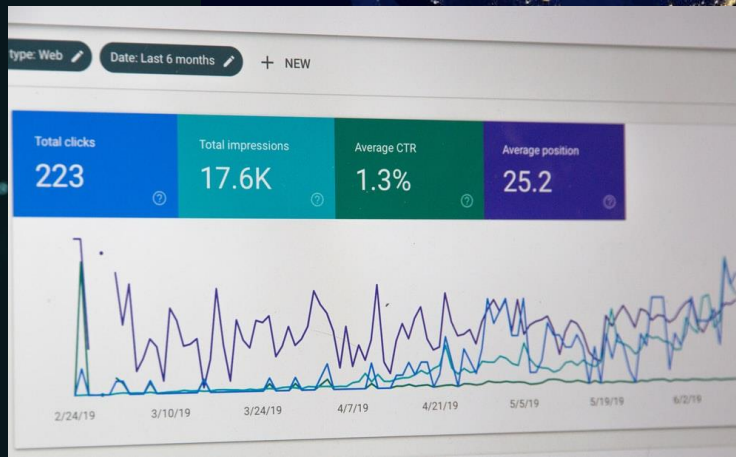




Πολιτισμική Πληροφορική

Παρουσίαση και
οπτικοποίηση

kkarpou@panteion.gr - @kkarpou



Information overload problem (IOP)

- Η **παρουσίαση** των αποτελεσμάτων της ανάλυσης σχετίζεται άμεσα με το **μήνυμα** που θέλουμε να περάσουμε
 - Digital Humanities as **narrative**
- Τα δεδομένα που έχουμε στη διάθεσή μας χαρακτηρίζονται από το μεγάλο τους **όγκο** και τις **διαφορές** στα **χαρακτηριστικά** (Big Data)
- Big Data → αναζήτηση σχέσεων ή τάσεων, συνδυασμός διαφορετικών ειδών δεδομένων, μηχανική μάθηση

Information overload problem (IOP)

- Πρόβλημα: τα πρωτογενή δεδομένα μπορεί να μην έχουν **αξία** αν δεν τα **επεξεργαστούμε**
 - Αλλά πρέπει να τα **διατηρήσουμε** για να δοκιμάσουμε κάτι διαφορετικό στο μέλλον
 - Χρόνος και πόροι για την **αποθήκευση** και **διατήρηση**

5Vs of Big Data

- **Velocity:** η ταχύτητα με την οποία δημιουργούνται καινούρια δεδομένα κάθε λεπτό
 - Κάθε μέρα, δημιουργούμε βίντεο διάρκειας 3000 ετών στο Facebook και βλέπουμε 4 δισ. βίντεο
 - 500 εκατ. χρήστες επισκέπτονται τα Instagram stories καθημερινά
 - 9.281 tweets ανά δευτερόλεπτο (!)
 - 1/3 χρήστες του TikTok ανεβάζουν βίντεο καθημερινά
- Οι φορητές συσκευές παράγουν δεδομένα τοποθεσίας, κινητικότητας και χρήσης εφαρμογών
- Internet of things
- Open data

5Vs of Big Data

- **Volume**

- Τα δεδομένα συσσωρεύονται και πρέπει (;) να διατηρηθούν
- Χρόνος για την επισημείωση και την ανάλυση/επεξεργασία
- Μεγαλύτερος όγκος δε σημαίνει απαραίτητα και μεγαλύτερη αξία

- **Value**

- Πώς γνωρίζω την αξία των δεδομένων αν δεν έχω σαφές ερώτημα;
- Τι γίνεται με τα μελλοντικά ερωτήματα;
- Ιδιοκτησία δεδομένων; Ηθικά ζητήματα;

- **Variety**

- Διαφορετικά είδη → διαφορετικές μέθοδοι διατήρησης και επεξεργασίας
- Μπορεί να δώσουν και διαφορετικά αποτελέσματα
- Σίγουρα θα απαιτήσουν άλλη μέθοδο παρουσίασης

