



Theatina

Καλησπέρα σας.

Είμαστε οι Θεατίνα Κυλάφη και Τζουλιάνα Πιργιάσι και θα σας παρουσιάσουμε το άρθρο με τίτλο «A Game – Theoretic Approach to Word Sense Disambiguation», το οποίο αφορά τον τομέα της επεξεργασίας φυσικής γλώσσας και συγκεκριμένα την άρση της αμφισημίας των λέξεων (ΑΑΛ), εφαρμόζοντας τεχνικές της θεωρίας παιγνίων.

# Introduction

- **New model** for Word Sense Disambiguation (WSD) formulated in terms of evolutionary game theory
- **Word Sense Disambiguation:** the task of identifying the intended meaning of a word based on the context in which it appears
- “**bank**”
  - Financial sense
  - Naturalistic sense

Theatina

Παρουσιάζεται λοιπόν ένα νέο μοντέλο, που αξιοποιεί την εξελικτική προσέγγιση του τομέα.

Αρχικά θα ορίσουμε το WSD. Κάθε γλώσσα περιλαμβάνει λέξεις οι οποίες μπορεί να είναι αμφίσημες, δηλαδή να έχουν περισσότερες από μια έννοιες, ανάλογα με το περικείμενο στο οποίο βρίσκονται. Η αποσαφήνιση αυτών των έννοιών είναι η διαδικασία η οποία εντοπίζει τη συγκεκριμένη έννοια που λαμβάνει η λέξη μέσα σε αυτήν την πρόταση δηλαδή το περικείμενο.

Για παράδειγμα, η λέξη *bank* είναι αμφίσημη. Οι κύριες έννοιες τις οποίες μπορεί να λάβει είναι η οικονομική και η περιβαλλοντική / νατουραλιστική .

Η ΑΑΛ είναι ακριβώς αυτό, προσπαθεί να βρει ποια από τις δύο σημασίες κυριαρχεί κάθε φορά ανάλογα με την πρόταση μέσα στην οποία βρίσκεται η λέξη «*bank*»

## #Notes

Το άρθρο παρουσιάζει ένα νέο μοντέλο που προσεγγίζει το θέμα της αποσαφήνισης των εννοιών μιας λέξης από τη σκοπιά της θεωρίας των παιγνίων (Game theory) και συγκεκριμένα της εξελικτικής θεωρίας (evolutionary game theory).

## Background

- Approaches based on
  - 1. Learning models
    - Supervised
    - Unsupervised
    - Semi-supervised
  - 2. Techniques
    - Heuristic
  - 3. Algorithms
    - Graph-based
    - Knowledge-based
- ✓ Remain on the surface of the word, compromising the coherence of the data to be analyzed

### Theatina

Το θέμα του WSD είχε προσεγγιστεί ήδη από τις αρχές του τομέα επεξεργασίας φυσικής γλώσσας (NLP), ωστόσο οι μελέτες αυτές παρέμεναν στην επιφάνεια του θέματος, παραμελώντας τη συνοχή του κειμένου το οποίο εξέταζαν.

Οι μελέτες προσεγγίζονταν από τη σκοπιά επιβλεπόμενων, μη επιβλεπόμενων ή ημι-επιβλεπόμενων μοντέλων μάθησης, από ευρετικές τεχνικές (και αλγορίθμους) καθώς και αλγορίθμους βασισμένους σε γραφήματα (graph based) ή γνώση (knowledge based).

### #Notes

**Supervised** = μοντέλα τα οποία μαθαίνουν από ένα σύνολο δεδομένων με παραδείγματα και μετά είναι ικανά να χαρακτηρίσουν νέα παραδείγματα.

**Unsupervised** = μέσα από μη χαρακτηρισμένα/επισημειωμένα δεδομένα προσπαθούν να ανακαλύψουν μια πιθανή δομή.

**Semi-supervised** = Ένα μικρό ποσό χαρακτηρισμένων και ένα μεγάλο μη χαρακτηρισμένων παραδειγμάτων.

**Heuristics** = είναι οποιαδήποτε προσέγγιση στην επίλυση που χρησιμοποιεί μια πρακτική μέθοδο που δεν είναι εγγυημένη ότι είναι βέλτιστη αλλά είναι ωστόσο επαρκής για την επίτευξη ενός άμεσου, βραχυπρόθεσμου στόχου.

**Knowledge based** = χρησιμοποιώντας μια υπάρχουσα λεξική βάση γνώσεων, χρησιμοποιεί αλγόριθμους γραφημάτων για να αποσαφηνίσει τις λέξεις με βάση τις σχέσεις που έχουν έννοιες αυτών των λέξεων στο network.

**Graph based** = μοντελοποιεί τις σχέσεις μεταξύ λέξεων και αισθήσεων ενός κειμένου, σε γραφήματα, που αντιπροσωπεύουν λέξεις και αισθήσεις ως κόμβους και τις σχέσεις μεταξύ τους ως άκρα

# Game Theory

- **Mathematical approach to study the interaction between two or more individuals**
  - Benefits and costs depend on the strategies of each other
  - **Players**  $I = \{1, \dots, n\} \rightarrow N$ : total number of players
  - **Pure player strategy set**  $M_i = \{m_{i1}, m_{i2}, \dots, m_{i|M_i|}\}$ ,  $|M_i|$ : total number of pure strategies of player  $i$
  - **Mixed player strategy vector**  $x_i = (x_{i1}, x_{i2}, \dots, x_{i|M_i|})$ ,  $x_{ih}$ : probability of player  $i$  to choose its  $h$ -th pure strategy
    - $\Delta = \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 1, \text{ where } x_i \geq 0 \text{ for } i = 0, 1, \dots, n\} \rightarrow \text{probability simplex}$
    - $n$ : number of pure strategies of player  $i$
  - **Payoff** → value associated with a possible outcome of a game (set of strategies)
  - **Utility function**  $u_i : M_1 \times M_2 \times \dots \times M_N \rightarrow \mathbb{R}$ 
    - strategies → payoffs
    - combination of strategies played by all players
  - **Nash equilibrium** → set of strategy profiles in which each strategy is a **best response** to the strategy of the co-players and no player has the incentive to deviate from their decision (changing strategy  $\not\rightarrow$  payoff increase)
- **Classical game theory VS Evolutionary game theory**
  - Static VS Dynamic strategies
  - **Evolutionary game theory** → dynamics of strategy change through repeated games

## Theatina

Η θεωρία παιγνίων ασχολείται με την ανταγωνιστική αλληλεξάρτηση μεταξύ δύο ή περισσότερων αντιπάλων, ώστε να μελετηθεί και ιδανικά να προβλεφθεί ο τρόπος με τον οποίο γίνεται η λήψη των αποφάσεων, υπό συγκεκριμένες συνθήκες. Το παίγνιο, αποτελείται από ένα σύνολο παικτών, των αντίστοιχων τους στρατηγικών καθώς και τις αποδόσεις τους.

