ABSINTH
A small world approach to word sense induction.

Victor Zimmermann[1], Maja Hoffmann[2]
[1]Webis Group, Leipzig University
[2]Leipzig University of Applied Sciences (HTWK)

Wednesday, 14th September 2022
Konvens 2022

Preliminary summary

- Exploration of small world property of coöccurence networks.
- Transfer of sentiment propagation to word sense induction.
- Extension of Veronis (2004) [1].
- (This is workshopped from a student project and some of the larger limitations stem from that fact.)

- Word sense induction:
  - Task description
  - Graph-based approaches
  - Other approaches

- Word sense induction:
    - Task description
    - Graph-based approaches
    - Other approaches
- Root hub detection
    - Small world property
    - Hyperlex [1]
    - Minimum spanning trees

- Word sense induction:
  - Task description
  - Graph-based approaches
  - Other approaches
- Root hub detection
  - Small world property
  - Hyperlex [1]
  - Minimum spanning trees
- Root hub propagation
  - Sentiment propagation [2]
  - Toy example
  - Disambiguation

- Word sense induction:
  - Task description
  - Graph-based approaches
  - Other approaches
- Root hub detection
  - Small world property
  - Hyperlex [1]
  - Minimum spanning trees
- Root hub propagation
  - Sentiment propagation [2]
  - Toy example
  - Disambiguation
- ABSINTH
  - Experiments
  - Results
  - Limitations

Word sense induction

Wednesday, 14th September 2022
Konvens 2022

- Word sense induction on search results.
- Given query (search string) and list of 100 results (w/ title, url and snippet), cluster results by sense (from Wikipedia).

ID	47.6
url	*http://us.imdb.com/title/tt0120169/*
title	Soul Food (1997)
snippet	Directed by George Tillman Jr.. With Vanessa Williams, Vivica A. Fox,...

Table 1: Example dataset entry for 'soul food'.

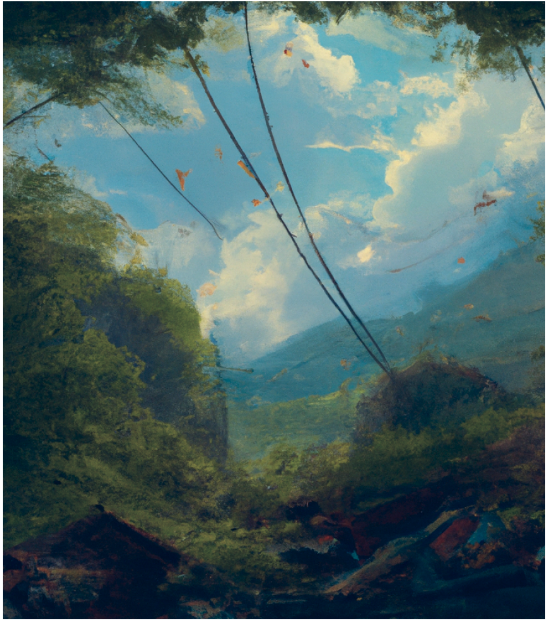
Identifying sense-components in coöccurence graphs:

- Hyperlex: Root hub detection & minimum spanning trees. [1]
- Chinese Whispers: Randomised spreading of senses through network. [3]
- SquaT++: Highly connected graph-patterns as stable senses. [4]

Topic models, vector-space segmentation, document encoding, etc:

- LDA, topic models. [5]
- Topic models + word2vec. [6]
- Bert, transformers. [7]