5Vs of Big Data

- **Veracity**

- Τι μου δείχνουν τα δεδομένα;
- Όσο πλούσιο κι αν είναι ένα σώμα δεδομένων, παραμένει ένα «στιγμιότυπο» του πραγματικού κόσμου

- **Anchoring Bias**

- Δίνω μεγαλύτερη βάση στο πρώτο δεδομένο που βλέπω
- Και κρίνω τα υπόλοιπα με βάση αυτό

- **Attention Bias**

- Δίνω μεγαλύτερη βάση σε συγκεκριμένα δεδομένα ή χαρακτηριστικά τους (π.χ. το κυρίαρχο χρώμα μιας εικόνας)

- **Confirmation Bias**

- Επιλέγω τα δεδομένα ή το νόημα που επιβεβαιώνει τη θέση μου

Πώς θα επιλέξω τον τρόπο παρουσίασης;

- Πριν προχωρήσω, πρέπει να απαντήσω κάποια κρίσιμα ερωτήματα
- Πώς θα χρησιμοποιήσω τα δεδομένα για να πω την **ιστορία** που θέλω;
- Πώς θα ορίσω ποιες πληροφορίες είναι **σχετικές**;
- Ποια **μορφή παρουσίασης** και **αλληλεπίδρασης** είναι η καλύτερη για να λύσω κάποιο πρόβλημα ή να πάρω μια απόφαση;

Οπτικοποίηση

- Visualization
- Σε επιφανειακό επίπεδο, η διεργασία παραγωγής μιας γραφικής παράστασης ή μιας εικόνας που περιγράφει μια πτυχή των δεδομένων μου
- Στην πραγματικότητα, παράγει ένα **μοντέλο** των δεδομένων
 - Δημιουργεί μια **νοητική εικόνα** που βοηθά στην κατανόηση
- Στόχος: **ενόραση** (insight), όχι μόνο εμφάνιση
 - Ανακάλυψη σχέσεων, λήψη αποφάσεων, αιτιολόγηση

Οπτικοποίηση

- “Information visualization is the use of computer-supported, **interactive, visual representation of abstract data** in order to amplify **cognition**. Its purpose is not the pictures themselves, but **insight** (or rapid information **assimilation** or **monitoring** large amounts of data). Information visualization is part of the new media made possible by the development of real-time visual computer.” (p.637)
 - Stuart K. Card, Jock D. Mackinlay, Ben Shneiderman (1999). Readings in information visualization: using vision to think.

Οπτικοποίηση

- “Information visualization is the use of computer-supported, **interactive, visual representation of abstract data** in order to amplify **cognition**. Its purpose is not the pictures themselves, but **insight** (or rapid information **assimilation** or **monitoring** large amounts of data). Information visualization is part of the new media made possible by the development of real-time visual computer.” (p.637)
- Η χωρική συσχέτιση **δεν είναι πάντα εφικτή**
 - Γραφική παράσταση tweets ανά θέμα/ημέρα
 - Πολλές διαστάσεις των δεδομένων (περισσότερες από 2/3/4)

Γιατί οπτικοποίηση;

- **Διάθεση** στατικών και δυναμικών δεδομένων σε διαφορετικούς καταναλωτές
- Επιτρέπει τη **διαμόρφωση υποθέσεων** και την **επικύρωσή** τους
- Ανακάλυψη **τάσεων** και **προτύπων**
- **Εποπτική επεξεργασία** μεγάλων σωμάτων δεδομένων
- Εντοπισμός και διαχείριση **απόντων ή αβέβαιων δεδομένων**
 - Π.χ. η τοποθεσία μιας φωτογραφίας

Χαρακτηριστικά

- Οι οπτικές αναπαραστάσεις ποικίλλουν ανάλογα με το **είδος** των δεδομένων και το **ερώτημα**
 - **Χρονική εξέλιξη** δεδομένων, πολλαπλές διαστάσεις
 - **Ιεραρχική** παρουσίαση (π.χ. οπτικοακουστικά μέσα → κινηματογραφικές ταινίες → είδος)
 - Δίκτυα και γράφοι (π.χ. χρήστες ενός κοινωνικού δικτύου)
- Ειδικές περιπτώσεις
 - Δεδομένα που αλλάζουν **δυναμικά** (π.χ. τιμές μιας μετοχής)
 - Βίντεο, χρήση ανά **συσκευή**
- Πεδίο εφαρμογής
 - Humanities, κοινωνικές επιστήμες, λογοτεχνία, κλπ.
 - Storytelling?

