Σύνθετα Δίκτυα

com+plex: with+ -fold (having parts)

Διδάσκων –
Δημήτριος Κατσαρός

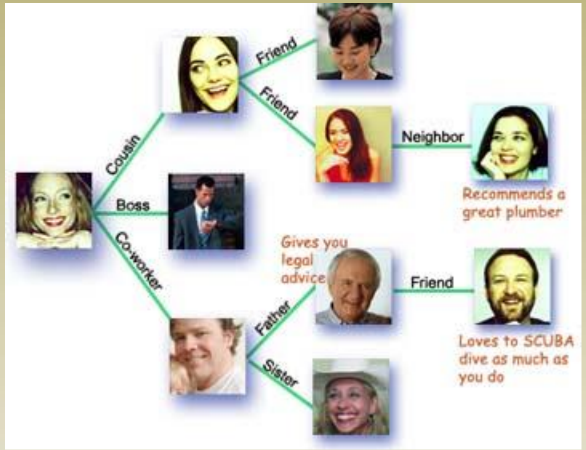
Διάλεξη 14η

Influence maximization

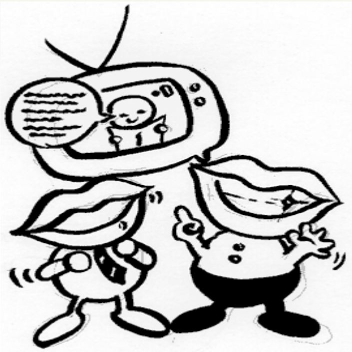
Μεγιστοποίηση επιρροής

Κοινωνικά δίκτυα και διάδοση επιρροής

- Τα κοινωνικά δίκτυα διαδραματίζουν θεμελιώδη ρόλο ως μέσα για τη διάδοση της ΕΠΙΡΡΟΗΣ μεταξύ των μελών τους
  - Γνώμες, ιδέες, πληροφορίες, καινοτομία …



- Το direct marketing εκμεταλλεύεται το “word-of-mouth” για να αυξήσει σημαντικά τα ωφέλη (Gmail, Tupperware popularization, Microsoft Origami …)



Διατύπωση του προβλήματος

- Δίνεται
  - Ένας περιορισμένος προϋπολογισμός B για αρχική διαφήμιση (π.χ., δωρεάν δείγματα του προϊόντος)
  - Εκτιμήσεις για την επιρροή μεταξύ των ατόμων
- Στόχος
  - Κινητοποίηση ενός μεγάλου καταρράκτη επιρροής (π.χ., περεταίρω υιοθέτηση του προϊόντος)
- Ερώτηση
  - Σε ποιο υποσύνολο B των ατόμων να στοχεύσουμε;
- Εφαρμογές πέραν του marketing προϊόντων
  - Διάδοση καινοτομιών
  - Ανίχνευση ιστοριών σε blogs

Τι χρειαζόμαστε

- Μοντέλα επιρροής στα κοινωνικά δίκτυα
- Δεδομένα για τέτοια δίκτυα (για να εκτιμήσουμε την δια-προσωπική επιρροή)
- Επινόηση αλγορίθμου για τη διάδοση της επιρροής

Μοντέλα επιρροής

- Τα πρώτα μαθηματικά μοντέλα
  - [Schelling '70/'78, Granovetter '78]
- Μετέπειτα, αρκετές εργασίες:
  - [Rogers '95, Valente '95, Wasserman/Faust '94]
- Δυο βασικές ομάδες μοντέλων διάχυσης: threshold και cascade
- Γενική θεώρηση:
  - Το κοινωνικό δίκτυο αναπαρίσταται ως γράφημα
  - Οι κόμβοι ξεκινούν είτε ως active είτε ως inactive
  - Ένας active κόμβος node μπορεί να εκκινήσει activation των γειτονικών κόμβων
  - Υπόθεση : οι active κόμβοι ποτέ δεν γίνονται deactivate

Linear Threshold Model

- Ένας κόμβος v έχει τυχαίο κατώφλι $\theta_v \sim U[0,1]$
- Ένας κόμβος v υφίσταται επιρροή από κάθε γείτονα w σύμφωνα με το *weight* b_{vw} τέτοιο ώστε