Οι παίκτες, ως αυτόνομη μονάδα ο καθένας και βασιζόμενοι σε κανόνες και πληροφορίες, προσπαθούν να λάβουν μια απόφαση που θα βελτιστοποιήσει τη δική τους ευημερία έναντι των αντιπάλων τους.

Οι κανόνες που ορίζουν τις εφικτές επιλογές των παικτών, αποτελεί το σύνολο των δυνατών **στρατηγικών** τους, που καθορίζονται είτε απόλυτα και ονομάζονται καθαρές στρατηγικές, είτε μέσω μιας κατανομής πιθανοτήτων επί των αντίστοιχων καθαρών και σε αυτή την περίπτωση ονομάζονται μικτές.

Η «απόδοση» εκφράζει το αριθμητικό αποτέλεσμα του παιγνίου για έναν συνδυασμό στρατηγικών.

Η λύση του παιγνίου ή αλλιώς ισορροπία κατά Nash, είναι ο συνδυασμός εκείνος των στρατηγικών όπου η απόδοση της αντίστοιχης στρατηγικής για κάθε παίκτη είναι η καλύτερη απόκριση στη στρατηγική των άλλων παικτών και κανένας από αυτούς δεν θα επωφεληθεί από κάποια αλλαγή (στρατηγικής του) .

**Η εξελικτική μέθοδος της θεωρίας παιγνίων** έχει δυναμικό χαρακτήρα και σε αντίθεση με την παραδοσιακή που ορίζει «λύση» σε παιγνιο που παίζεται μόνο μια φορά, εδώ οι στρατηγικές των παικτών ανταγωνίζονται μεταξύ τους καθώς ο χρόνος εξελίσσεται, προσομοιώνεται μια μέθοδος φυσικής επιλογής και τελικά επικρατούν αυτές με τα πιο σωστά «εξελικτικά» χαρακτηριστικά.

#Notes

Simplex -> πλέγμα που εκφράζει τον χώρο μεικτών στρατηγικών του παίκτη και οι γωνίες του τις καθαρές στρατηγικές του

# Proposed System

## Evolutionary game theory framework

- First attempt in the specific NLP task of WSD

## WSD

- Sense - labelling task → sense assignment to words
- Constraint satisfaction problem

## Game-theoretical approach

- **Players** → words ( to be disambiguated )
- **Strategies** → senses ( evolving population )
- **Payoff matrices** → sense similarity
- **Interactions** → weighted graph
- **Nash equilibrium** → consistent word-sense assignment
- **Selection process**
  - Iterative process of **fitness increase** ( candidates / senses with certain features )
- **Best candidates** ( senses with higher fitness ) in the population

## Theatina

Το σύστημα που προτείνει η παρούσα έρευνα λοιπόν, εφαρμόζει για πρώτη φορά τεχνικές της θεωρίας παιγνίων και συγκεκριμένα της εξελικτικής, με σκοπό την επίλυση του προβλήματος της ΑΑΛ.

Για να μπορέσει να μοντελοποιηθεί, μετατρέπεται σε εργασία εννοιολογικής (σημασιολογικής) επισημείωσης, με σκοπό τη συμμόρφωση σε προκαθορισμένους περιορισμούς.

Σε όρους θεωρίας παιγνίων, έχουμε τις αντιστοιχίες που φαίνονται στη διαφάνεια. Για παράδειγμα, αντί για τις γενικές έννοιες «παίκτης», «στρατηγική» και «απόδοση», έχουμε αντίστοιχα τη «λέξη», την «έννοια» και την «ομοιότητα» μεταξύ αυτών.

Όπως προαναφέρθηκε, κατά την επαναληπτική εκτέλεση παιγνίων μεταξύ των λέξεων, λαμβάνει χώρα μια διαδικασία επιλογής των πιο κατάλληλων εννοιών, οι οποίες επικρατούν έναντι των λιγότερο κατάλληλων, οδηγώντας στην ανάθεση μίας μόνο έννοιας σε κάθε λέξη

# Proposed System

**Consistent final labelling of the data**

**The solution of the problem is always found**

- System convergence to the nearest Nash equilibrium  
(Nash Theorem 1951)

**Most appropriate sense association**

- Target word

**WSD**

- Continuous optimization problem
- Exploitation of contextual information in a dynamic way  
(evolutionary game theoretic framework)

**Versatile approach**

- Adaptive to different scenarios and tasks
- Unsupervised / Semi-supervised

Julie

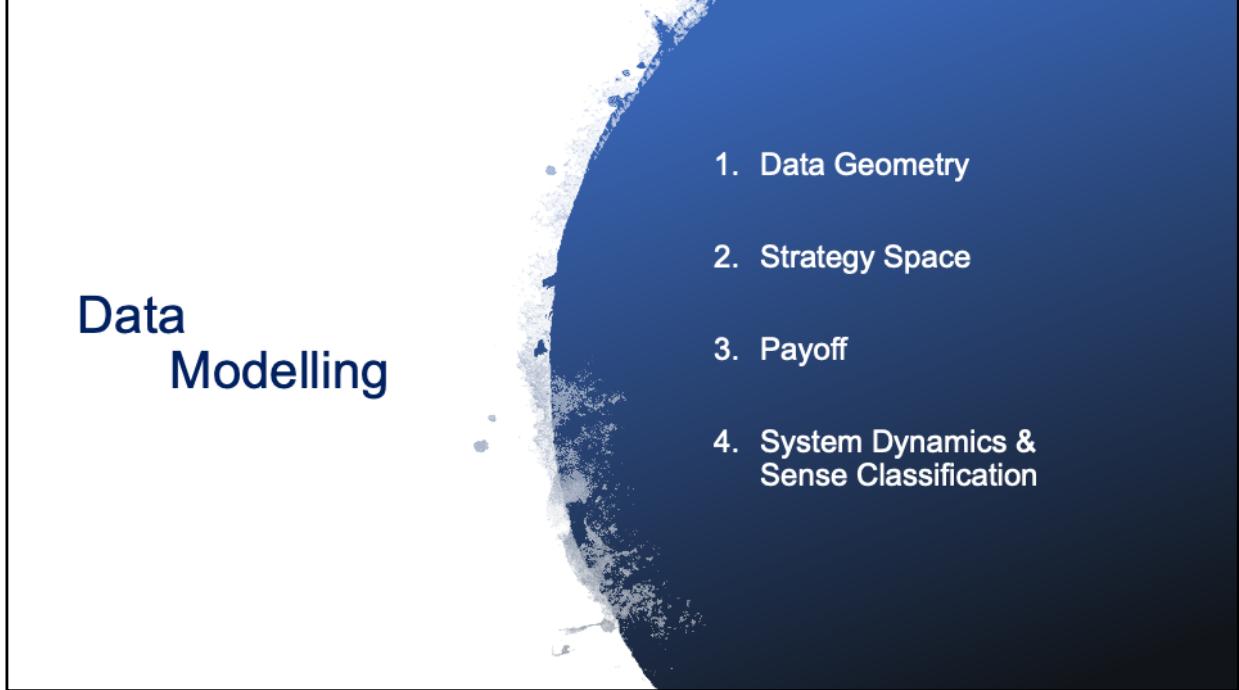
Η συγκεκριμένη προσέγγιση βασίζεται στην σταθερή επισημείωση των δεδομένων αλλά και στο Θεώρημα του Nash κατά το οποίο για κάθε πεπερασμένο παίγνιο υπάρχει τουλάχιστον ένα σημείο ισορροπίας ( αν όχι με καθαρές, με μεικτές στρατηγικές ), οπότε και η λύση του προβλήματος επιτυγχάνεται σε κάθε περίπτωση (Θεώρημα Nash, 1951).