Root hub detection



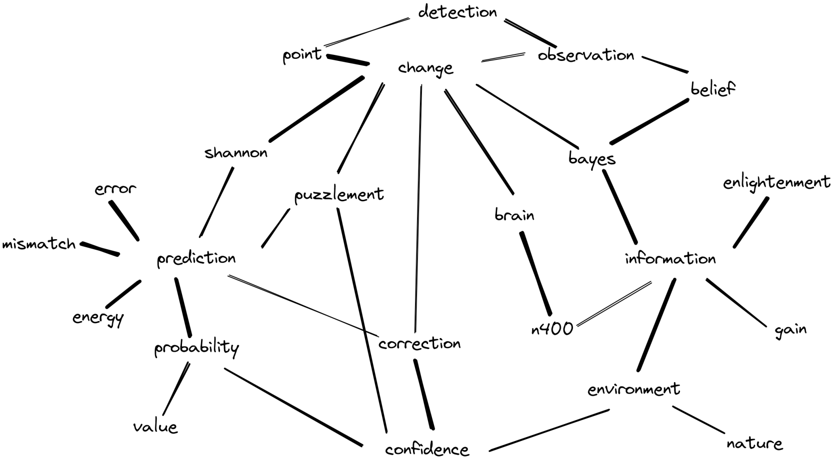
Wednesday, 14th September 2022
Konvens 2022

- Most nodes are not neighbours of each other (high clustering coëfficient: $C >> C_{rand}$).
- But high likelihood being the "neighbour of a neighbour" (short average path length: $L \sim L_{rand}$).
- Common in social science [8], political science [9], but also coöccurence networks [2].

- Still cited as "state-of-the-art" roughly until the advent of the transformer. [10][11]
- Root hub detection algorithm on (pruned) coöccurence graph:
- Step 1: Find node with highest number of neighbours[1], mark as root hub.
- Step 2: Remove root hub and all neighbours from network.
- Repeat step 1 and 2 as long as nodes with high enough degree remain.

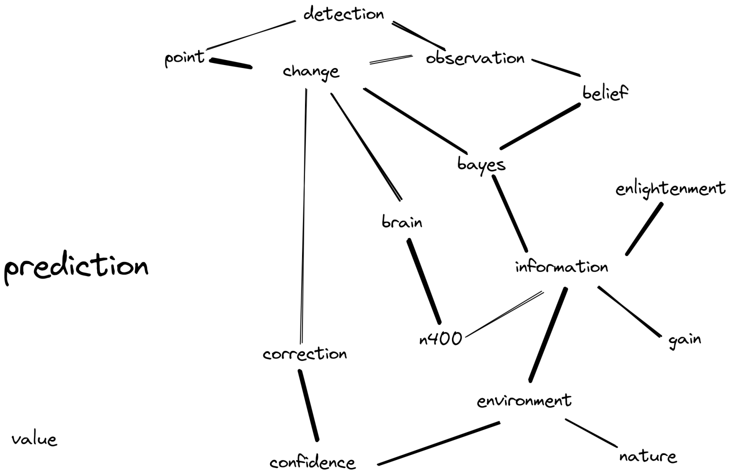
[1] and under a mean distance threshold (here 0.9.)

Toy example (root hubs): *surprise*



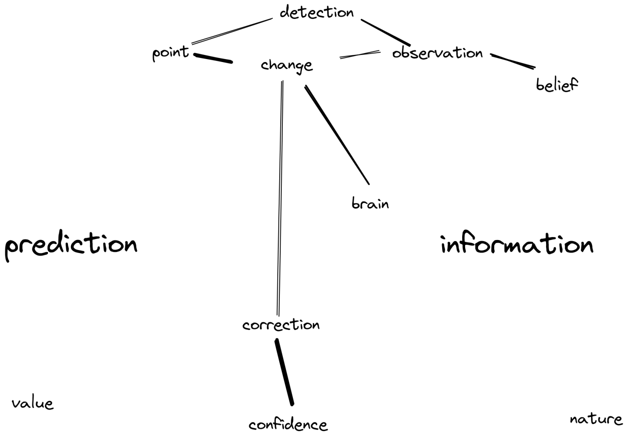
Zimmermann & Hoffmann (Konvens 2022)

Toy example (root hubs): *surprise*



Zimmermann & Hoffmann (Konvens 2022)

Toy example (root hubs): *surprise*



Zimmermann & Hoffmann (Konvens 2022)