Διαδραστικότητα

- Κυρίως για τη διαχείριση των διαφορετικών **διαστάσεων** και του μεγάλου **όγκου** των δεδομένων
- Φιλτράρισμα, επιλογή
 - Π.χ. εμφάνιση tweets με συγκεκριμένο hashtag ή σε περιορισμένο χρονικό διάστημα
- Μεταβολή οπτικής αναπαράστασης
 - Από 2D γραφική παράσταση σε χάρτη που χρησιμοποιεί πληροφορία τοποθεσίας
- Κλίμακα
 - Μέγεθος αναπαράστασης, πυκνότητα, ποσότητα δεδομένων
- Πρώτα παρουσιάζουμε **εποπτικά** → επιλέγουμε το **εύρος** → παρέχουμε συγκεκριμένες πληροφορίες

Visual analytics

- Το πρόβλημα ξεκινάει από την ανάγκη (;) να οπτικοποιήσουμε μεγάλα σώματα δεδομένων
- Δεν υπάρχει η δυνατότητα να οπτικοποιήσουμε την πληροφορία από 9.281 tweets ανά δευτερόλεπτο
 - Ούτε οι υπολογιστές την έχουν (not enough pixels!)
 - Ούτε εμείς μπορούμε να την επεξεργαστούμε και να την κατανοήσουμε
- Επιλογή και φιλτράρισμα
 - Συνδυασμός ανθρώπων και μηχανών



Data Storage

Computing Power

Search, Filter, ...

Planning

Diagnosis

Logics

Predictions

Domain Knowledge

General Knowledge

Perception

Creativity

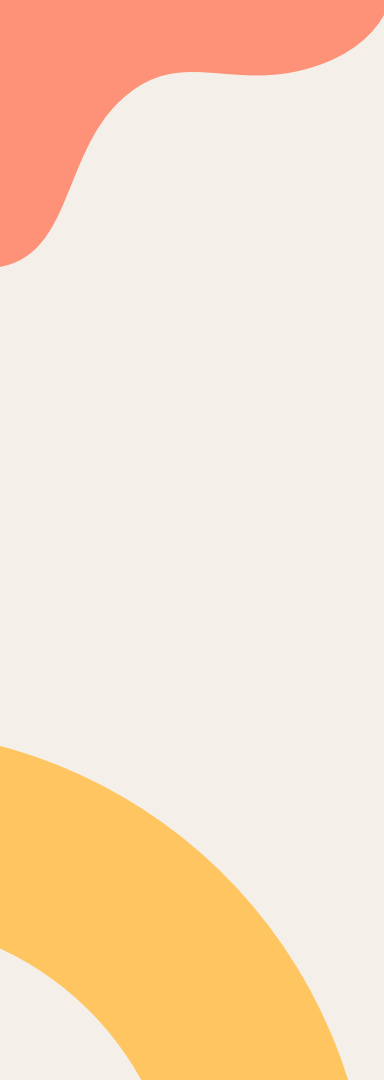


Visual analytics

- “Detect the expected and discover the unexpected”
- Συνδυάζει την οπτικοποίηση της πληροφορίας με τη διαχείρισή της και τις τεχνικές επεξεργασίας και μηχανικής μάθησης
- **Αυτόματη** ανάλυση
 - Αν είναι εφικτό
- **Οπτικοποίηση**
- Επεξεργασία από τους **χρήστες**
 - Αν δεν είναι εφικτή η αυτόματη ανάλυση ή αν απαιτείται η επεξεργασία των ενδιάμεσων αποτελεσμάτων