Επίσης, κάθε λέξη σχετίζεται με το περικείμενό της.

Λόγω της θεωρίας παιγνίων, το task της ΑΑΛ, μετατρέπεται σε πρόβλημα (-κατάσταση) δυναμικής βελτιστοποίησης, εκμεταλλευόμενη τις πληροφορίες του περικειμένου.

Το προτεινόμενο σύστημα είναι πολύπλευρο, καθώς μπορεί να χρησιμοποιηθεί και από μη επιβλεπόμενα μοντέλα μάθησης αλλά και από ημι-επιβλεπόμενα.

# Data Modelling

- 
1. Data Geometry
  2. Strategy Space
  3. Payoff
  4. System Dynamics & Sense Classification

Julie

Η διαδικασία μοντελοποίησης των δεδομένων, έγινε βηματικά και περιγράφεται στις επόμενες διαφάνειες.

# Data Geometry

- List  $I$  of  $N$  words from the text
  - $I = \{i_1, i_2, \dots, i_N\}$
- Word similarity matrix  $W_{N \times N}$ 
  - Pairwise similarities among words → players' interactions
  - $w_{kj} = \text{sim}(i_k, i_j), \forall k, j = 1, 2 \dots, N \text{ and } k \neq j$
  - Weighted adjacency matrix of the graph
- Similarity measure: strength of co-occurrence between words  $i, j$ 
  - Semantically correlated words
  - 8 association measures: Dice coefficient (dice), modified Dice coefficient (mDice), pointwise mutual information (pmi), t-score measure (t-score), z-score measure (z-score), odds ratio (odds-r), chi-squared test (chi-s), chi-squared correct (chi-s-c).
- Proximity relations with  $n$ -neighbours (similarity augmentation)
  - Sentence structure
  - Size of  $n$ : fixed expressions / semantic concepts



Julie

Αρχικά, συλλέγονται Ν λέξεις, των οποίων υπολογίζεται η συσχέτιση ανά δύο, δημιουργώντας έτσι έναν πίνακα γειτνίασης ενός γράφου με βάρη, μέσω 8 διαφορετικών μεθόδων, η απόδοση των οποίων και αξιολογείται στη συνέχεια.

Έπειτα, η ομοιότητα μεταξύ ζευγών λέξεων επαυξάνεται με βάση τη συντακτική δομή της πρότασης και την απόσταση της κάθε λέξης από τους ο γείτονές της (δεξιά και αριστερά).

Όπως είναι φυσικό, με μικρές τιμές του η αποτυπώνονται στερεότυπες φράσεις, ενώ αντίθετα με μεγάλες τιμές του η, ευρύτερες σημασιολογικές ιδέες.

# Data Geometry

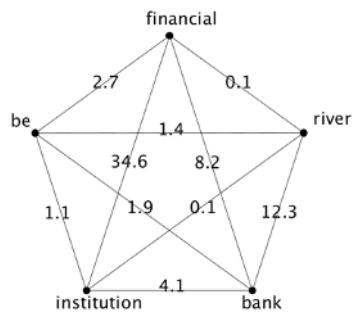


Figure 1. Co-occurrence graph

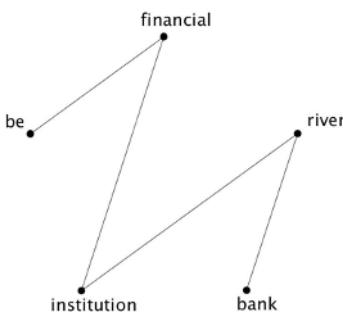


Figure 2. n-gram graph ( $n=1$ )

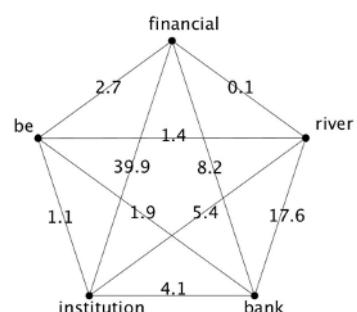


Figure 3. Similarity n-gram graph

Julie

Οι διαδικασίες που περιγράφηκαν προηγουμένως, παράγουν τα 3 διαγράμματα που βλέπουμε.

Το σχήμα 1 αποτελεί γραφική αναπαράσταση του πίνακα συσχέτισης μεταξύ των λέξεων, το σχήμα 2 αναπαριστά τις σχέσεις τους με τις η γειτονικές τους λέξεις (εδώ  $n=1$ ) και ο γράφος της ενισχυμένης ομοιότητας μεταξύ των γειτόνων τους που φαίνονται στο n-gram, όταν αυτές δεν περιλαμβάνουν stop-words ( πχ. το "be" ), παρουσιάζεται στο σχήμα 3.

# Strategy Space



- **Sense inventories** of each word  $i \in I$ 
  - $M_i = \{m_{i1}, m_{i2}, \dots, m_{ik}\}$ , where  $k = |M_i|$  the number of **senses** associated with word  $i$
  - WordNet and BabelNet as knowledge base
- **Unique concepts** in sense inventories
  - Game space  $C = \{c_1, c_2, \dots, c_k\}$ , where  $k = |C|$  the total number of unique concepts
  - Assignment of **probability distribution** over the senses in  $C$ , for each word  $i \in I$
- **Player strategy space**
  - $s_p = (s_{p1}, s_{p2}, \dots, s_{pn})$ ,  $n = |C|$
  - $s_{pi}$  : **probability** that player  $p$  chooses the  $i - th$  sense / concept
  - Graphical representation: regular **polygon** of radius 1, where the distance from the centre to any node - mixed strategy, represents the probability associated with a particular word sense - pure strategy ( for  $i \in M_i$  )



## Theatina

Στο επόμενο βήμα συλλέγονται οι κέννοιες που συνδέονται με την εκάστοτε λέξη  $i$ , ώστε να δημιουργηθεί ο χώρος στρατηγικής του παιχνιδιού, αποτελούμενος από την ένωση των προαναφερθέντων συνόλων και να υπολογιστεί η πιθανότητα επιλογής της εκάστοτε έννοιας από την  $i$ -οστή λέξη.