Toy example (root hubs): *surprise*

change

belief

prediction

information

value

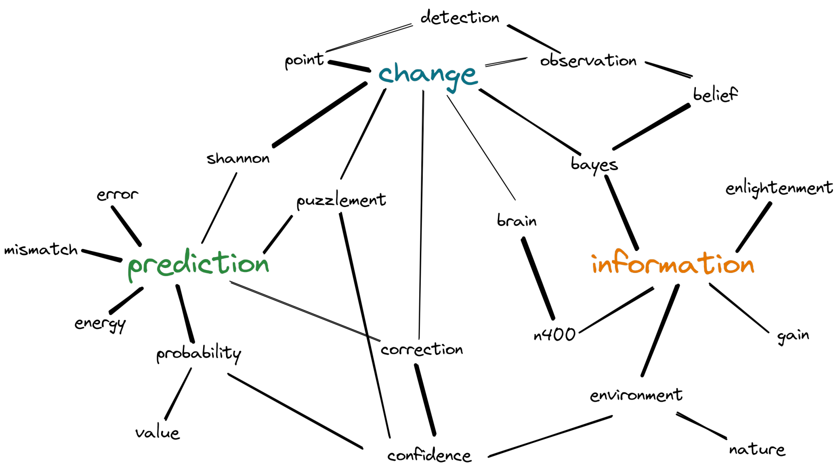
confidence

nature

Zimmermann & Hoffmann (Konvens 2022)

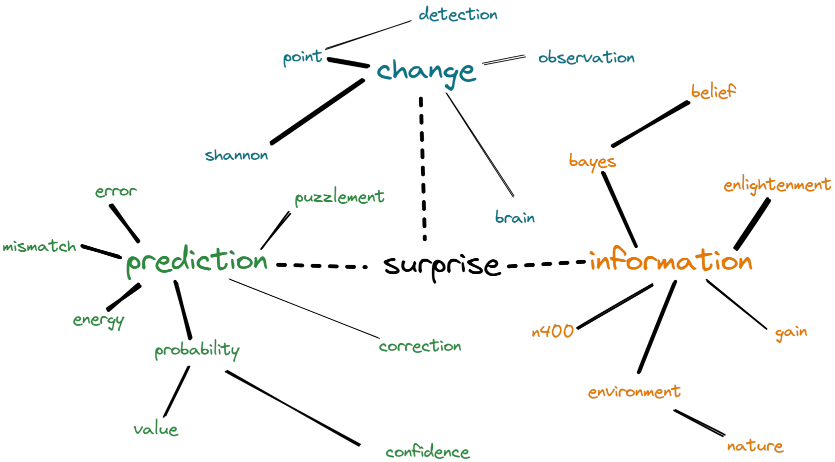
- Minimise graph to tree with minimal path length [12].
- Connect root hubs with new node at distance 0, apply MST algorithm.
- Resulting trees under root hubs represent sense lexicon. [1]

Toy example (minimum spanning tree): *surprise*



Zimmermann & Hoffmann (Konvens 2022)

Toy example (minimum spanning tree): *surprise*



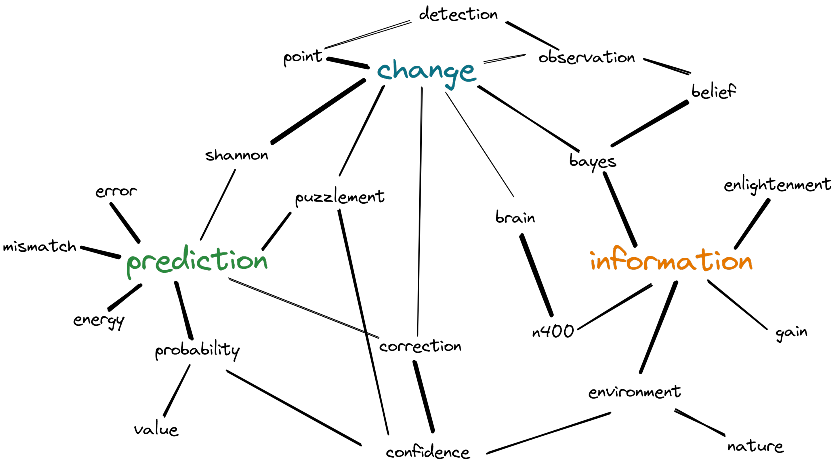
Zimmermann & Hoffmann (Konvens 2022)