Σύνδεσμοι

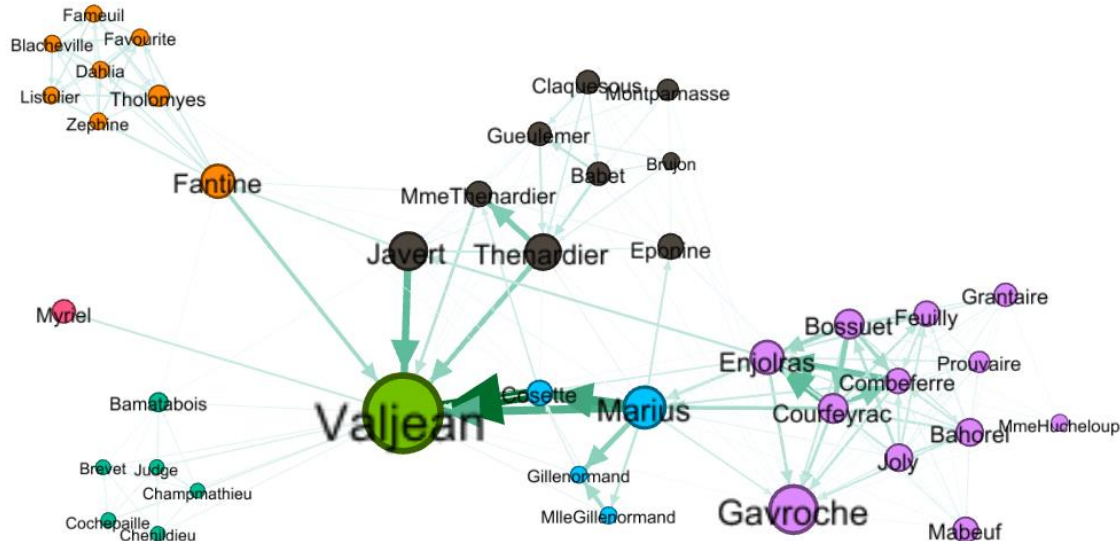
- Παραδείγματα οπτικοποίησης,
<https://philogb.github.io/jit/demos.html>
- Spence, R. (2001). Information visualization (Vol. 1). New York: Addison-Wesley.
- Keim, D., Andrienko, G., Fekete, J. D., Görg, C., Kohlhammer, J., & Melançon, G. (2008). Visual analytics: Definition, process, and challenges. In Information visualization (pp. 154-175). Springer, Berlin, Heidelberg.
- Workshop on Visualization for the Digital Humanities,
<http://www.vis4dh.org/>
- Drucker, J. (2015). Graphical Approaches to the Digital Humanities. A New Companion to Digital Humanities, 238–250,
<https://doi.org/10.1002/9781118680605.ch17>