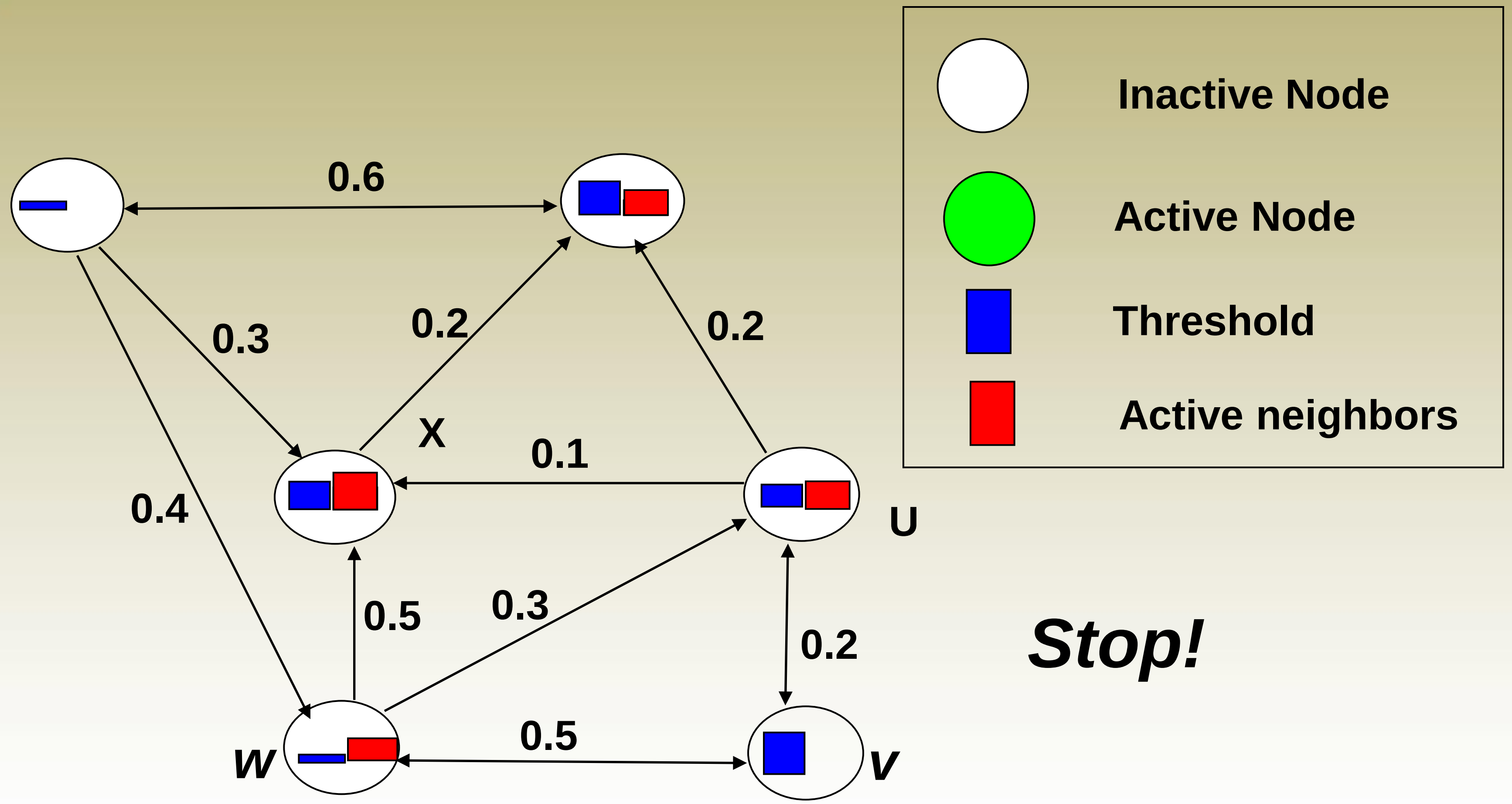
$$\sum_{w \text{ neighbor of } v} b_{v,w} \leq 1$$

