βάση το περιεχόμενο: γλωσσολογική ανάλυση...

*Το περιεχόμενο μιας τυπικής ψευδούς είδησης διαφέρει με **μετρήσιμο τρόπο** ως προς το περιεχόμενο μιας τυπικής αληθινής είδησης, με βάση σαφώς καθορισμένα χαρακτηριστικά του **λόγου**, της **γλώσσας** και του **συναισθήματος**.*

- **N-grams** (ή Shingles): Αναζήτηση συχνά εμφανιζόμενων **φράσεων** σε ψευδείς ειδήσεις.
- **P**art-**O**f-**S**peech Tags: Μελέτη συχνότητας διαφορετικών **μερών του λόγου** στο κείμενο.
- **P**robabilistic **C**ontext-**F**ree **G**rammars: Μελέτη συχνότητας των **συντακτικών κανόνων** παραγωγής φράσεων.
- Κατηγοριοποιητές: Support Vector Machines.

Μέθοδοι Ανίχνευσης Ψευδών Ειδήσεων

...με βάση το περιεχόμενο: γλωσσολογική ανάλυση...

*Το περιεχόμενο μιας τυπικής ψευδούς είδησης διαφέρει με **μετρήσιμο τρόπο** ως προς το περιεχόμενο μιας τυπικής αληθινής είδησης, με βάση σαφώς καθορισμένα χαρακτηριστικά του **λόγου**, της **γλώσσας** και του **συναισθήματος**.*

- Η γλωσσολογική ανάλυση ισχυρότερη από μεθόδους που βασίζονται σε διανύσματα χαρακτηριστικών.

Μέθοδοι Ανίχνευσης Ψευδών Ειδήσεων

...με βάση το περιεχόμενο: γλωσσολογική ανάλυση...

*Το περιεχόμενο μιας τυπικής ψευδούς είδησης διαφέρει με **μετρήσιμο τρόπο** ως προς το περιεχόμενο μιας τυπικής αληθινής είδησης, με βάση σαφώς καθορισμένα χαρακτηριστικά του **λόγου**, της **γλώσσας** και του **συναισθήματος**.*

- Η γλωσσολογική ανάλυση ισχυρότερη από μεθόδους που βασίζονται σε διανύσματα χαρακτηριστικών.
- Η σημειολογία (N-grams) πιο ισχυρή από τη γραμματική (POS) και τη σύνταξη (PCFG) της γλώσσας.

Μέθοδοι Ανίχνευσης Ψευδών Ειδήσεων

...με βάση το περιεχόμενο: γλωσσολογική ανάλυση...

*Το περιεχόμενο μιας τυπικής ψευδούς είδησης διαφέρει με **μετρήσιμο τρόπο** ως προς το περιεχόμενο μιας τυπικής αληθινής είδησης, με βάση σαφώς καθορισμένα χαρακτηριστικά του **λόγου**, της **γλώσσας** και του **συναισθήματος**.*

- Η γλωσσολογική ανάλυση ισχυρότερη από μεθόδους που βασίζονται σε διανύσματα χαρακτηριστικών.
- Η σημειολογία (N-grams) πιο ισχυρή από τη γραμματική (POS) και τη σύνταξη (PCFG) της γλώσσας.
- Αδυναμία αποτύπωσης της πλήρους σημασιολογίας ενός κειμένου.

Μέθοδοι Ανίχνευσης Ψευδών Ειδήσεων

...με βάση το περιεχόμενο: τεχνικές βαθιάς μάθησης...

*Το περιεχόμενο μιας τυπικής ψευδούς είδησης διαφέρει με **μετρήσιμο τρόπο** ως προς το περιεχόμενο μιας τυπικής αληθινής είδησης, με βάση σαφώς καθορισμένα χαρακτηριστικά του **λόγου**, της **γλώσσας** και του **συναισθήματος**.*

- **C**onvolutional **N**eural **N**etworks.

Μέθοδοι Ανίχνευσης Ψευδών Ειδήσεων

...με βάση το περιεχόμενο: τεχνικές βαθιάς μάθησης...

*Το περιεχόμενο μιας τυπικής ψευδούς είδησης διαφέρει με **μετρήσιμο τρόπο** ως προς το περιεχόμενο μιας τυπικής αληθινής είδησης, με βάση σαφώς καθορισμένα χαρακτηριστικά του **λόγου**, της **γλώσσας** και του **συναισθήματος**.*

- **C**onvolutional **N**eural **N**etworks.
- **R**ecurrent **N**eural **N**etworks.

Μέθοδοι Ανίχνευσης Ψευδών Ειδήσεων

...με βάση το περιεχόμενο: τεχνικές βαθιάς μάθησης...