Οπτικοποίηση δικτύων

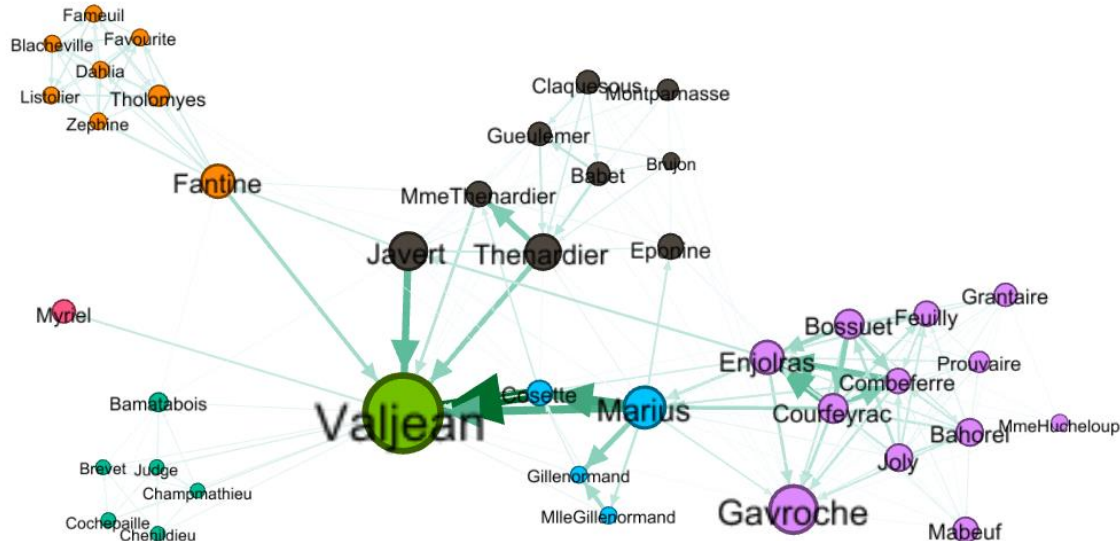
Ανάλυση δικτύων

- **Δίκτυο:** οργάνωση δεδομένων με οντότητες που συνδέονται με κάποια σχέση
- **Οντότητες** → Κόμβοι, **Σχέσεις** → Ακμές



Ανάλυση δικτύων

- Οι ακμές μπορεί να είναι **κατευθυνόμενες** (directed, Twitter) ή μη **κατευθυνόμενες** (undirected, Facebook)



Τι μπορώ να μετρήσω σε ένα δίκτυο;

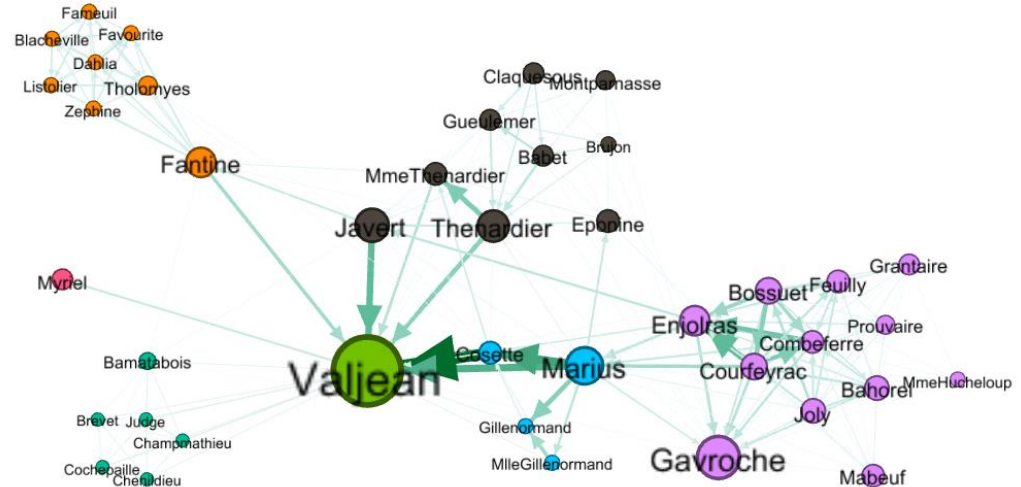
- **Power** (== centrality)
 - Η θέση ενός κόμβου στο δίκτυο
- **Degree**
 - Αριθμός συνδέσεων με άλλους κόμβους
 - In-degree: εισερχόμενες ακμές (Twitter mentions)
 - Out-degree: εξερχόμενες ακμές (ο χρήστες κάνει mention άλλους)
- **Closeness**
 - Απόσταση από το υπόλοιπο δίκτυο
- **Betweenness**
 - Πόσο συχνά ένας κόμβος βρίσκεται στη γραμμή που συνδέει δύο κόμβους
 - Π.χ. κοινός φίλος στο FB, ηθοποιός που έχει παίξει σε κοινές σειρές με άλλους δύο, κλπ.