- Ένας κόμβος v γίνεται active όταν τουλάχιστον

(weighted) θ_v κλάσμα των γειτόνων είναι active

$$\sum_{w \text{ active neighbor of } v} b_{v,w} \geq \theta_v$$

Παράδειγμα

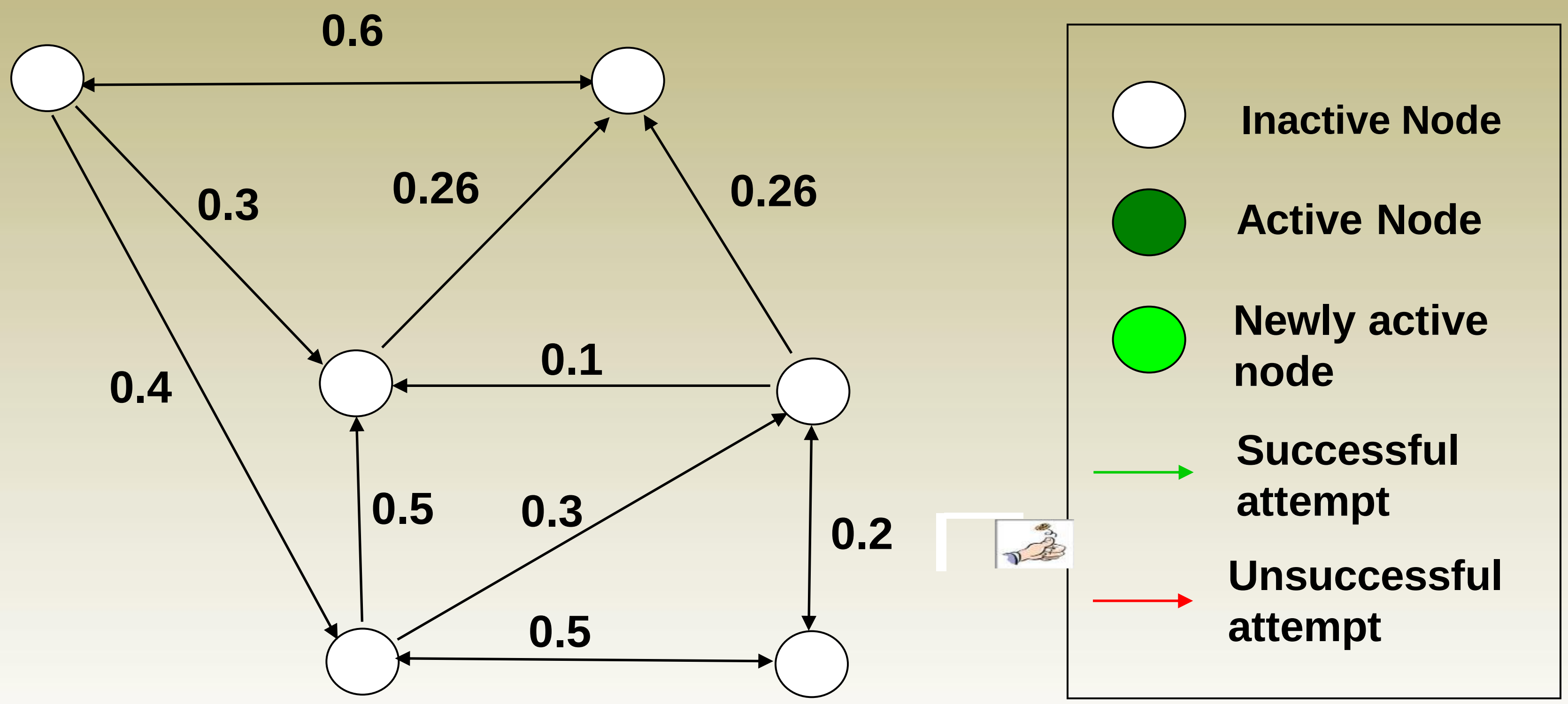


Stop!

Independent Cascade Model

- Όταν ο κόμβος v γίνεται active, έχει ΜΙΑ ΜΟΝΟ (single) ευκαιρία να ενεργοποιήσει (activating) κάθε ΤΩΡΑ inactive γείτονα w
- Η προσπάθεια activation επιτυγχάνει με πιθανότητα p_{vw}

Παράδειγμα



Stop!