*Το περιεχόμενο μιας τυπικής ψευδούς είδησης διαφέρει με **μετρήσιμο τρόπο** ως προς το περιεχόμενο μιας τυπικής αληθινής είδησης, με βάση σαφώς καθορισμένα χαρακτηριστικά του **λόγου**, της **γλώσσας** και του **συναισθήματος**.*

- Ακόμα και με τις πιο εξελιγμένες μορφές βαθιάς μάθησης για προσδιορισμό κρίσιμων χαρακτηριστικών της γλώσσας, η ανίχνευση ψευδών ειδήσεων είναι πολύ δύσκολη.
  - Το ψέμα είναι κατασκευασμένο ώστε να προσομοιάζει με την αλήθεια, με σκοπό την εξαπάτηση.
  - Αποκλειστικά με ανάλυση περιεχομένου σε πραγματικά σύνολα δεδομένων, ακόμη και με τεχνικές βαθιάς μάθησης, οδηγεί σε ποσοστά ανίχνευσης το πολύ έως 70%.

Μέθοδοι Ανίχνευσης Ψευδών Ειδήσεων

...με βάση τις αναδράσεις: χειρωνακτικός ορισμός χαρακτηριστικών...

*Η διαδιδόμενη είδηση έχει πολλές **μετρήσιμες μορφές** ανάδρασης (πχ, **αντιδράσεις** χρηστών, **σχολιασμός** άρθρων, **μοτίβα διάδοσης**, κ.λπ.) που παρέχουν επιπρόσθετη πληροφόρηση προς την κατεύθυνση της κατηγοριοποίησής της ως αληθινής ή ψευδούς.*

1. «Χειρωνακτικός» προσδιορισμός **χαρακτηριστικών** που σχετίζονται με ανάδραση χρηστών:
   - Προφίλ χρηστών που (ανα)μεταδίδουν την είδηση (πχ, ηλικία, αριθμός συνδέσμων, πλήθος δημοσιεύσεων).
   - Τυπικός μορφότυπος μεταδιδόμενων ειδήσεων (πχ, ποσοστό tweets που αναφέρουν συχνά άλλους χρήστες, μέσω του '@').
   - Μοτίβα χωρικής / χρονικής εξάπλωσης είδησης (πχ, βάθος δένδρου διάδοσης, χρονοσειρά αντιδράσεων σε είδηση).
2. Τροφοδότηση χαρακτηριστικών σε έναν κατηγοριοποιητή για εκπαίδευση και στη συνέχεια κατηγοριοποίηση.

Μέθοδοι Ανίχνευσης Ψευδών Ειδήσεων

...με βάση τις αναδράσεις: ανάλυση μοτίβων διάδοσης...

*Η διαδιδόμενη είδηση έχει πολλές **μετρήσιμες μορφές** ανάδρασης (πχ, **αντιδράσεις** χρηστών, **σχολιασμός** άρθρων, **μοτίβα διάδοσης**, κ.λπ.) που παρέχουν επιπρόσθετη πληροφόρηση προς την κατεύθυνση της κατηγοριοποίησής της ως αληθινής ή ψευδούς.*

- Πυρήνες δένδρου διάδοσης / γραφήματος τυχαίου περιπάτου. Κάθε κόμβος αφορά κάποια αντίδραση σε είδηση, με χρονοσφραγίδα, τύπο, κείμενο. Συνδυασμός μετρικών ομοιότητας ανάλογα με το υπό εξέταση χαρακτηριστικό.
- Αναδρομικά Νευρωνικά Δίκτυα: Συσσώρευση πληροφορίας από τη ρίζα προς τα φύλλα, ή αντίστροφα.
- Μοντελοποίηση Διεργασίας Διάδοσης: 4 κατηγορίες χρηστών S(usceptible)-E(xposed)-Infected-Z(skeptic) με συγκεκριμένους ρυθμούς μετάβασης.

Σκελετός Ομιλίας

1. Εισαγωγικά: Το Πρόβλημα των Ψευδών Ειδήσεων
   - Ορισμός, Ιδιότητες, Προκλήσεις, Εμπλεκόμενοι
   - Υπάρχουσες Τεχνικές Αντιμετώπισης Ψευδών Ειδήσεων
2. Μέθοδοι Ανίχνευσης Ψευδών Ειδήσεων
   - Μέθοδοι Αξιοποίησης Περιεχομένου
   - Μέθοδοι Αξιοποίησης Αναδράσεων
3. Μελέτη Περίπτωσης: Ακολουθιακό Μοντέλο Διάδοσης
   - Περιγραφή Μοντέλου
   - Ρόλος των Χρηστών
   - Ρόλος της Πλατφόρμας

Μελέτη Περίπτωσης

Ένα Ακολουθιακό Μοντέλο Διάδοσης [Papanastasiou (2019)]

