Efficient Graph-based Word Sense Induction by Distributional Inclusion Vector Embeddings

Haw-Shiuan Chang[1], Amol Agrawal[1], Ananya Ganesh[1],
Anirudha Desai[1], Vinayak Mathur[1], Alfred Hough[2], Andrew McCallum[1]

[1]UMass Amherst, [2]Lexalytics

Word Sense Disambiguation

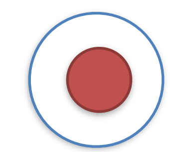
Plaintext corpus

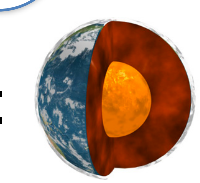
... these cold dense **core** be the site of future star formation ...

... both basic cpus and standard product built around a CPU **core** ...

... the innovation of the common **core** , a educational strategy ...

Target word: core

sense 1: the most essential part 

sense 2: the central part of a planet 

sense 3: CPU 

......

Word Sense Induction (WSI)

Plaintext corpus

... these cold dense core be the site of future star formation ...

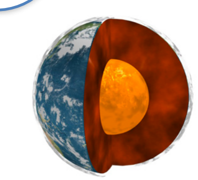
... both basic cpus and standard product built around a CPU core ...

... the innovation of the common core , a educational strategy ...