Social network analysis

- Ειδική περίπτωση δικτύων
 - Η σχέση μεταξύ δύο κόμβων αντιστοιχεί σε σχέση στο Μέσο Κοινωνικής Δικτύωσης ή IRL
- Μπορώ να υπολογίσω
 - Πιθανότητα σύνδεσης (people you may know)
 - Κοινωνικό κύκλο
 - Προτάσεις για σύνδεση
 - Το δίκτυο ενός κόμβου
 - Κοινότητες χρηστών (clusters, cliques ή υποδίκτυα)

Μετρήσιμα δεδομένα

- Αριθμός συνδέσεων κόμβου (degree)
- Μέσος αριθμός συνδέσεων
- Διάμετρος → μέγιστη απόσταση ανάμεσα σε δύο κόμβους
- Εγγύτητα → μέση απόσταση ενός κόμβου με τους υπόλοιπους
- Κοινότητες (modularity)



Κέντρο του δικτύου

- Μέτρηση degree/closeness/betweenness
- Γενικά, θα συσχετίζονται θετικά – αν όχι:

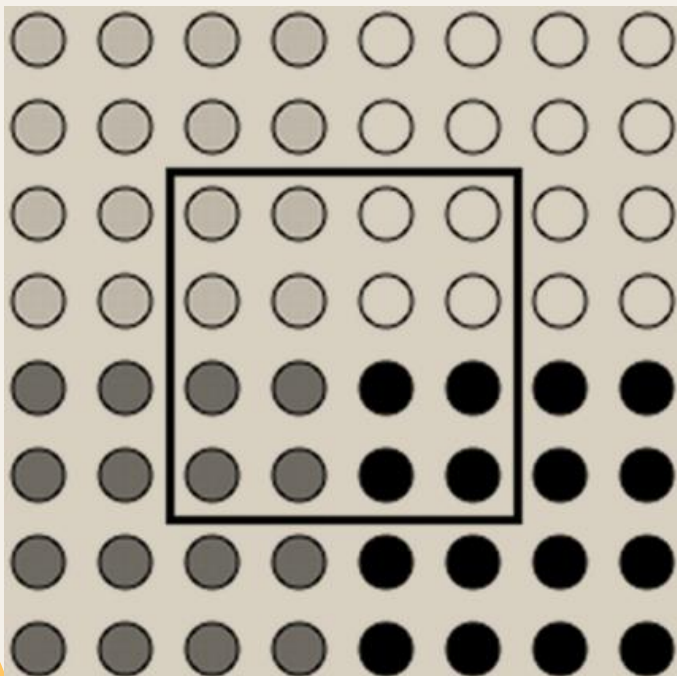
	Low degree	Low closeness	Low betweenness
High degree		Πυκνό υποδίκτυο, μακριά από το υπόλοιπο δίκτυο	Μακριά από την ουσιαστική επικοινωνία των υπολοίπων κόμβων
High closeness	Κεντρικός κόμβος, συνδεδεμένος με άλλους σημαντικούς		Πολλαπλά μονοπάτια στο δίκτυο, πολλές πιθανότητες επικοινωνίας
High betweenness	Λίγες συνδέσεις, αλλά σημαντικές για τη ροή πληροφοριών	Ο κόμβος «ελέγχει» τις συνδέσεις από μια ομάδα κόμβων στο υπόλοιπο δίκτυο	

Connection bias

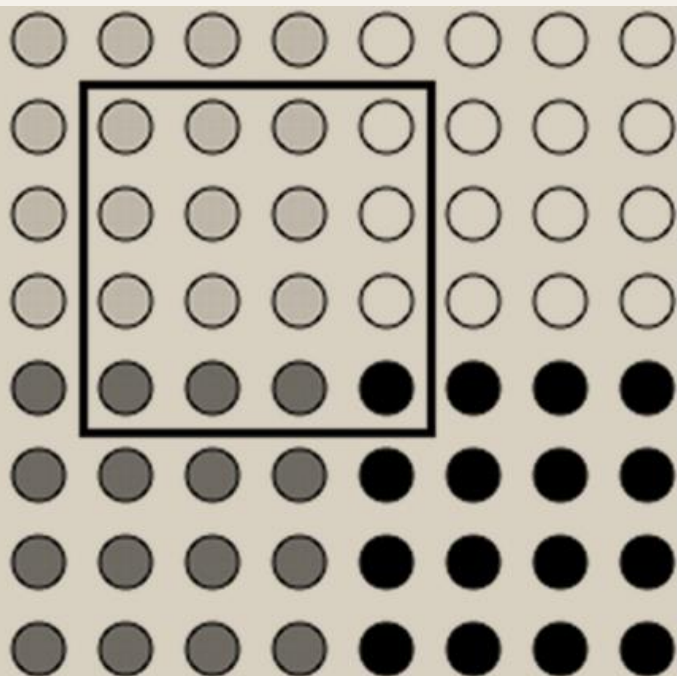
- Στη γενική περίπτωση, τα δεδομένα αποτελούν ένα **στιγμιότυπο** του δικτύου
- **Population sampling**: με δεδομένο έναν πληθυσμό (π.χ. Έλληνες χρήστες του twitter), να επιλέξουμε ένα υποσύνολο γνωρίζοντας την πιθανότητα να συμπεριλάβουμε έναν συγκεκριμένο κόμβο
 - Δημοσκοπήσεις
- Στα Μέσα Κοινωνικής Δικτύωσης, δεν είναι προφανές ότι μπορούμε να **απαριθμήσουμε** τον πληθυσμό ή τα χαρακτηριστικά του