- $\Theta \in \{Y, N\}$: Η **αντικειμενική αλήθεια** για μια είδηση (όχι άμεσα παρατηρήσιμη).
- $mesg \in \{y, n\}$: Η **υποκειμενική θέση** της είδησης.
- $v \in \{T(rue), F(ake)\}$: Ο **τύπος** της είδησης.
- Κάθε είδηση παράγεται από το εξής πιθανοτικό μοντέλο, όπου $a \in (0.5, 1)$ και $p_y \in [0, 1]$:

$\mathbb{P}[\ mesg \mid \Theta\ ]$	*truthful*		*fake*	
	$\Theta = Y$	$\Theta = N$	$\Theta = Y$	$\Theta = N$
$mesg = y$	a	$1 - a$	p_y	p_y
$mesg = n$	$1 - a$	a	$1 - p_y$	$1 - p_y$

- Αρχικές πεποιθήσεις ενός χρήστη $i \geq 1$:
  - για αντικειμενική αλήθεια μιας είδησης: $\boxed{b_{i,0} = \mathbb{P}[\ \Theta = Y\ ]}$
  - για τύπο μηνύματος: $\boxed{q_0 = \mathbb{P}[\ v = F\ ]}$

Μελέτη Περίπτωσης

Ένα Ακολουθιακό Μοντέλο Διάδοσης [Papanastasiou (2019)]

- Επικαιροποίηση αρχικών πεποιθήσεων.

$$
\begin{aligned}
b_{i,1} &= \mathbb{P}[\ \Theta = Y \mid m = y\ ] \\
&= \mathbb{P}[\ m = y \mid \Theta = Y\ ] \cdot \frac{\mathbb{P}[\Theta = Y]}{\mathbb{P}[m = y]} \\
&= \frac{a \cdot b_{i,0}}{\mathbb{P}[m=y \mid v=T] \cdot \mathbb{P}[v=T] + \mathbb{P}[m=y \mid v=F] \cdot \mathbb{P}[v=F]} \\
&= \frac{a \cdot b_{i,0}}{\mathbb{P}[m=y \mid v=T] \cdot (1-q_0) + \mathbb{P}[m=y \mid v=F] \cdot q_0}
\end{aligned}
$$

  - Για $q_0 = 0$: $\boxed{b_{i,1} = \frac{a \cdot b_{i,0}}{\mathbb{P}[m=y \mid v=T]} = \frac{a b_{i,0}}{a b_{i,0} + (1-a)(1-b_{i,0})} > b_{i,0}}$
  - Για $q_0 = 1$: $\boxed{b_{i,1} = \frac{a \cdot b_{i,0}}{\mathbb{P}[m=y \mid v=F]} = \frac{p_y b_{i,0}}{p_y b_{i,0} + p_y (1-b_{i,0})} = b_{i,0}}$

Μελέτη Περίπτωσης

Ένα Ακολουθιακό Μοντέλο Διάδοσης [Papanastasiou (2019)]

- Περιγραφή ακολουθίας διάδοσης μηνύματος:
  - Άπειρο πλήθος χρηστών, οργανωμένοι σε κοινωνικό δίκτυο με τη δομή ΜΟΝΟΠΑΤΙΟΥ.
  - $\forall i \geq 1$, χρήστης i που παραλαμβάνει το μήνυμα τη στιγμή i, έχει τρεις επιλογές:

    CHECK Να ελέγξει το μήνυμα για την αλήθεια του. Εφόσον διαπιστωθεί ότι είναι αληθινό μήνυμα, το μεταδίδει στον χρήστη $i+1$. Διαφορετικά, τερματίζει τη μετάδοση του μηνύματος.

    SEND Να μεταδώσει το μήνυμα στον χρήστη $i+1$, δίχως να ελέγξει για την αλήθεια του.

    KILL Να τερματίσει τη διάδοση του μηνύματος, δίχως να ελέγξει για την αλήθεια του.

Μελέτη Περίπτωσης

Ένα Ακολουθιακό Μοντέλο Διάδοσης [Papanastasiou (2019)]

- **Συνάρτηση Ωφέλειας** του χρήστη $i \geq 1$ $(0 < K < 0.5)$:
  $\forall v \in \{T, F\}, \forall a_i \in \{CHECK, SEND, KILL\}$,

$$U_i^v(a_i) = \begin{cases} 1, & (v = T, a_i = SEND) \vee (v = F, a_i = KILL) \\ 0, & (v = T, a_i = KILL) \vee (v = F, a_i = SEND) \\ 1 - K, & a_i = CHECK \end{cases}$$