Το πρόβλημα μεγιστοποίησης επιρροής

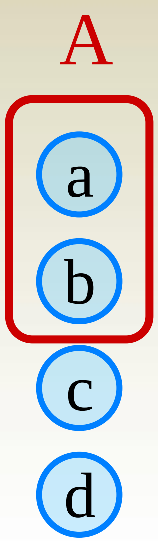
- Επιρροή (influence) του συνόλου κόμβων S: f(S)
  - expected αριθμός των active κόμβων στο τέλος, εάν το σύνολο S είναι το αρχικό active σύνολο
- Πρόβλημα:
  - Δεδομένης μια παραμέτρου k (budget), βρες ένα k-node set S που να μεγιστοποιεί το f(S)
  - Constrained optimization problem με το f(S) ως την αντικειμενική συνάρτηση

Set functions και energy functions

any set function with $|V| = n$.

... is a function on binary vectors!

$$F : 2^V \to \mathbb{R} \qquad\qquad f : \{0,1\}^n \to \mathbb{R}$$



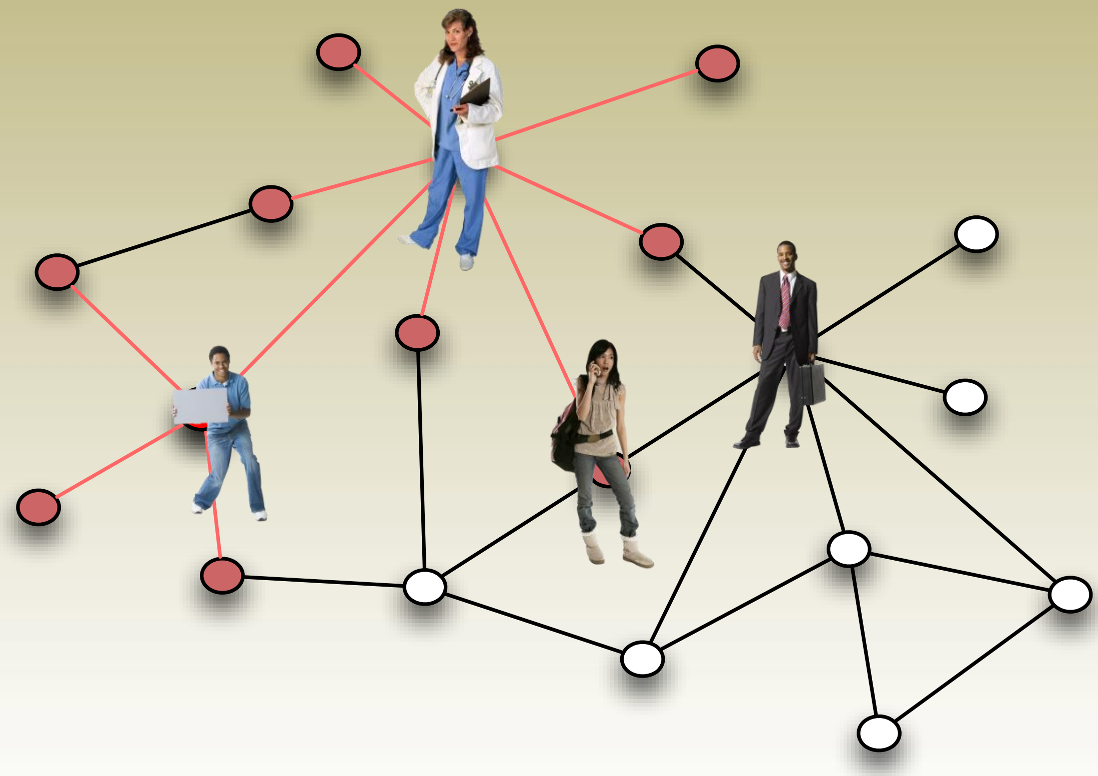
$$\hat{=}$$

$$x = 1_A$$

$$\begin{array}{cc} \boxed{1} & a \\ \boxed{1} & b \\ \boxed{0} & c \\ \boxed{0} & d \end{array}$$

binary labeling problems = subset selection problems!