Connection bias

Αντιπροσωπευτικό δείγμα



Μη αντιπροσωπευτικό δείγμα



Netnography

- Network + Ethnography
 - **Εθνογραφική μελέτη** των **ψηφιακών** αλληλεπιδράσεων
 - Ποιοτική ανάλυση των social media
- Τι αναζητούμε;
 - Ψηφιακές **κοινότητες** (communities): ομάδες ατόμων με κοινούς δεσμούς ή πρακτικές, που αλληλεπιδρούν σε ένα ψηφιακό μέσο
 - **Consociality**: ταυτόχρονη παρουσία ατόμων στον πραγματικό ή τον ψηφιακό κόσμο → ευκαιρία για κοινωνική αλληλεπίδραση

Προσεγγίσεις netnography

- **Ενεργητική (active):** δημιουργία δεδομένων μέσα από την αλληλεπίδραση με τους κόμβους του δικτύου
 - Κυρίως συνεντεύξεις και ερωτηματολόγια
 - Πυκνότερες αναπαραστάσεις
 - Ενσωμάτωση γνώσης και εμπειρίας
 - Αντικειμενικότητα
- **Παθητική (passive):** παρατήρηση και καταγραφή των αλληλεπιδράσεων σε μια ψηφιακή κοινότητα ή στην πραγματική ζωή
 - Κατανόηση του περιβάλλοντος
 - Μικρότερο κόστος
 - Περισσότερο προσβάσιμη

Σύνδεσμοι

- Fenton & Procter (2019) "Studying social media communities: Blending methods with netnography." Sage Research Methods Cases.
- Weingart (2011) "Demystifying Networks, Parts I & II." Journal of Digital Humanities
- Kozinets (2018) "Netnography for management and business research." The Sage Handbook of Qualitative Business and Management Research Methods.
- Costello, L., McDermott, M. L. & Wallace, R. (2017). Netnography: Range of practices, misperceptions, and missed opportunities. International Journal of Qualitative Methods, 16(1), <https://doi.org/10.1177/1609406917700647>

Ερωτήσεις αυτοαξιολόγησης

- Μπορώ να επιλέξω διαφορετικές μεθόδους οπτικοποίησης για το ίδιο σύνολο πρωτογενών δεδομένων και γιατί;
- Πώς σχετίζεται η επιλογή μεθόδου οπτικοποίησης με τη συσκευή στην οποία θα καταναλωθεί;
- Πώς μεταβάλλεται η προσβασιμότητα των αποτελεσμάτων σε σχέση με την οπτικοποίηση;
- «Περισσότερα δεδομένα → σαφέστερη οπτικοποίηση» – σωστό ή λάθος και γιατί;
- Ποια ερωτήματα μπορώ να απαντήσω με ένα δίκτυο μουσικών/συνθετών/τραγουδιστών και πώς;
- Πώς επηρεάζει η αντιπροσωπευτικότητα την ποιότητα των ποιοτικών απαντήσεων που δίνουμε στην ανάλυση κοιν. δικτύων;