- $S_i = \mathbb{P}[a_i = SEND]$, $C_i = \mathbb{P}[a_i = CHECK]$, $N_i = \mathbb{P}[a_i = KILL]$
- Επικαιροποίηση πεποιθήσεων για τύπο μηνύματος: $\forall i \geq 1$,

$$q_i = \mathbb{P}[\ v = F \mid H_i\ ,\ m = y\ ]$$

L1 $\forall i \geq 1$, $\boxed{q_i = \frac{p_y q_0 w_i}{p_y q_0 w_i + [a b_{i,0} + (1-a)(1-b_{i,0})](1-q_0)}}$ όπου:

$$w_1 = 1,\ w_i = w_{i-1} \cdot \frac{S_{i-1}}{S_{i-1} + C_{i-1}} = \frac{\prod_{k=1}^{i-1} S_k}{\prod_{k=1}^{i-1} (S_k + C_k)}$$

Μελέτη Περίπτωσης

Ένα Ακολουθιακό Μοντέλο Διάδοσης [Papanastasiou (2019)]

- Λελογισμένη συμπεριφορά χρηστών: Επιλέγεται από τον $i \geq 1$ η δράση a_i που μεγιστοποιεί το **αναμενόμενο προσωπικό όφελος** $\mathbb{E}[U_i^v(a_i)]$, με βάση την **τρέχουσα πεποίθηση** q_i του i για τον τύπο $v \in \{T, F\}$ του μηνύματος.

Μελέτη Περίπτωσης

Ένα Ακολουθιακό Μοντέλο Διάδοσης [Papanastasiou (2019)]

- Λελογισμένη συμπεριφορά χρηστών: Επιλέγεται από τον $i \geq 1$ η δράση a_i που μεγιστοποιεί το **αναμενόμενο προσωπικό όφελος** $\mathbb{E}[U_i^v(a_i)]$, με βάση την **τρέχουσα πεποίθηση** q_i του i για τον τύπο $v \in \{T, F\}$ του μηνύματος.

P1 $\forall i \geq 1$, η βέλτιστη δράση για τον i με βάση την *τρέχουσα πεποίθηση* q_i, είναι η εξής:

$$a_i = \begin{cases} KILL, & b_{i,0} \leq Z_i^{low} \\ CHECK, & Z_i^{low} < b_{i,0} \leq Z_i^{high} \\ SEND, & Z_i^{high} < b_{i,0} \end{cases}$$

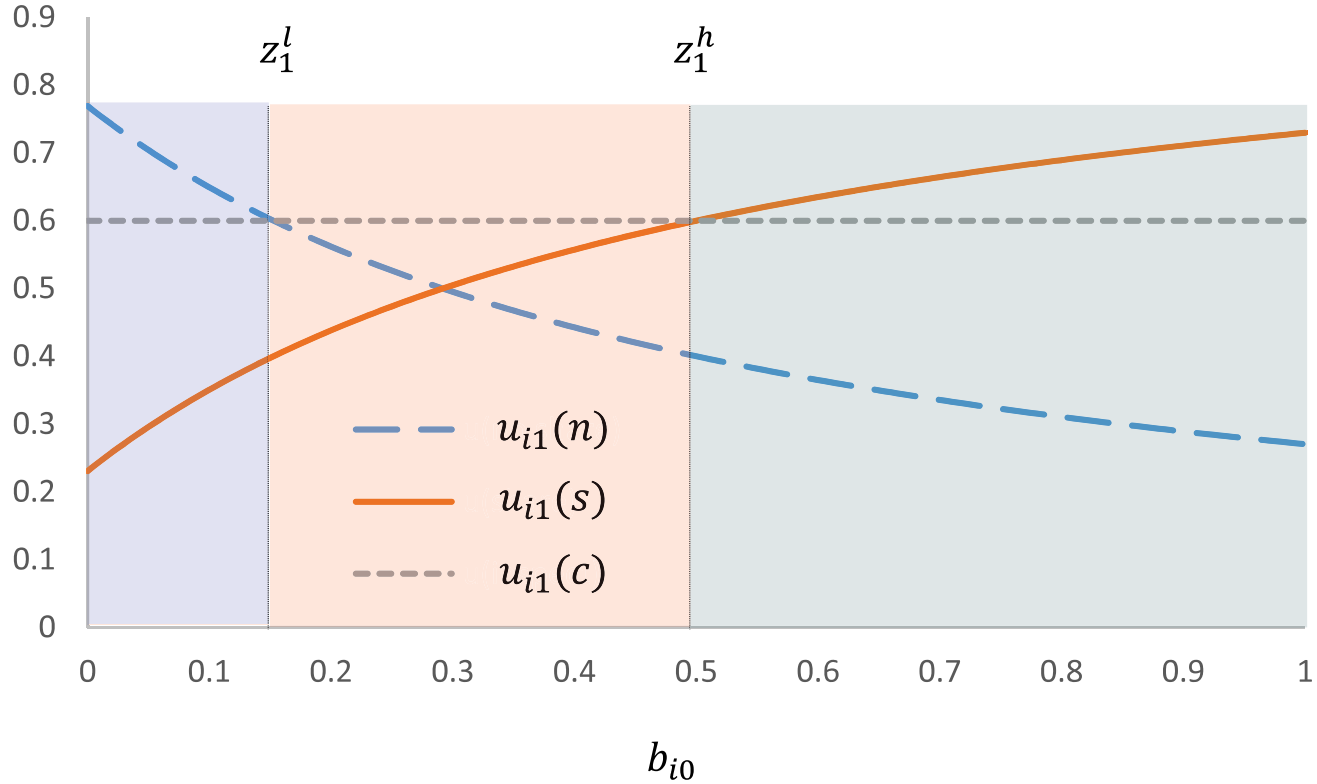
όπου:

- $Z_i^{low} = \left[\frac{K}{1-K} p_y q_0 w_i - (1-a)(1-q_0)\right] \ / \ [(2a-1)(1-q_0)]$
- $Z_i^{high} = \left[\frac{1-K}{K} p_y q_0 w_i - (1-a)(1-q_0)\right] \ / \ [(2a-1)(1-q_0)]$

Μελέτη Περίπτωσης

Ένα Ακολουθιακό Μοντέλο Διάδοσης [Papanastasiou (2019)]

- Λελογισμένη συμπεριφορά χρηστών: Επιλέγεται από τον $i \geq 1$ η δράση a_i που μεγιστοποιεί το **αναμενόμενο προσωπικό όφελος** $\mathbb{E}[U_i^v(a_i)]$, με βάση την **τρέχουσα πεποίθηση** q_i του i για τον τύπο $v \in \{T, F\}$ του μηνύματος.



Μελέτη Περίπτωσης

Ένα Ακολουθιακό Μοντέλο Διάδοσης [Papanastasiou (2019)]

L2 $\forall i \geq 1$, η **τρέχουσα πεποίθηση** q_i ότι το μήνυμα είναι FAKE είναι *φθίνουσα* στην τιμή του (χρόνου) i.

Μελέτη Περίπτωσης

Ένα Ακολουθιακό Μοντέλο Διάδοσης [Papanastasiou (2019)]

L2 $\forall i \geq 1$, η **τρέχουσα πεποίθηση** q_i ότι το μήνυμα είναι FAKE είναι *φθίνουσα* στην τιμή του (χρόνου) i.

R Οι τιμές των **κατωφλίων** Z_i^{low} και Z_i^{high} είναι (αυστηρά, αν $C_i > 0$) *φθίνουσες στην τιμή του χρόνου* i, και ανεξάρτητη από τις προσωπικές πεποιθήσεις $b_{i,0}$ των χρηστών για την αντικειμενική αλήθεια της πληροφορίας.

Μελέτη Περίπτωσης

Ένα Ακολουθιακό Μοντέλο Διάδοσης [Papanastasiou (2019)]

L2 $\forall i \geq 1$, η **τρέχουσα πεποίθηση** q_i ότι το μήνυμα είναι FAKE είναι *φθίνουσα* στην τιμή του (χρόνου) i.

R Οι τιμές των **κατωφλίων** Z_i^{low} και Z_i^{high} είναι (αυστηρά, αν $C_i > 0$) *φθίνουσες στην τιμή του χρόνου* i, και ανεξάρτητη από τις προσωπικές πεποιθήσεις $b_{i,0}$ των χρηστών για την αντικειμενική αλήθεια της πληροφορίας.

P3 COMPUTABILITY OF ACTION PROBABILITIES: Έστω ότι οι αρχικές πεποιθήσεις $b_{i,0}$ κατανέμονται ανεξάρτητα, σύμφωνα με μια συνεχή και γνησίως αύξουσα συνάρτηση κατανομής πιθανότητας $F(\cdot)$, όπου $F(0) = 0,\ F(1) = 1$. Οι πιθανότητες για επιλογή δράσης από τον χρήστη $i \geq 1$ είναι πλήρως προσδιορισμένες:

$$N_i = F(Z_i^{low}) \qquad S_i = 1 - F(Z_i^{high}) \qquad C_i = 1 - S_i - N_i$$

Μελέτη Περίπτωσης

Ένα Ακολουθιακό Μοντέλο Διάδοσης [Papanastasiou (2019)]

P2 SHARING CASCADE: Υπάρχει χρονικός ορίζοντας $T_c \geq 1$ τ.ώ. $\forall i \geq T_c,\ S_i = 1$.