Ο κάθε παίκτης διαθέτει ένα σύνολο μεικτών στρατηγικών, το οποίο εκφράζεται ως ένα διάνυσμα  $s_p$  διάστασης  $n$ , δηλαδή με ο μη αρνητικές συντεταγμένες, οι οποίες αθροίζουν στη μονάδα, όπως θα δούμε αργότερα και εκφράζουν την πιθανότητα επιλογής της αντίστοιχης έννοιας με την οποία αυτή η λέξη σχετίζεται.

Για την καλύτερη κατανόηση του χώρου των στρατηγικών, θα δούμε τη γραφική αναπαράστασή του, η οποία αποδίδεται ως ένα κανονικό πολύγωνο ακτίνας 1, με τις ακμές που ενώνουν το κέντρο του με κάθε κορυφή να δηλώνουν την πιθανότητα επιλογής της έννοιας που αντιπροσωπεύει η αντίστοιχη κορυφή.

# Strategy Space



- **System strategy space**  $S_{N \times |C|}$  initialization
  - $s_{ih}$  : **probability** of player  $i$  to choose the  $h - th$  pure strategy to play
  - **Uniform** distribution
    - **unsupervised** learning: no prior knowledge
  - $s_{ih} = \begin{cases} |M_i|^{-1}, & \text{if sense } h \text{ is in } M_i \\ 0, & \text{otherwise} \end{cases}$
  - **Geometric** distribution
    - **semi-supervised** learning: exploits information from prior knowledge
  - $pr_{ih} = \begin{cases} p(1-p)^{r_h}, & \text{if sense } h \text{ is in } M_i \\ 0, & \text{otherwise} \end{cases}$ 
    - ✓ **higher** probability on senses that have a **high frequency**  $\rightarrow r_h$  : sense rank
  - $pr_{ih} = \begin{cases} p(1-p)^{r_o}, & \text{if sense } h \text{ is in cluster } o \\ 0, & \text{otherwise} \end{cases}$ 
    - ✓ **equal** probability to the senses belonging to a determined **cluster**  $\rightarrow r_o$  : cluster rank
  - $s_{ih} = \frac{pr_{ih}}{\sum_{j \in M_i} pr_{ij}} \rightarrow \sum_{h \in C} s_{ih} = 1$

## Theatina

Για την αρχικοποίηση λοιπόν του χώρου στρατηγικών του παιγνίου, αποδίδεται σε κάθε λέξη μια τιμή πιθανότητας επιλογής των εννοιών της.

Στην περίπτωση της μη επιβλεπόμενης μάθησης, ακολουθείται η ομοιόμορφη κατανομή, καθώς δεν υπάρχει πρότερη γνώση για τις προτιμήσεις των λέξεων σχετικά με την επιλογή εννοιών, ενώ στην ημι-επιβλεπόμενη μάθηση, αξιοποιούνται πληροφορίες από εννοιολογικά επισημειωμένα δεδομένα και ακολουθείται η γεωμετρική κατανομή.

Με αυτό τον τρόπο, είτε λαμβάνουν μεγαλύτερη πιθανότητα οι συχνότερες έννοιες που αφορούν την i-οστή λέξη, είτε μετά από τεχνικές συσταδοποίησης, αποδίδονται ίσες τιμές πιθανότητας στις έννοιες εντός μιας συστάδας.

## #Notes

Ομοιόμορφη κατανομή ( μέση τιμή:  $(a+b)/2$ , διασπορά:  $(n^2 - 1)/12$  ),  $n=οι\;έννοιες$   
Μικρό rank  $r_h$  μεγάλη συχνότητα, και όσο μικραίνει το rank, τόσο μεγαλώνει και η πιθανότητα γιατί είναι φθίνουσα κατανομή ( μέση τιμή:  $(1-p)/p$ , διασπορά:  $(1-p)/(p^2)$  )

Όσο μικραίνει το  $\rho$ , τόσο μεγαλύτερη διασπορά τιμών έχουμε  
Διαιρούμε με το άθροισμα του συνόλου των τιμών, ώστε να αθροίζει στη μονάδα

## Payoff

- **Sense similarity matrix**  $Z_{|C| \times |C|}$ 
  - Pairwise similarity among senses in  $C \rightarrow$  partial payoff matrices of each game
  - $z_{ij} = ssim(c_i, c_j), \forall c_i, c_j \in C, i \neq j$
- **Partial payoff matrix**  $Z_{|M_i| \times |M_j|}$ :
  - Games played between two words  $i, j$
  - Dimension  $|M_i| \times |M_j|, |M_k|$ : cardinality of senses set of word  $k$
- **Semantic similarity**
  - Relations of **likeness** ("is-a")
  - wup: path depth of senses  $c_i, c_j$  and msa
    - $ssim_{wup}(c_i, c_j) = 2 * \frac{\text{depth(msa)}}{\text{depth}(c_i) + \text{depth}(c_j)}$
    - msa : "most specific ancestor"
  - jcn measure: corpus statistics and structural properties
    - $ssim_{jcn}(c_i, c_j) = IC(c_i) + IC(c_j) - 2 IC(msa)$
    - $IC(c) = -\log(p(c)) = \log \frac{1}{p(c)}$
    - Information Content (IC): level of "surprise" of a particular outcome

Julie

Οι ομοιότητες ανά ζεύγη εννοιών χρησιμοποιούνται για να φτιαχτεί ο πίνακας  $Z$ . Οι αποδόσεις (των στρατηγικών) για κάθε παιχνίδι ανάμεσα σε μια λέξη  $i$  και  $j$ , λαμβάνονται από τον αντίστοιχο υποπίνακα  $Z_{|M_i| \times |M_j|}$ .  
( διάστασης  $m_i \times m_j$  όπου τα  $m_i$  και  $m_j$  είναι ο πληθικός αριθμός των συνόλων εννοιών  $M_i$  και  $M_j$ , αντίστοιχα. )

Για τον υπολογισμό της σημασιολογικής ομοιότητας των εννοιών, χρησιμοποιήθηκαν τα μέτρα wup και jcn.  
Αυτά, υπολογίζουν την ομοιότητα μεταξύ των δύο εννοιών  $c_i, c_j$  σύμφωνα με το βάθος των δύο εννοιών στη λεξική βάση δεδομένων και τον πιο συγκεκριμένο πρόγονο κόμβο (msa) των δύο εννοιών.