Root hub propagation

Wednesday, 14th September 2022
Konvens 2022

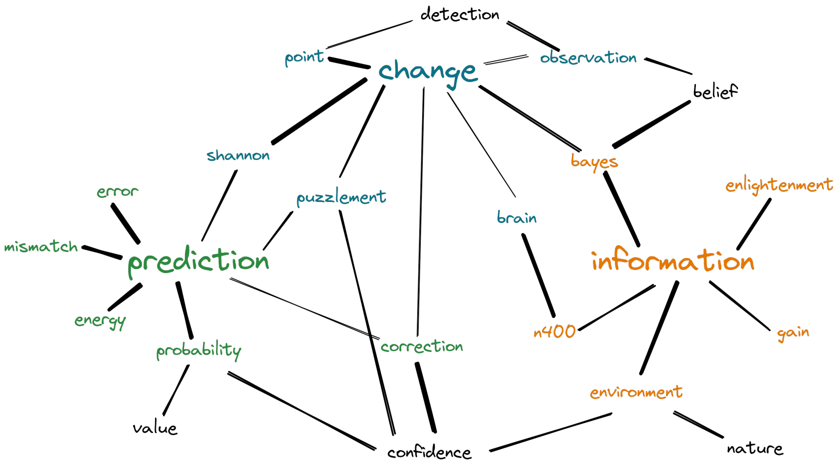
- Similar algorithm to Chinese Whispers [13].
- In sentiment: manual annotation of seeds [2].
- In WSI: root hubs as seeds.
- Step 1: sum edge-weighted senses of neighbours.
- Step 2: assign sense with highest value.
- Sense vector of node is the edge-weighted sum of its neighbours.

Toy example (label propagation): *surprise*



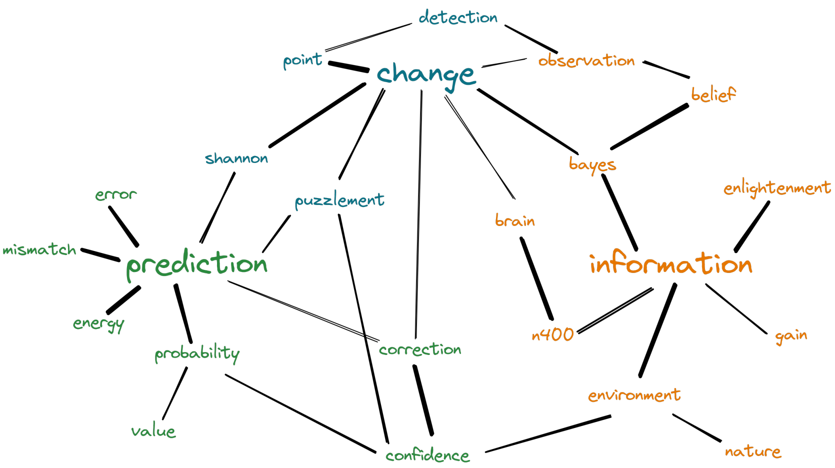
Zimmermann & Hoffmann (Konvens 2022)

Toy example (label propagation): *surprise*



Zimmermann & Hoffmann (Konvens 2022)

Toy example (label propagation): *surprise*



Zimmermann & Hoffmann (Konvens 2022)

Disambiguation

- Assign each word in query the vector of the corresponding node's senses.
- Weigh sense value by its distance to the respective root hub.
- Sum word vectors of entire query.
- Choose sense with the highest value.