Μελέτη Περίπτωσης

Ένα Ακολουθιακό Μοντέλο Διάδοσης [Papanastasiou (2019)]

P2 SHARING CASCADE: Υπάρχει χρονικός ορίζοντας $T_c \geq 1$ τ.ώ. $\forall i \geq T_c,\ S_i = 1$.

- Αν το μήνυμα επιβιώσει για ένα κρίσιμο χρονικό διάστημα, τότε γίνεται viral (δλδ, συνεχίζει με αδιαμφισβήτητες διαμοιράσεις χωρίς κανέναν έλεγχο).

Μελέτη Περίπτωσης

Ένα Ακολουθιακό Μοντέλο Διάδοσης [Papanastasiou (2019)]

P2 SHARING CASCADE: Υπάρχει χρονικός ορίζοντας $T_c \geq 1$ τ.ώ. $\forall i \geq T_c,\ S_i = 1$.

- Αν το μήνυμα επιβιώσει για ένα κρίσιμο χρονικό διάστημα, τότε γίνεται viral (δλδ, συνεχίζει με αδιαμφισβήτητες διαμοιράσεις χωρίς κανέναν έλεγχο).
- Η πεποίθηση των όλων χρηστών για $i \geq T_c$ παραμένει **σταθερή** και ίση με q_{T_c}.

Μελέτη Περίπτωσης

Ένα Ακολουθιακό Μοντέλο Διάδοσης [Papanastasiou (2019)]

P2 SHARING CASCADE: Υπάρχει χρονικός ορίζοντας $T_c \geq 1$ τ.ώ. $\forall i \geq T_c,\ S_i = 1$.

- Αν το μήνυμα επιβιώσει για ένα κρίσιμο χρονικό διάστημα, τότε γίνεται viral (δλδ, συνεχίζει με αδιαμφισβήτητες διαμοιράσεις χωρίς κανέναν έλεγχο).
- Η πεποίθηση των όλων χρηστών για $i \geq T_c$ παραμένει **σταθερή** και ίση με q_{T_c}.
- Η κρίσιμη χρονική στιγμή T_c είναι συνάρτηση των παραμέτρων a, q_0, K του μοντέλου, *ανεξάρτητη από την ετερογένεια* των χρηστών (δλδ, τα $b_{i,0}$), και μπορεί εκ των προτέρων να υπολογιστεί για οποιονδήποτε συνδυασμό των παραμέτρων.

Μελέτη Περίπτωσης

Ένα Ακολουθιακό Μοντέλο Διάδοσης [Papanastasiou (2019)]

- Συμμετοχή πλατφόρμας:
  - $R > 0$ = όφελος (ανά μετάδοση) για διαμοίραση αληθινών μηνυμάτων.
  - $P > 0$ = ζημιά (ανά μετάδοση) για διαμοίραση ψεύτικων μηνυμάτων ($P > R$).
  - $K_p > 0$ = κόστος για διεξαγωγή ελέγχου μηνύματος και κοινοποίηση αποτελέσματος σε όλους.
  - $b_{p,0} = \int_0^1 b \; dF(b)$. /* η πλατφόρμα ασπάζεται τις πεποιθήσεις των χρηστών */
  - $q_{p,i} = \frac{p_y q_0 w_i}{p_y q_0 w_i + [a b_{p,0} + (1-a)(1-b_{p,0})](1-q_0)}$

Μελέτη Περίπτωσης

Ένα Ακολουθιακό Μοντέλο Διάδοσης [Papanastasiou (2019)]

- Συμμετοχή πλατφόρμας:
  - $R > 0$ = όφελος (ανά μετάδοση) για διαμοίραση αληθινών μηνυμάτων.
  - $P > 0$ = ζημιά (ανά μετάδοση) για διαμοίραση ψεύτικων μηνυμάτων ($P > R$).
  - $K_p > 0$ = κόστος για διεξαγωγή ελέγχου μηνύματος και κοινοποίηση αποτελέσματος σε όλους.
  - $b_{p,0} = \int_0^1 b \; dF(b)$. /* η πλατφόρμα ασπάζεται τις πεποιθήσεις των χρηστών */
  - $q_{p,i} = \frac{p_y q_0 w_i}{p_y q_0 w_i + [a b_{p,0} + (1-a)(1-b_{p,0})](1-q_0)}$

- Δράσεις πλατφόρμας (σε κάθε βήμα, πριν τη δράση του τρέχοντος χρήστη):

  GLOBAL CHECK Έλεγχος τύπου μηνύματος και κοινοποίηση του αποτελέσματος.

  NO CHECK Μη παρέμβαση πλατφόρμας.