Το Information Content είναι η πληροφορία ενός περιεχομένου μιας ευρύτερης έννοιας  $c$  που προέρχεται από ένα corpus και πρόκειται για εναλλακτική έκφραση πιθανότητας που δηλώνει το βαθμό έκπληξης σχετικά με την εμφάνιση μιας λέξης.

## #Notes

Οι γραμμές επιλέγονται με βάση τις έννοιες που σχετίζονται με τη λέξη i και οι στήλες με τις έννοιες που σχετίζονται με τη λέξη j.

Ο πινακας μερικής απόδοσης περιέχει παιχνίδια ανάμεσα σε 2 λέξεις i, j.

## Payoff

- Semantic **relatedness**
  - Similarity among the **definitions** of two concepts
  - **Wider range** of relations ( "is-a-part-of", "is-the-opposite-of" )
  - Definitions derived from glosses of the synsets in WordNet
  - **Co-occurrence** vector  $v_i = (w_{i1}, w_{i2}, \dots, w_{in})$ ,  $i$  : concept,  $w$  : word gloss occurrences,  $n$  : total words
- **Cosine similarity**
  - $sim(v_i, v_j) = \cos \theta = \frac{v_i \cdot v_j}{\|v_i\| \|v_j\|}$ ,  $i, j$  : concepts
  - $\|v_i\| = \sqrt{\sum_{j=1}^n w_{ij}^2}$
  - cosine of the angle  $\theta$  between the two co-occurrence vectors  $v_i, v_j$
- **4 variations** (construction of vectors)
  - Difference in co-occurrence calculation, corpus use and relation source
  - tf-idf, tf-idf\_ext, vec and vec\_ext

Julie

Η συσχέτιση, είναι η ομοιότητα ανάμεσα στους ορισμούς δύο ευρύτερων εννοιών, οπότε και τα μέτρα συσχέτισης είναι πιο γενικά.

Οι ορισμοί των εννοιών χρησιμοποιούνται για να φτιάξουν ένα διάνυσμα συνεμφάνισης  $v_i = (w_{i1}, w_{i2} \dots, w_{in})$  όπου για κάθε έννοια  $i$ , το  $w$  αντιπροσωπεύει τον αριθμό των φορών που εμφανίζεται η λέξη και η  $i$  είναι ο συνολικός αριθμός διαφορετικών λέξεων (τύπων) στο corpus. Αυτή η αναπαράσταση επιτρέπει την προβολή τους στο διανυσματικό χώρο για τη διεξαγωγή πειραμάτων και μελετών. Στα πειράματα της παρούσας έρευνας, χρησιμοποιήθηκε το μέτρο ομοιότητας συνημιτόνου της γωνίας  $\theta$  ανάμεσα στα διανύσματα συνεμφάνισης δύο διαφορετικών ευρύτερων εννοιών  $i, j$ , η οποία εκφράζει τις κοινώς συνεμφανιζόμενες λέξεις των 2 concepts.

#Notes

Και  $|| v ||$  είναι η νόρμα των διανυσμάτων.

Όσο περισσότερο μεγαλώνει η γωνία τόσο μεγαλύτερη γίνεται κάθε φορά η

ομοιότητα.

Οι ορισμοί των οποίων προέρχονται από τα glosses των synsets [ όπου gloss είναι μια σύντομη σημειογραφία, της έννοιας μιας λέξης και synset είναι ένα σύνολο λέξεων που μοιράζονται παρόμοιες έννοιες, δηλαδή ένα σύνολο συνωνύμων].

Συνημίτονο της γωνίας θ ανάμεσα σε δύο διανύσματα συνεμφάνισης δύο διαφορετικών λέξεων  $v_i$ ,  $v_j$ .

Στην περίπτωσή μας, επιστρέφει τιμές που κυμαίνονται από 0 έως 1 επειδή οι τιμές στα διανύσματα συνεμφάνισης είναι όλες θετικές

Όσο πιο κοντά στο 1 τόσο πιο όμοια είναι τα διανύσματα.

# System Dynamics & Sense Classification

- Computation of the **Nash equilibrium**
  - **Payoff of sense  $h$  ( word  $i$  )**
    - $u_i(e^h, x) = \sum_{j \in N_i} (w_{ij} Z_{ij} x_j)_h$ , where  $N_i$  the neighbours (context) of word  $i$
    - Pure strategy  $h : e^h$
    - Similarity with word  $j : w_{ij}$
    - Similarities among senses of  $i, j : Z_{ij}$
    - Sense preference of word  $j : x_j (s_{jh})$
  - **Average payoff of player (  $u_i(x)$  )**
- **Replicator Dynamics Equation**
  - $x_i^h(t+1) = x_i^h(t) \frac{u_i(e^h, x)}{u_i(x)}$ ,  $\forall h \in M_i$  ( $x_i^h : s_{jh}$ )
  - Player strategy update
  - Depends on performance → measured relatively to the average payoff

## Theatina

Όπως αναφέρθηκε και στις εισαγωγικές διαφάνειες, για την εύρεση του σημείου ισορροπίας του παιγνίου, πραγματοποιείται μια διαδικασία επιλογής των επικρατέστερων εννοιών.

Αναλυτικότερα, σε κάθε επανάληψη, κάθε λέξη «παίζει» ένα παιχνίδι με το σύνολο των γειτονικών της.

Με αυτό τον τρόπο ενημερώνει την πιθανότητα επιλογής της κάθε συσχετιζόμενης με αυτήν έννοιας, λαμβάνοντας υπόψιν την ομοιότητα με τις γειτονικές της, την ομοιότητα μεταξύ των εννοιών τους καθώς και τις προτιμήσεις του συμπαίκτη της για τη συγκεκριμένη έννοια.

Καθώς λοιπόν εξελίσσεται ο χρόνος, σε κάθε επόμενη χρονική στιγμή οι λέξεις «μιμούνται» επιτυχημένες στρατηγικές συμπαικτών τους, δηλαδή αυξάνουν την πιθανότητα επιλογής μιας έννοιας μέσω της εξίσωσης αντιγραφής, δίνοντας μεγαλύτερο βάρος σε εκείνη που επιφέρει διαφορά απόδοσής της από τη μέση με θετικό πρόσημο, ενώ παράλληλα μειώνει την πιθανότητα στην αντίθετη περίπτωση (όταν δηλαδή η διαφορά από τη μέση απόδοση έχει αρνητικό πρόσημο).

Έτσι, ενισχύονται και επικρατούν συνολικά οι έννοιες εκείνες που προσφέρουν μεγαλύτερο κέρδος στις αντίστοιχες λέξεις.