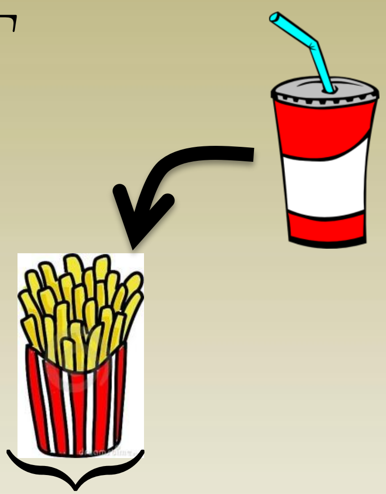
Influential υποσύνολα



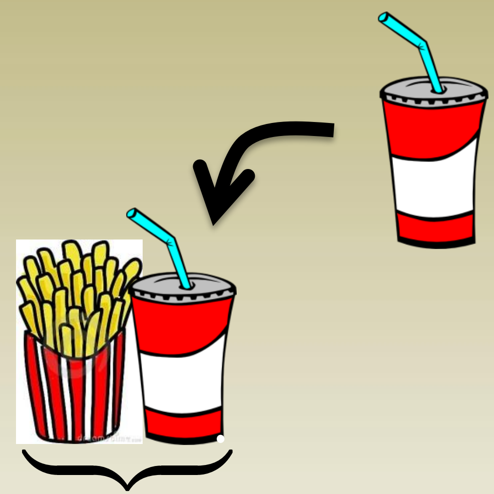
$$F(S) = \text{spread}$$

Submodularity

$S \subseteq T$



S



T

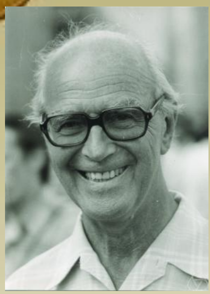
$$F(S \cup s) - F(S) \quad \geq \quad F(T \cup s) - F(T)$$