Μελέτη Περίπτωσης

Ένα Ακολουθιακό Μοντέλο Διάδοσης [Papanastasiou (2019)]

- Συνάρτηση ωφέλειας της πλατφόρμας: $\forall t \geq 1$,

$$\begin{aligned}\mathbb{E}\left[U_p(t)\right] &= \sum_{i=1}^{t-1} \prod_{j=1}^{i-1}(1-N_j)\cdot\delta^{i-1} \\ &\quad \cdot\left[(R-P)S_i q_{p,i} + R(S_i+C_i)(1-q_{p,i})\right] \\ &\quad + \prod_{j=1}^{t-1}(1-N_j)\cdot\delta^{t-1}\cdot\left[\frac{R}{1-\delta}(1-q_{p,t}) - K_p\right]\end{aligned}$$

Μελέτη Περίπτωσης

Ένα Ακολουθιακό Μοντέλο Διάδοσης [Papanastasiou (2019)]

- Συνάρτηση ωφέλειας της πλατφόρμας: $\forall t \geq 1$,

$$\begin{aligned}\mathbb{E}\left[U_p(t)\right] &= \sum_{i=1}^{t-1} \prod_{j=1}^{i-1}(1-N_j)\cdot\delta^{i-1} \\ &\quad \cdot\left[(R-P)S_i q_{p,i} + R(S_i+C_i)(1-q_{p,i})\right] \\ &\quad + \prod_{j=1}^{t-1}(1-N_j)\cdot\delta^{t-1}\cdot\left[\frac{R}{1-\delta}(1-q_{p,t}) - K_p\right]\end{aligned}$$

L3 Έστω ότι η πλατφόρμα επιλέγει να μην κάνει έλεγχο στα πρώτα $T_c - 1$ βήματα της διάδοσης. Τότε:

if $q_{p,T_c} > \frac{K_p(1-\delta)}{P-R}$
then $a_p = T_c$ /* διεξαγωγή ελέγχου τη στιγμή T_c */
else $a_p = \infty$ /* καμιά διεξαγωγή ελέγχου */

Μελέτη Περίπτωσης

Ένα Ακολουθιακό Μοντέλο Διάδοσης [Papanastasiou (2019)]

- Υπολογισμός βέλτιστης στρατηγικής για την πλατφόρμα: Πρόβλημα πεπερασμένου χρονικού ορίζοντα $\{1, \ldots, T_c\}$...

Μελέτη Περίπτωσης

Ένα Ακολουθιακό Μοντέλο Διάδοσης [Papanastasiou (2019)]

- Υπολογισμός βέλτιστης στρατηγικής για την πλατφόρμα: Πρόβλημα πεπερασμένου χρονικού ορίζοντα $\{1,\ldots,T_c\}$...

T1 Έστω $T_c = \min\{t \geq 1 : S_t = 1\}$, και

$$u_{T_c} = \max\left\{\frac{1-q_{p,T_c}}{1-\delta}\cdot R - K_p, \frac{R - q_{p,T_c}\cdot P}{1-\delta}\right\}$$

Για κάθε $1 \leq t \leq T_c - 1$,

$$u_t = \max\left\{\begin{array}{l} \frac{1-q_{p,t}}{1-\delta}\cdot R - K_p, \\ C_t\cdot(1-q_{p,t})\cdot R + S_t\cdot(R - q_{p,t}\cdot P) \\ +[S_t + C_t\cdot(1-q_{p,t})]\cdot\delta\cdot u_{t+1} \end{array}\right\}$$

Για $\tau = \inf\{t \leq T_c : u_t = \frac{1-q_{p,t}}{1-\delta}\cdot R - K_p\}$, η βέλτιστη στρατηγική της πλατφόρμας ορίζεται ως εξής:

if $\tau = \infty$ **then** κανένας έλεγχος.
else έλεγχος τη στιγμή $\tau < \infty$.

Μελέτη Περίπτωσης

Ένα Ακολουθιακό Μοντέλο Διάδοσης [Papanastasiou (2019)]

- Η αναμενόμενη ωφέλεια της πλατφόρμας είναι:
  - Φθίνουσα ως προς την τιμή P της ζημιάς για μετάδοση ψευδών ειδήσεων.
  - Αύξουσα ως προς την τιμή K_p του κόστους για διεξαγωγή ελέγχου.