#Notes

$e^h \rightarrow$  μοναδιαίο διάνυσμα

$u_i(x) = \sum_{j \in N_i} x_i^T (w_{ij} Z_{ij} x_j) \rightarrow$  average fitness of population i

# System Dynamics & Sense Classification

## • Classification

- Mixed strategy  $s_{ih} \in S \rightarrow$  each word  $i \in I, h \in M_i$
- The strategy with the **highest probability** is chosen
  - $\varphi_i = \operatorname{argmax}_{h \in [1, 2, \dots, |C|]} s_{ih}$ , where  $|C|$  the total number of senses
  - Each word  $i \in I$  is mapped to exactly one sense  $c \in C$
- Word **fails** to be disambiguated
  - Unable to update its strategy space
  - Strategy space initialized with a uniform distribution
    - Zero entry payoff ( no similar senses with co-players)
    - Not connected with other nodes in the graph

Julie

Τελικά, μετά το πέρας της προηγούμενης επαναληπτικής διαδικασίας, προκύπτει μία και μόνο επικρατούσα έννοια που αποδίδεται σε κάθε λέξη-παίκτη.

Αυτό επιτυγχάνεται μέσω της συνάρτησης argmax, η οποία επιστρέφει την έννοια με τη μεγαλύτερη πιθανότητα από το διάνυσμα των μεικτών στρατηγικών της κάθε λέξης.

Επίσης, όπως πειραματικά αποδείχθηκε, σε ελάχιστες περιπτώσεις δεν επιτυγχάνεται η επίλυση της αμφισημίας μιας λέξης, λόγω αδυναμίας ενημέρωσης του χώρου των μεικτών στρατηγικών της, γεγονός που μπορεί να οφείλεται για παράδειγμα στην ανυπαρξία κοινών έννοιών με άλλες λέξεις.

# System Dynamics & Sense Classification

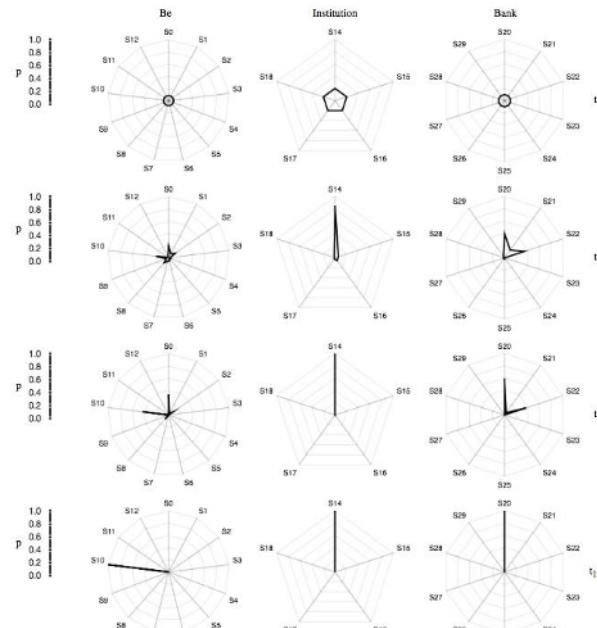


Figure 4. System dynamics for words "be", "institution" and "bank"



(1. Για την καλύτερη κατανόηση του χώρου των στρατηγικών, θα δούμε τη γραφική αναπαράστασή του, η οποία αποδίδεται ως ένα κανονικό πολύγωνο ακτίνας 1, με τις ακμές που ενώνουν το κέντρο του με κάθε κορυφή να δηλώνουν την πιθανότητα επιλογής της έννοιας που αντιπροσωπεύει η αντίστοιχη κορυφή. )

Theatina

2. Όπως είδαμε και πριν, στο σχήμα αυτό παρουσιάζεται γραφικά ο στρατηγικός χώρος των λέξεων, δηλαδή τα κανονικά πολύγωνα με τις αντίστοιχες πιθανότητες επί των συσχετιζόμενων έννοιών τους.

Εδώ λοιπόν γίνεται πιο ξεκάθαρα αντιληπτή η εξέλιξη των στρατηγικών στο χρόνο που αναφέρθηκε στην προηγούμενη διαφάνεια, για τις λέξεις "be", "institution" και "bank".

Συγκεκριμένα, στη χρονική στιγμή t1, δηλαδή την 1η γραμμή, όλες οι έννοιες που σχετίζονται με τις προηγούμενες λέξεις είναι ισοπίθανες.

Στο επόμενο χρονικό βήμα t2 απαλοίφονται αρκετές έννοιες, είτε λόγω μικρής συσχέτισης, είτε λόγω κυρίαρχης επίδρασης κάποιων άλλων, όπως ισχύει για τη λέξη «institution», η οποία λόγω των λέξεων «financial» & «bank», κατευθύνεται αμέσως προς μια συγκεκριμένη έννοια (οικονομική).

Ένα ενδιαφέρον παράδειγμα από την άλλη, είναι εκείνο της αμφίσημης λέξης «bank», η οποία

"παιζοντας παιχνίδια" με τις λέξεις "financial" & "institution" κατευθύνεται προς την οικονομική έννοια, ενώ όταν παίζει με τη λέξη «river», ενισχύεται η περιβαλλοντική / νατουραλιστική της έννοια.

Στο χρονικό βήμα t12 το σύστημα τελικά συγκλίνει (δηλαδή οι αλλαγές στις καταστάσεις του παιχνιδιού είναι πλέον αμελητέες), και η έννοια της αμφίσημης λέξης «bank» με τη μεγαλύτερη πιθανότητα καταλήγει να είναι η περιβαλλοντική (s20) έναντι της οικονομικής (s22), επιλογή που όπως περιγράφηκε σε προηγούμενες διαφάνειες, οφείλεται σε ποικίλους παράγοντες και μέτρα, ισχυροποιώντας τη σταθερότητα του τελικού αποτελέσματος.

## Parameter Tuning

- Two datasets to **evaluate** our algorithm in **different scenarios**
- From each data set
  - 50 different data sets to simulate a situation in which the system must work on texts of different sizes and on different domains
- The **parameters** that will be **tuned**
  - Association, similarity and relatedness measures to weight the similarity among word and senses
  - The  $n$ -gram graph to increase the weights of near words ( proximity )
  - The  $p$  of the geometric distribution used by their semi-supervised system

### Theatina

Οι μεταβλητές που πρόκειται να παραμετροποιηθούν είναι:

1. Τα μέτρα που αφορούν τη συσχέτιση και τη σημασιολογία, ώστε να σταθμιστεί η ομοιότητα ανάμεσα σε λέξεις και έννοιες.
2. Το  $n$ -gram γράφημα που αξιοποιείται για την αύξηση των βαρών των κοντινών λέξεων (proximity-εγγύτητα) και
3. Η παράμετρος  $p$  της γεωμετρικής κατανομής, από την οποία εξαρτάται η διασπορά των τιμών.