extra cost:
one drink

extra cost:
free refill ☺

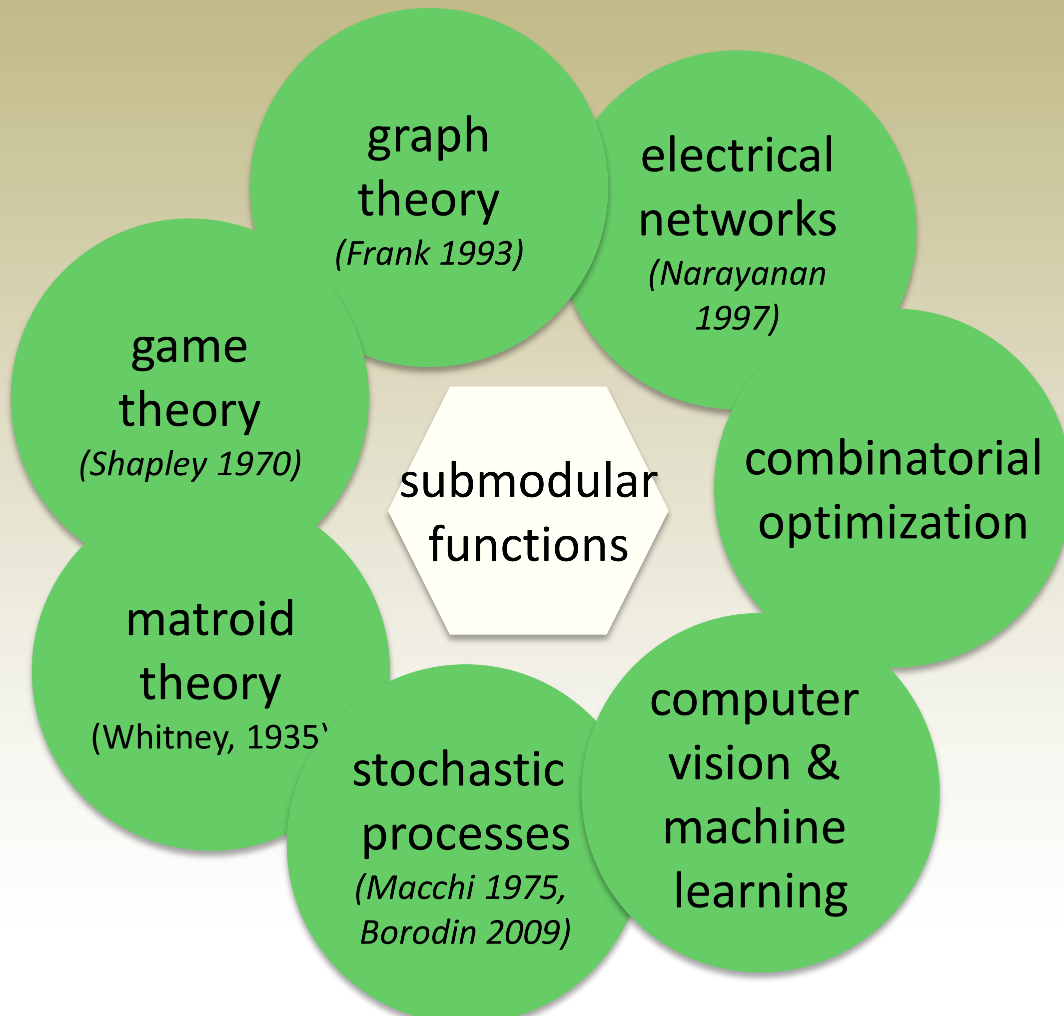
diminishing marginal costs

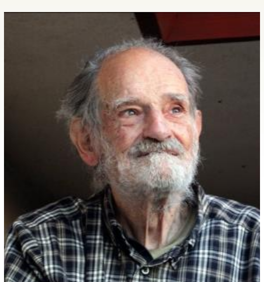
Η μεγάλη εικόνα



G. Choquet