Disambiguation algorithm

procedure DISAMBIGUATE

  $S \leftarrow$ context *string*

  $G \leftarrow$ labelled *graph*

  $H \leftarrow$ *list* of root hubs

  $v \leftarrow$ score *vector* with length H

  for $token \in S$ **do**

    if $token \in G$ **then**

      for $h \in H$ **do**

        $v_h \leftarrow v_h + token.\omega_h \cdot \frac{1}{1+d(token,h)}$

  return $\arg\max(v)$

Absinth

Wednesday, 14th September 2022
Konvens 2022

- English Wikipedia dump from 2014,
- without disambiguation pages.
- We do not use a web scraper (or the URLs provided).
- We fine-tuned our system on a sub-set of four samples from the development set and tested on the remaining development set (110 queries).

- ABSINTH: root hubs + label propagation (+ minimum spanning tree[2]).
  - w/o MST: discard unlabelled nodes.
  - w/o labelling: Hyperlex [1].
- Baseline: 10 most frequent tokens as hubs + label propagation (+ MST).
- Singletons: All nodes distinct senses.
- All-in-one: All nodes one sense.

[2] Backup: early stopping produces unlabelled nodes, avg. 2% of nodes labelled by MST.

System	F_1	JI	RI	ARI
ABSINTH	**55.21**	31.73	54.73	6.98
w/o MST	53.57	33.00	**56.21**	**9.08**
w/o labelling	50.13	**46.20**	53.63	5.51
Baseline	49.87	42.52	51.76	3.26
Singletons	**68.66**	0.00	49.00	-0.07
All-in-one	47.42	**51.00**	51.00	0.00

Table 2: Results for F_1-score, Jaccard index (JI), Rand index (RI) and adjusted Rand index (ARI).

Output: *the_colour_of_magic*

Topic: the_colour_of_magic.
Nodes: 156 Edges: 471.
Characteristic path length: 3.93.
Global clustering coefficient: 0.69.
Mean cluster length (arithmetic): 20.0.
Mean cluster length (harmonic): 5.42.
Mean node degree: 6.03.
Number of clusters: 3.
Tuples gained through merging: 0.

Sense inventory:
-> pratchett: terry, discworld, book, series.
-> game: discworld, computer, mobile.
-> sean: astin, comments, album, home.

Output: *ghost*

Topic: ghost.

Nodes: 868 Edges: 2785.

Characteristic path length: 4.47.

Global clustering coefficient: 0.39.

Mean cluster length (arithmetic): 7.87.

Mean cluster length (harmonic): 3.65.

Number of clusters: 8.

Tuples gained through merging: 3.

Sense inventory:

-> christmas: carol, scrooge, past, dickens.

-> film: horror, story, films, american.

-> album: band, song, single, records.

-> holy: church, father, son, catholic.

-> player: game, players, time, mode.

-> house: story, box, night, julian.

-> series: television, episode, tv, season.

-> town: county, united, states, population.

-> james: story, stories, r., m., horror.

-> rolls: royce, silver, cars, phantom.

-> family: moths, world, hepialidae.

-> rider: marvel, blaze, comics, vengeance.

Zimmermann & Hoffmann (Konvens 2022)

Output: *prince_of_persia*

Topic: prince_of_persia.
Nodes: 200 Edges: 674.
Characteristic path length: 3.55.
Global clustering coefficient: 0.66.
Mean cluster length (arithmetic): 17.33.
Mean cluster length (harmonic): 9.61.
Mean node degree: 6.74.
Number of clusters: 3.
Tuples gained through merging: 0.

Sense inventory:
-> arterton: sands, time, gyllenhaal.
-> game: video, ubisoft, sands, series.
-> creed: assassin, games, video, series.
-> screenshots: reviews, cheats, trailers.
-> %: reviews, score, pc, metacritic.

Output: *stephen_king*

Topic: stephen_king.

Nodes: 157 Edges: 527.

Characteristic path length: 3.49.

Global clustering coefficient: 0.49.

Mean cluster length (arithmetic): 43.5.