## Parameter Tuning - Results

### 1. Association, Semantic & Relatedness Measures

- The **relatedness** measures perform significantly **better** than the semantic **similarity** measures
- Particularly **suitable** measures for the algorithm
  - Modified Dice coefficient ( **mdice** )
  - Term Frequency – Inverse Document Frequency ( **Tf-idf** )

### 2. **n** - gram Graph

- Highest results on both data sets using  **$n = 5$**  ( **n** nearest neighbours)

### 3. **Geometric** Distribution:

- Best results obtained with  **$p = 0.4$**

Theatina

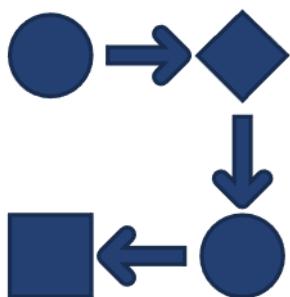
Τα αποτελέσματα, δείχνουν ότι όσον αφορά την σχετικότητα και την ομοιότητα, τα μέτρα της συσχέτισης (relatedness) λειτουργούν καλύτερα απ' ότι αυτά της σημασιολογικής ομοιότητας (semantic) .

Συγκεκριμένα, τα mdice & tf-idf αποδείχθηκαν ως τα πιο κατάλληλα.

Όσον αφορά το n-gram, όταν το  $n=5$  (5 κοντινότερες λέξεις - neighbours), τότε τα αποτελέσματα είναι τα βέλτιστα.

Τέλος, σχετικά με τη γεωμετρική κατανομή, τα καλύτερα αποτελέσματα παρουσιάζονται με  $p=0.4$

## Evaluation set-up



- Results are provided as balanced F-score ( $F_1$ ) measure
  - $F_1$ : determines the weighted harmonic mean of precision and recall
  - $$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \times 100 \quad (\%)$$
- Evaluation based on
  - Experiments with WordNet as Knowledge base
    - The best performance on nouns on all the data sets
    - Low results on verbs on all data sets
    - The use of prior knowledge is beneficial in general domain data sets
  - Experiments with BabelNet as Knowledge base
    - We used BabelNet to collect the sense inventories of each word to be disambiguated and NASARI to obtain the semantic representation of each sense
    - This data set contains highly ambiguous mentions

Julie

Στη συνέχεια για να λάβουμε τα ποσοστά της αξιολόγησης χρησιμοποιείται το μέτρο F-score και συγκεκριμένα το balanced, δηλαδή το F1, το οποίο καθορίζει τον σταθμισμένο αρμονικό μέσο της ακρίβειας και της ανάκλησης.

Αρχικά έγιναν πειράματα που είχαν ως knowledge base είτε το wordnet είτε το babelnet.

Τα κύρια αποτελέσματα είναι αυτά που φαίνονται στη διαφάνεια.

#Notes

Typically, it is appropriate for situations when the average of rates is desired.

[Όταν χρησιμοποιούταν το Wordnet:

την καλύτερη απόδοση είχαν τα ουσιαστικά και την χειρότερη τα ρήματα

ακόμα, η χρήση προηγούμενης γνώσης ήταν ευνοϊκή για τα dataset του γενικού τομέα τα καλύτερα αποτελέσματα είχαν τα semi-supervised

Όταν χρησιμοποιούταν το BabelNet:

φάνηκε να αποδίδει καλύτερα σε σχέση με Wordnet όταν γίνεται λόγος για named entities

Χρησιμοποιησε το NASARI ως τρόπο λήψης της σημασιολογικής αναπαράστασης από κάθε έννοια

Και τέλος, αυτό το dataset που χρησιμοποιήθηκε εδώ περιέχει τις περισσότερες αμφίσημες λέξεις και είναι δύσκολο να αποσημασιοποιηθούν χωρίς τη χρήση μεγάλων και καλά οργανωμένων knowledge based αλγορίθμων]

## Comparison to state-of-the-art algorithms

### WordNet:

- Our unsupervised system performs better than any other unsupervised algorithm in all datasets
- The performance of our system is more stable on the four datasets
- The comparison with semi-supervised systems shows that our system always performs better than the most frequent sense heuristic when we use information from sense-labeled corpora

### BabelNet:

- The performance of our system is close to the results obtained with Babelfy on S13 and substantially higher on KORE50
- It is also difficult to exploit distributional information on this data set because the sentences are short and, in many cases, cryptic.

Julie

Έπειτα έγινε σύγκριση του κάθε αλγορίθμου με τα state of the art συστήματα, όπου άντως φάνηκε στα αποτελέσματα να λειτουργούν καλύτερα αλλά και να παρουσιάζουν μια σταθερότητα στα ποσοστά τους σε κάθε dataset.

Κάτι αλλο που μπορεί να παρατηρηθεί είναι ότι το semi-supervised μοντέλο, έχει κοντινές τιμές με τα άλλα supervised συστήματα.

→ τα δείχνουμε παράλληλα με την εικόνα στην Διαφάνεια 21

## Comparison to state-of-the-art algorithms

	S7CG	S7CG (N)	S7	S3	S2
unsup.	<i>Nav10</i>	—	—	43.1	52.9
	<i>PPR<sub>w2w</sub></i>	80.1	83.6	41.7	57.9
	<i>WSD<sub>games</sub></i>	<b>80.4*</b>	<b>85.5</b>	<b>43.3</b>	<b>59.1</b>
semi sup.	<i>IRST-DDD-00</i>	—	—	58.3	—
	<i>MFS</i>	76.3	77.4	54.7	62.8
	<i>MRF-LP</i>	—	—	50.6*	58.6
	<i>Nav05</i>	<b>83.2</b>	84.1	—	60.4
	<i>PPR<sub>w2w</sub></i>	81.4	82.1	48.6	63.0
	<i>WSD<sub>games</sub></i>	82.8	<b>85.4</b>	<b>56.5</b>	<b>64.7*</b>
sup.	<i>Best</i>	82.5	82.3*	<b>59.1</b>	65.2
	<i>Zhong10</i>	<b>82.6</b>	—	58.3	<b>67.6</b>

**Figure 5.** Experiments with WordNet. Comparison with state-of-the-art algorithms: unsupervised (*unsup.*), semi-supervised (*semi sup.*), and supervised (*sup.*).

	S13	KORE50
<i>WSD<sub>games</sub></i>	<b>70.8</b>	<b>75.7</b>
<i>Babelfy</i>	69.2	71.5
<i>SUDOKU</i>	66.3	—
<i>MFS</i>	<b>66.5*</b>	—
<i>PPR<sub>w2w</sub></i>	60.8	—
<i>KORE</i>	—	<b>63.9*</b>
<i>GETALP</i>	<b>58.3</b>	—

**Figure 6.** Experiments with BabelNet. Comparison with state-of-the-art algorithms on WSD and entity linking. The results are provided as F1 for S13 and as accuracy for KORE50.