Μελέτη Περίπτωσης

Ένα Ακολουθιακό Μοντέλο Διάδοσης [Papanastasiou (2019)]

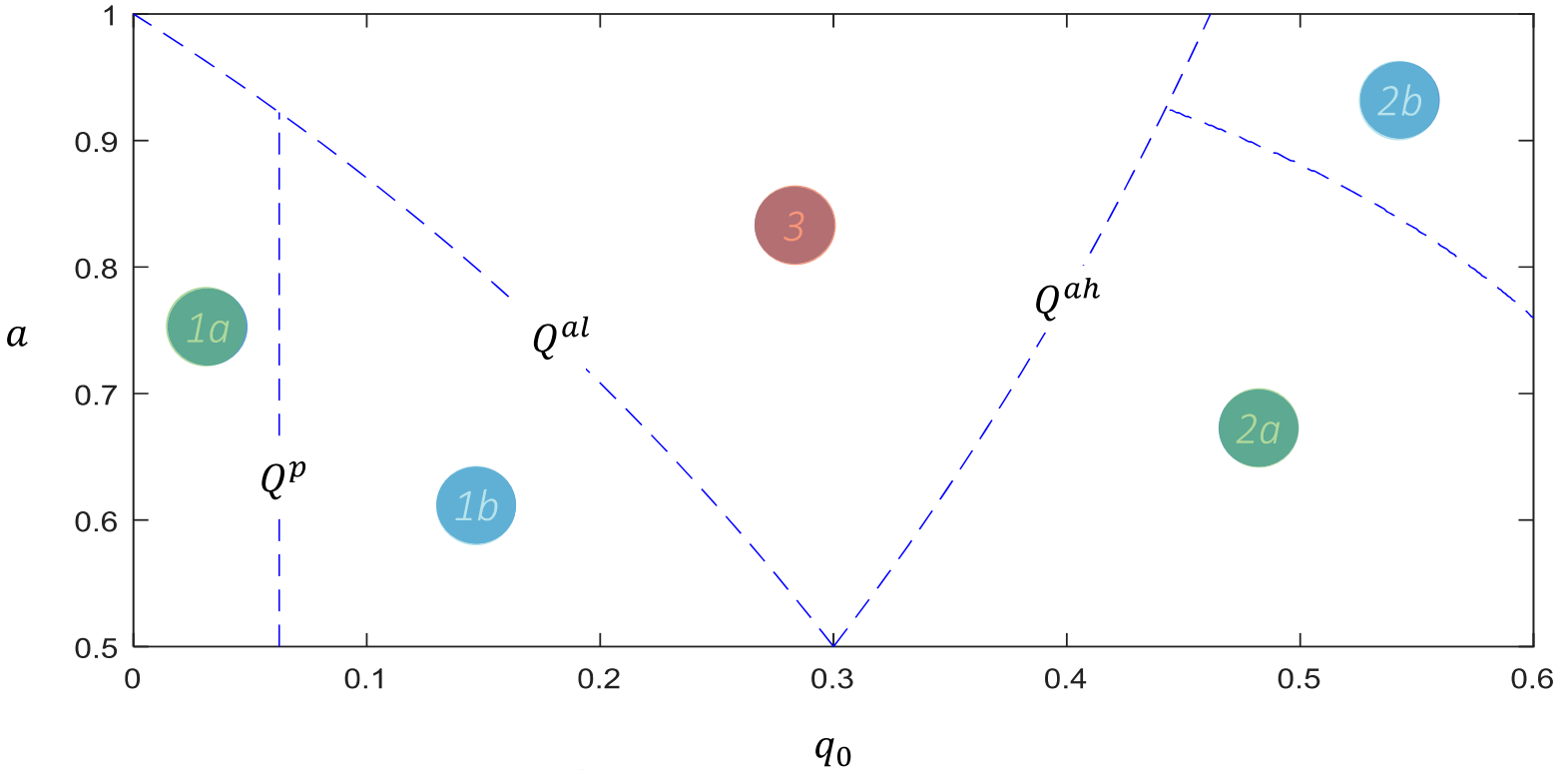
- Η αναμενόμενη ωφέλεια της πλατφόρμας είναι:
  - Φθίνουσα ως προς την τιμή P της ζημιάς για μετάδοση ψευδών ειδήσεων.
  - Αύξουσα ως προς την τιμή K_p του κόστους για διεξαγωγή ελέγχου.

- Υπάρχουν κατάλληλα κατώφλια ως προς το q_0, που καθορίζουν τη βέλτιστη στρατηγική για την πλατφόρμα.

Μελέτη Περίπτωσης

Ένα Ακολουθιακό Μοντέλο Διάδοσης [Papanastasiou (2019)]

- Η αναμενόμενη ωφέλεια της πλατφόρμας είναι:
  - Φθίνουσα ως προς την τιμή P της ζημιάς για μετάδοση ψευδών ειδήσεων.
  - Αύξουσα ως προς την τιμή K_p του κόστους για διεξαγωγή ελέγχου.



Ευχαριστώ για την προσοχή σας!

Μην αφήνεις ποτέ την αλήθεια να καταστρέψει μια ωραία ιστορία!

Ερωτήσεις / Σχόλια;