Mean cluster length (harmonic): 36.45.

Mean node degree: 6.71.

Number of clusters: 2.

Tuples gained through merging: 0.

Sense inventory:

-> novel: film, book, horror, series.

-> short: story, collection, stories.

Limitations

- Most recent & state-of-the-art work is on SemEval Task 13 (induction of senses for polysemous verbs, adjectives and nouns from WordNet).
- We could get our hands on the test queries, but not the gold test sense sets.[3]
- We can report a relative gain compared to Hyperlex on this task, but not much more.
- The coöccurence graphs still encoded textual similarities, not entity-conceptual similarities.

[3] If someone still has that somewhere lying around, we would be happy to send you our clustering and we'll publish the results to Gitlab.

Thanks!

Slides, resources and contact info:

victor.zimmermann@uni-leipzig.de

axtimhaus.eu

gitlab.com/axtimhaus

@dieaxtimhaus

[1] J. Véronis, "Hyperlex: Lexical cartography for information retrieval," *Computer Speech & Language*, vol. 18, no. 3, pp. 223–252, 2004. [Online]. Available: *https://doi.org/10.1016/j.csl.2004.05.002*.

[2] W. L. Hamilton, K. Clark, J. Leskovec, and D. Jurafsky, "Inducing domain-specific sentiment lexicons from unlabeled corpora," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, 2016, pp. 595–605. [Online]. Available: *http://aclweb.org/anthology/D/D16/D16-1057.pdf*.

[3] C. Biemann, ''Chinese whispers - an efficient graph clustering algorithm and its application to natural language processing problems,'' in *Proceedings of TextGraphs: the First Workshop on Graph Based Methods for Natural Language Processing*, New York City: Association for Computational Linguistics, Jun. 2006, pp. 73–80. [Online]. Available: `https://aclanthology.org/W06-3812`.

[4] R. Navigli and G. Crisafulli, ''Inducing word senses to improve web search result clustering,'' in *Proceedings of the 2010 conference on empirical methods in natural language processing*, 2010, pp. 116–126.

[5] R. K. Amplayo, S.-w. Hwang, and M. Song, ''Autosense model for word sense induction,'' in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 6212–6219.

[6] J. Wang, M. Bansal, K. Gimpel, B. D. Ziebart, and C. T. Yu, ''A sense-topic model for word sense induction with unsupervised data enrichment,'' *Transactions of the Association for Computational Linguistics*, vol. 3, pp. 59–71, 2015.

[7] A. Amrami and Y. Goldberg, ''Towards better substitution-based word sense induction,'' *arXiv preprint arXiv:1905.12598*, 2019.

[8] M. E. J. Newman, ''The structure and function of complex networks,'' *SIAM REVIEW*, vol. 45, pp. 167–256, 2003.

[9] J. H. Fowler, ''Turnout in a small world,'' *The social logic of politics: Personal networks as contexts for political behavior*, vol. 269, 2005.

[10] R. Tripodi and M. Pelillo, ''A game-theoretic approach to word sense disambiguation,'' *Computational Linguistics*, vol. 43, no. 1, pp. 31–70, 2017.

[11] D. Ustalov, A. Panchenko, and C. Biemann, ''Watset: Automatic induction of synsets from a graph of synonyms,'' in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vancouver, Canada: Association for Computational Linguistics, Jul. 2017, pp. 1579–1590. [Online]. Available: `https://aclanthology.org/P17-1145`.

[12] J. B. Kruskal, ''On the shortest spanning subtree of a graph and the traveling salesman problem,'' *Proceedings of the American Mathematical Society*, vol. 7, no. 1, pp. 48–50, 1956, ISSN: 00029939, 10886826. [Online]. Available: *http://www.jstor.org/stable/2033241*.

[13] C. Biemann, ''Chinese whispers-an efficient graph clustering algorithm and its application to natural language processing problems,'' in *Proceedings of TextGraphs: the first workshop on graph based methods for natural language processing*, 2006, pp. 73–80.