Julie

# Conclusion

## A new method for WSD

Evolutionary Game Theory  
Similarity measures that perform better  
Continuation of knowledge-based, graph-based approaches

## WSD as a constraint-satisfaction problem

Consistency in the assignment of senses to related words  
Development of contextual coherence on the assignment of senses  
(characteristic missing in many state-of-the-art systems)

## Replicator dynamics equation

Best labelling assignment

## Versatile

Unsupervised  
Semi-supervised

## Competitive compared to state-of-the-art systems

Considers the influence of each word on the others  
Imposes sense compatibility among each sense before assigning a meaning  
The meaning of a word depends only on words that share a proximity relation  
and on those that enjoy a high distributional similarity

Julie

Κλείνοντας, θα συνοψίσουμε τις μέχρι τώρα πληροφορίες.

Έγινε λόγος για μια νέα μέθοδο που άρει την αμφισημία των λέξεων, μέσω της εξελικτικής θεωρίας παιγνίων η οποία θέτει κάποια μέτρα ομοιότητας που αποδίδουν καλύτερα και μπορεί να θεωρηθεί η συνέχιση των knowledge-based & graph-based αλγορίθμων.

Η αποσαφήνιση των εννοιών των λέξεων αντιμετωπίστηκε ως ένα διαρκές παιχνίδι [έλεγε πρόβλημα] περιορισμών και ικανοποίησης αυτών, οδηγώντας σε σταθερή συσχέτιση εννοιών και λέξεων

Η ανάθεση της κάθε έννοιας στην αντίστοιχη λέξη βασίστηκε στο περικείμενο (στη κειμενική συνοχή), χαρακτηριστικό που λείπει από τα state of the art συστήματα.

Η χρήση του replicator dynamics equation είναι υπεύθυνη για την καλύτερη ανάθεση έννοιας σε κάθε λέξη

Το σύστημα αυτό όπως ειπώθηκε είναι πολύπλευρο και ευπροσάρμοστο

Και τέλος θεωρείται πλήρως ανταγωνιστικό σε σχέση με τα προϋπάρχοντα συστήματα. Η σημασία μιας λέξης εξαρτάται μόνο από λέξεις που μοιράζονται μια σχέση εγγύτητας (βρίσκονται πιο κοντά) και από εκείνες που απολαμβάνουν υψηλή ομοιότητα.

“ After so much **talking** we got **hungry**, therefore a **toast** is what we need the most ! ”

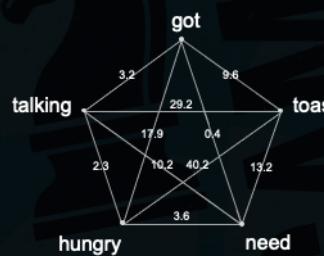


Figure 7. Co-occurrence graph

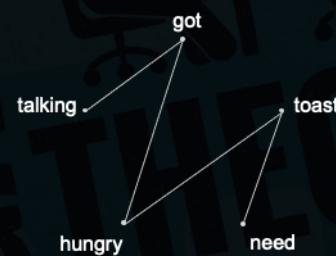


Figure 8.  $n$ -gram graph ( $n=1$ )



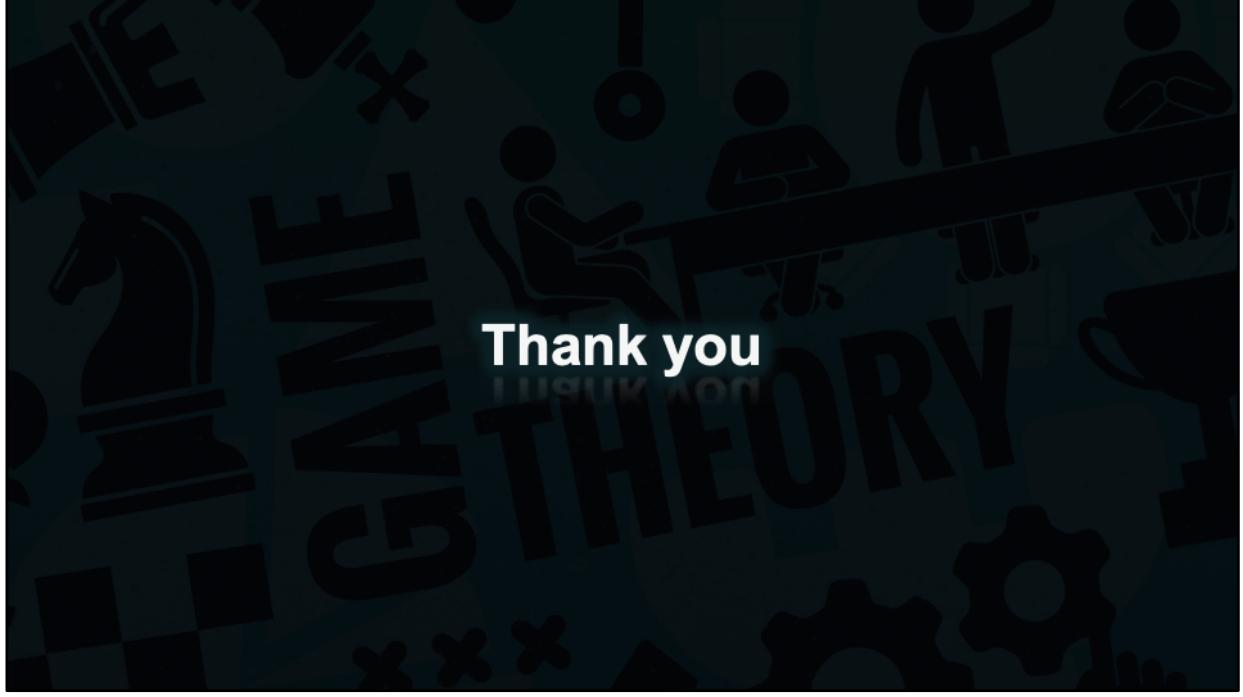
Figure 9. Similarity  $n$ -gram graph

After so much talking we got hungry, therefore a toast is what we need the most !

Context: talking, got, hungry, toast, need

Stop-words: got need

Ambiguous: toast



**Thank you**



## ! DISCLAIMER !

Αν και λατρέψαμε το θέμα και τον τρόπο προσέγγισής τους, αρκετές ήταν οι φορές που βρεθήκαμε σε αδιέξοδο λόγω εκφραστικών ή και μαθηματικών λαθών (τύποι, διαγράμματα, κλπ), όπως και ασάφεια ή έλλειψη συνοχής κατά την περιγραφή διαδικασιών και εννοιών, με ένα από τα κυριότερα προβλήματα να αποτελεί η έλλειψη συνέπειας στη σημειογραφία.

Επιπλέον, επειδή η θεωρία παιγνίων αποτελεί καινούριο αντικείμενο με το οποίο ασχολούμαστε πρώτη φορά, τα προαναφερθέντα λάθη είχαν μεγαλύτερη βαρύτητα.