Target word: core

sense 1: the most essential part

sense 2: the central part of a planet

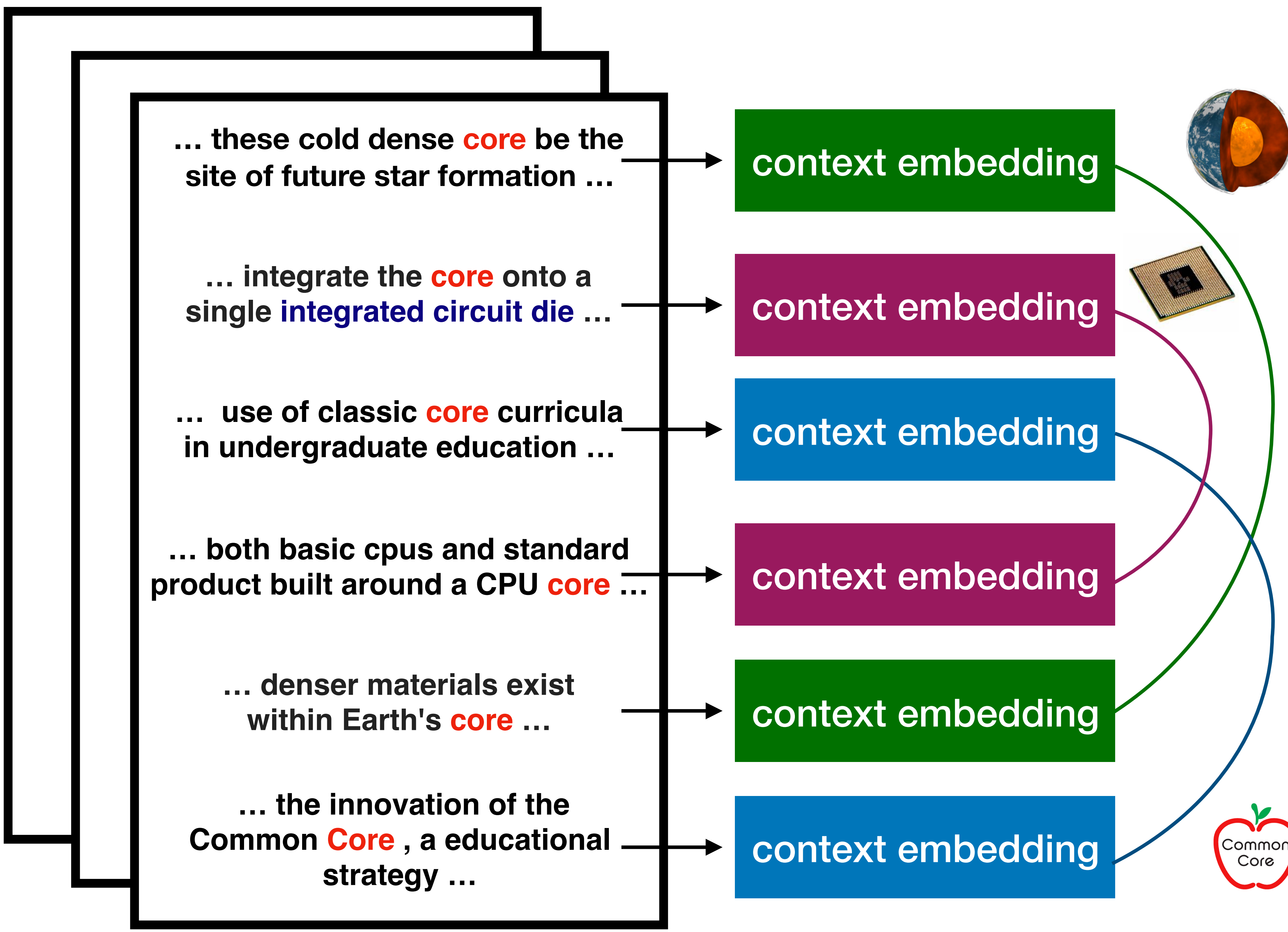


sense 3: CPU

......

Also known as
unsupervised word sense disambiguation

Related Work - 1: Clustering Mentions



Issues

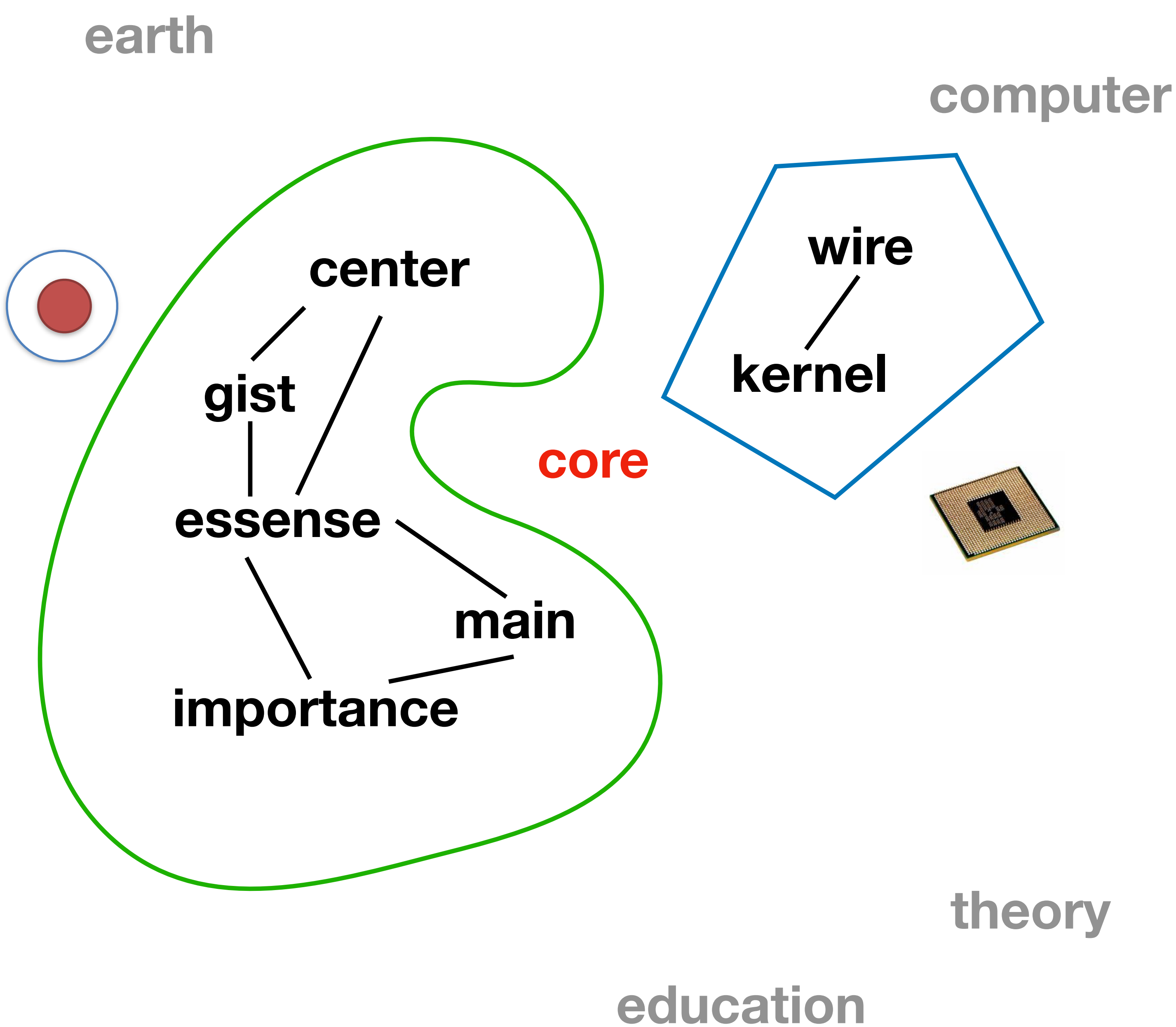
For each target word, need to cluster tens of thousands sentences