J. Edmonds

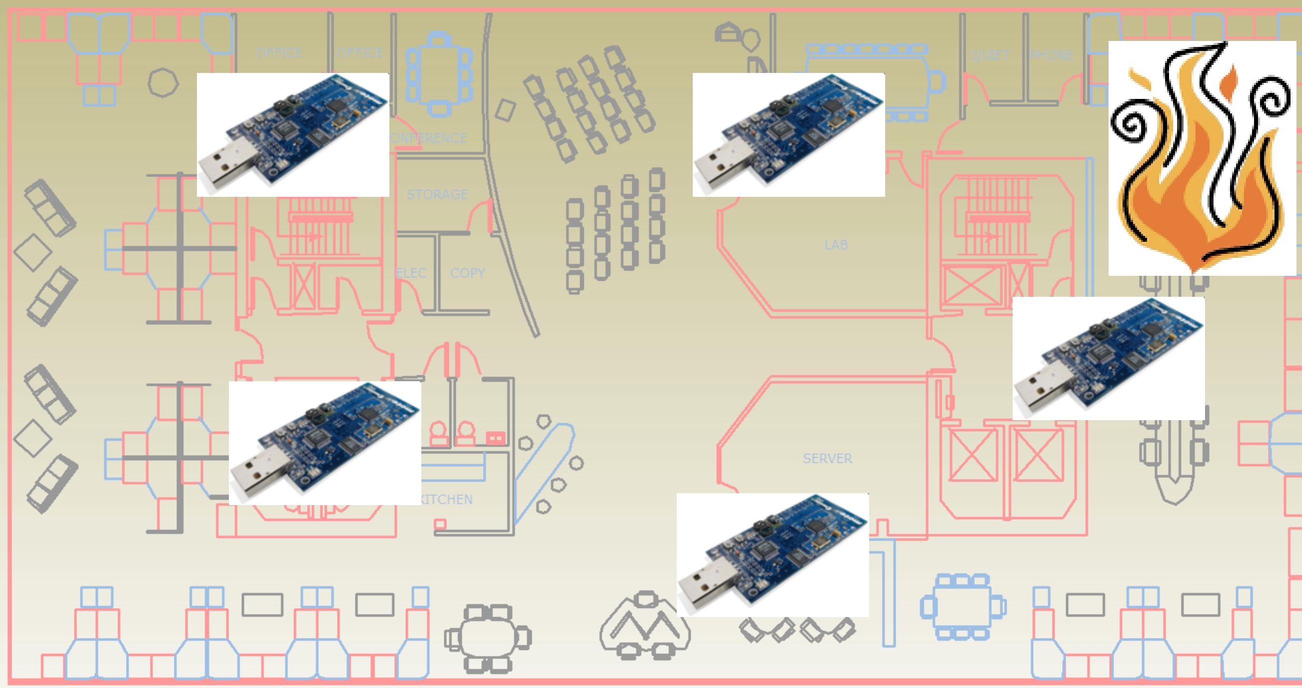




L.S. Shapley

L. Lovász

Παραδείγματα

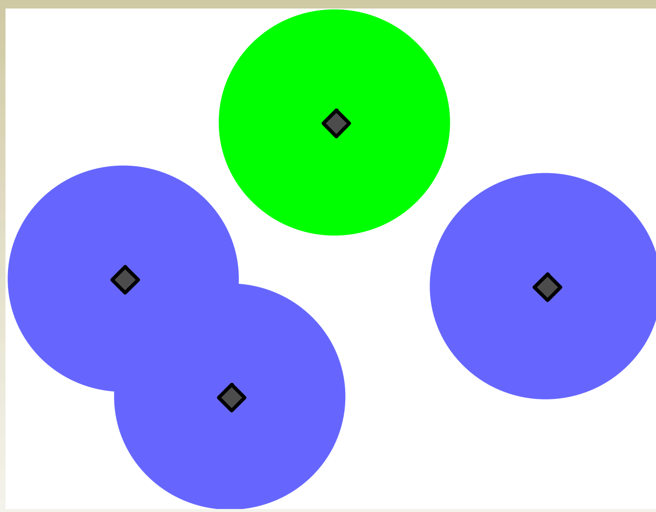


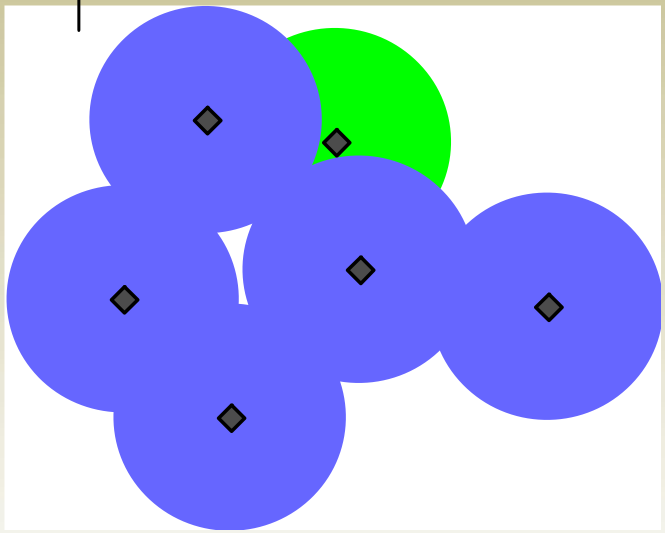
sensing:

F(S) = information gained from locations S

Παράδειγμα: cover

$$F(S) = \left| \bigcup_{v \in S} \text{area}(v) \right|$$

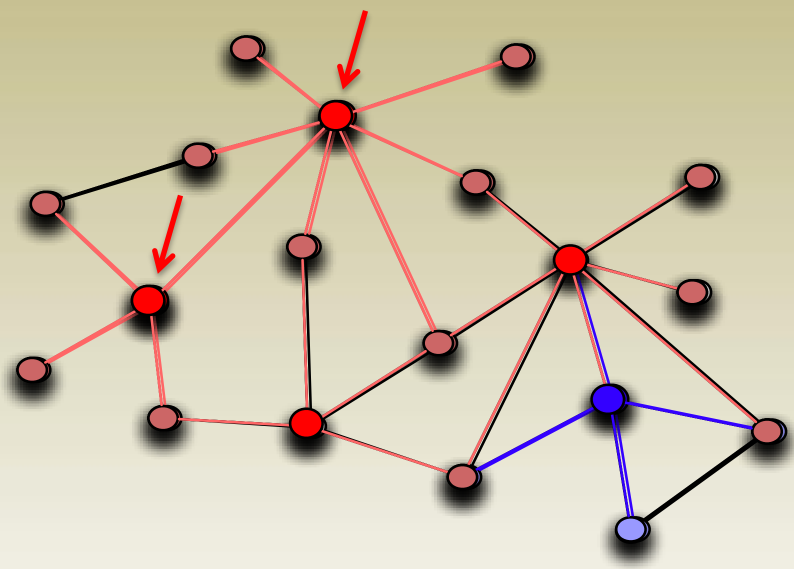




$$F(A \cup v) - F(A) \quad \geq \quad F(B \cup v) - F(B)$$

Μεγιστοποίηση επιρροής

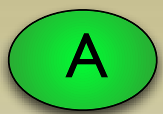
$$F(S) = \text{expected } \# \text{ infected nodes}$$



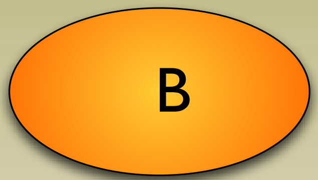
$$F(S \cup s) - F(S) \quad \geq \quad F(T \cup s) - F(T)$$

Submodular set functions

- Diminishing gains: for all $A \subseteq B$



+• e

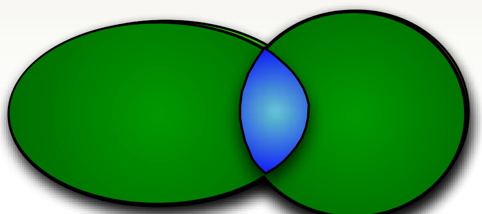


+• e

$$F(\underline{A \cup e}) - F(A) \quad \geq \quad F(B \cup e) - \underline{F(B)}$$

- Union-Intersection: for all

$$S, T \subseteq \mathcal{V}$$



$$\underline{F(S)} \; + \; \underline{F(T)} \quad \geq \quad F(S \cup T) \; + \; F(S \cap T)$$

f(S): ιδιότητες

- Μη αρνητική (προφανώς)
- Μονότονη: $f(S+v) \geq f(S)$
- Submodular:
  - Έστω N ένα πεπερασμένο σύνολο
  - Μια set συνάρτηση $f : 2^N \mapsto \Re$ είναι submodular *iff*

$$\forall S \subset T \subset N, \forall v \in N \setminus T,$$

$$f(S+v) - f(S) \geq f(T+v) - f(T)$$

(diminishing returns)

Τα άσχημα νέα

- Για μια submodular συνάρτηση *f*, εάν η *f* παίρνει μόνο μη-αρνητική τιμή, και είναι μονότονη, η εύρεση ενός συνόλου k στοιχείων, του S, για το οποίο το $f(S)$ βελτιστοποιείται είναι ένα πρόβλημα βελτιστοποίησης NP-hard
- Είναι NP-hard για να προσδιοριστεί το βέλτιστο για τη μεγιστοποίηση της επιρροής και για το independent cascade model και για το linear threshold model

Τα καλά νέα

- Μπορούμε να χρησιμοποιήσουμε κάποιον άπληστο αλγόριθμο!
  - Εκκίνηση με ένα κενό σύνολο S
  - Για k επαναλήψεις:

    Προσθέστε τον κόμβο v στο S που μεγιστοποιεί το $f(S+v) - f(S)$
- Πόσο καλό (κακό) είναι αυτό?
  - Θεώρημα: Ο άπληστος αλγόριθμος είναι μια $(1 - 1/e)$ προσέγγιση
  - Το προκύπτον σύνολο S activates τουλάχιστον $(1- 1/e) > 63\%$ του αριθμού των κόμβων που θα μπορούσε να ενεργοποιήσει οποιοδήποτε σύνολο S μεγέθους k

Ο άπληστος αλγόριθμος

Algorithm 1 Greedy alg. for influence maximization

Require: G, k, σ_m

Ensure: seed set S

1: $S \leftarrow \emptyset$
2: **while** $|S| < k$ **do**
3: $\quad u \leftarrow \arg\max_{w \in V \setminus S}(\sigma_m(S \cup \{w\}) - \sigma_m(S));$
4: $\quad S \leftarrow S \cup \{u\}$