Local refinement (e.g., EM) could be efficient, but the granularity usually need to specified at the beginning

Hard to do global optimization

Schutze, 1992; Reisinger and Mooney, 2010; Neelakantan et al., 2014; Tian et al., 2014; Pina and Johansson, 2015; Li and Jurafsky, 2015; Bartunov et al., 2016; Mu et al., 2017; Athiwaratkun Wilson, 2017, etc

Related Work - 2: Clustering Related Words



Issues

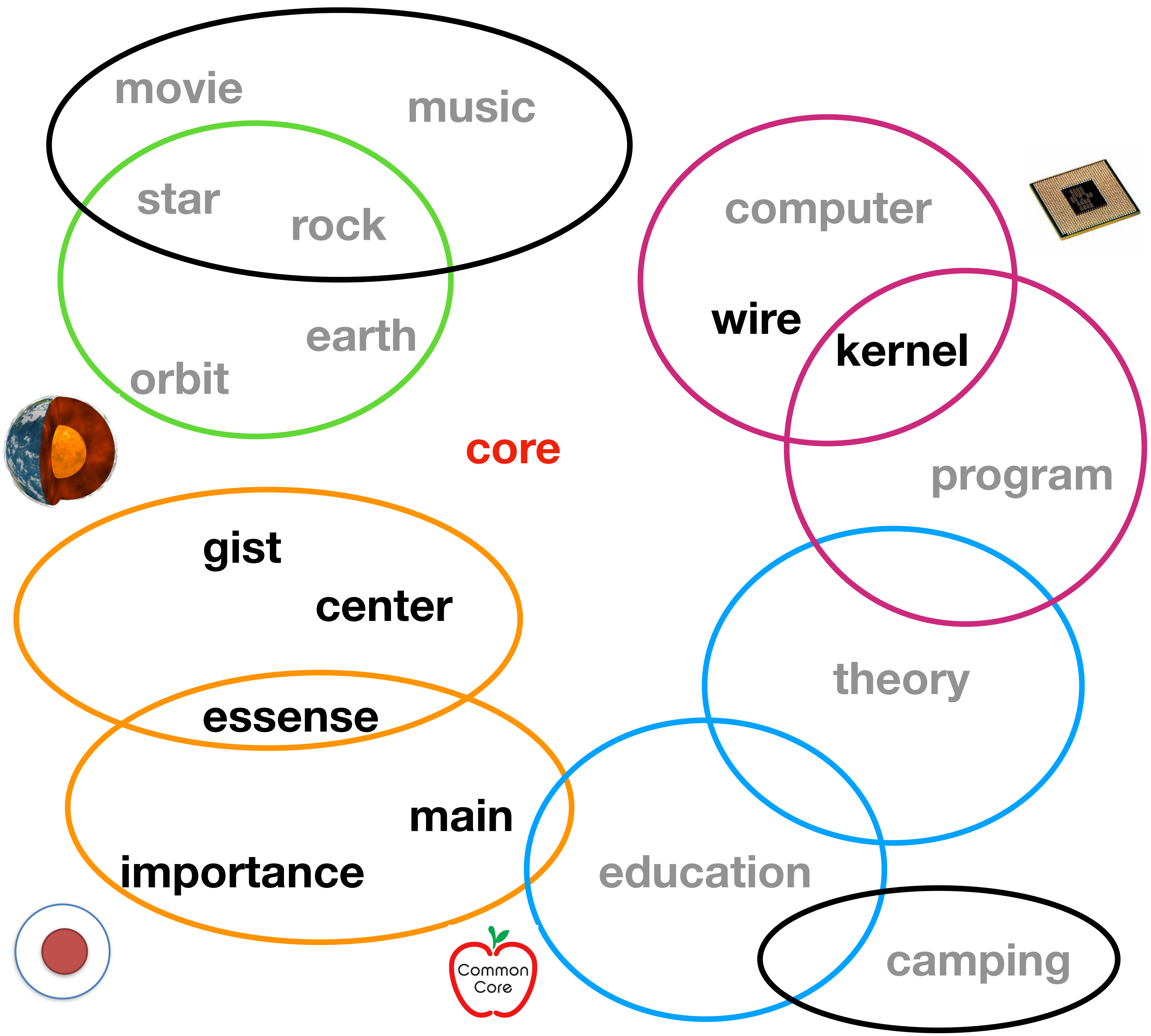
Time consuming to find related words[1]

Hard to know how many related words to be included

[1]Could do some approximated nearest neighbor search, but not sure how it will affect the performance

Lin et al., 1998; Pantel and Lin, 2002; Dorow and Widdows, 2003; Veronis, 2004; Agirre et al., 2006; Biemann, 2006; Navigli and Crisafulli, 2010; Lau et al.; 2012; Hope and Keller, 2013; Di Marco and Navigli, 2013; Mitra et al., 2014; Pelevina et al., 2016

Main Idea: Group Topics



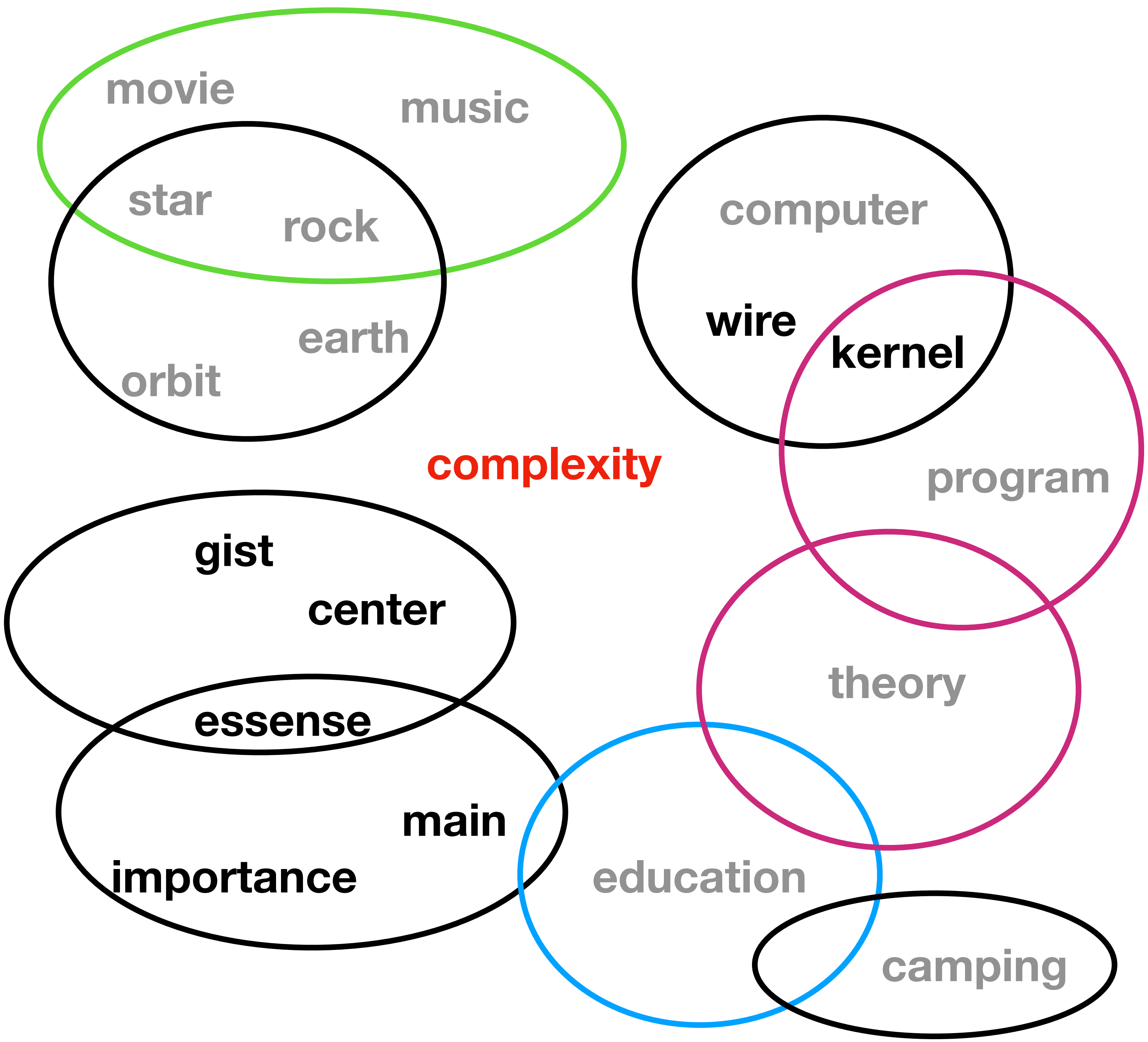
Motivations

More efficient, and different senses usually appear in different topics

Issues

Similarity between topics depends on the target word

Challenge: Similarity Changes



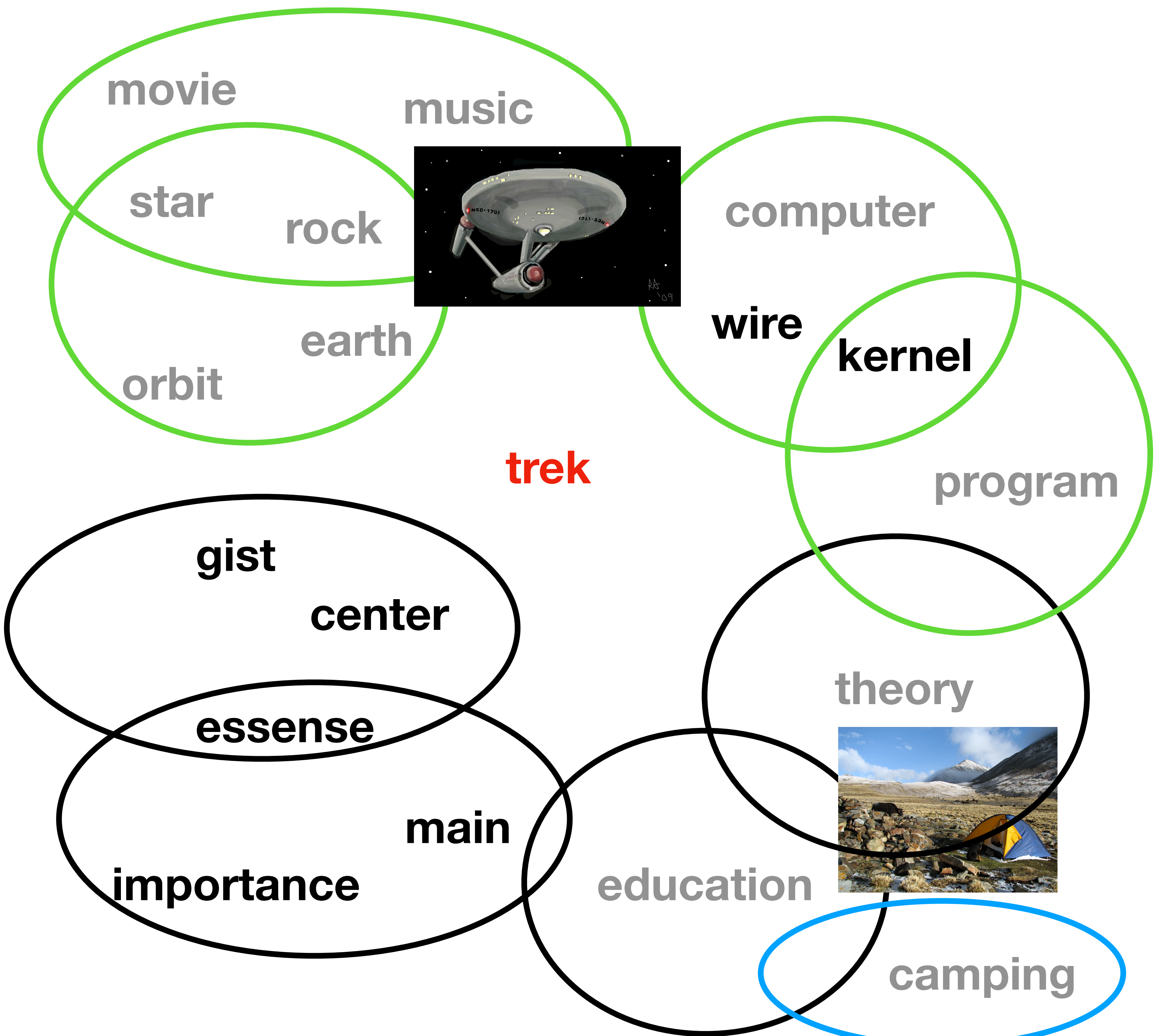
Motivations

More efficient, and different senses usually appear in different topics

Issues

Similarity between topics depends on the target word

Challenge: Similarity Changes



Motivations

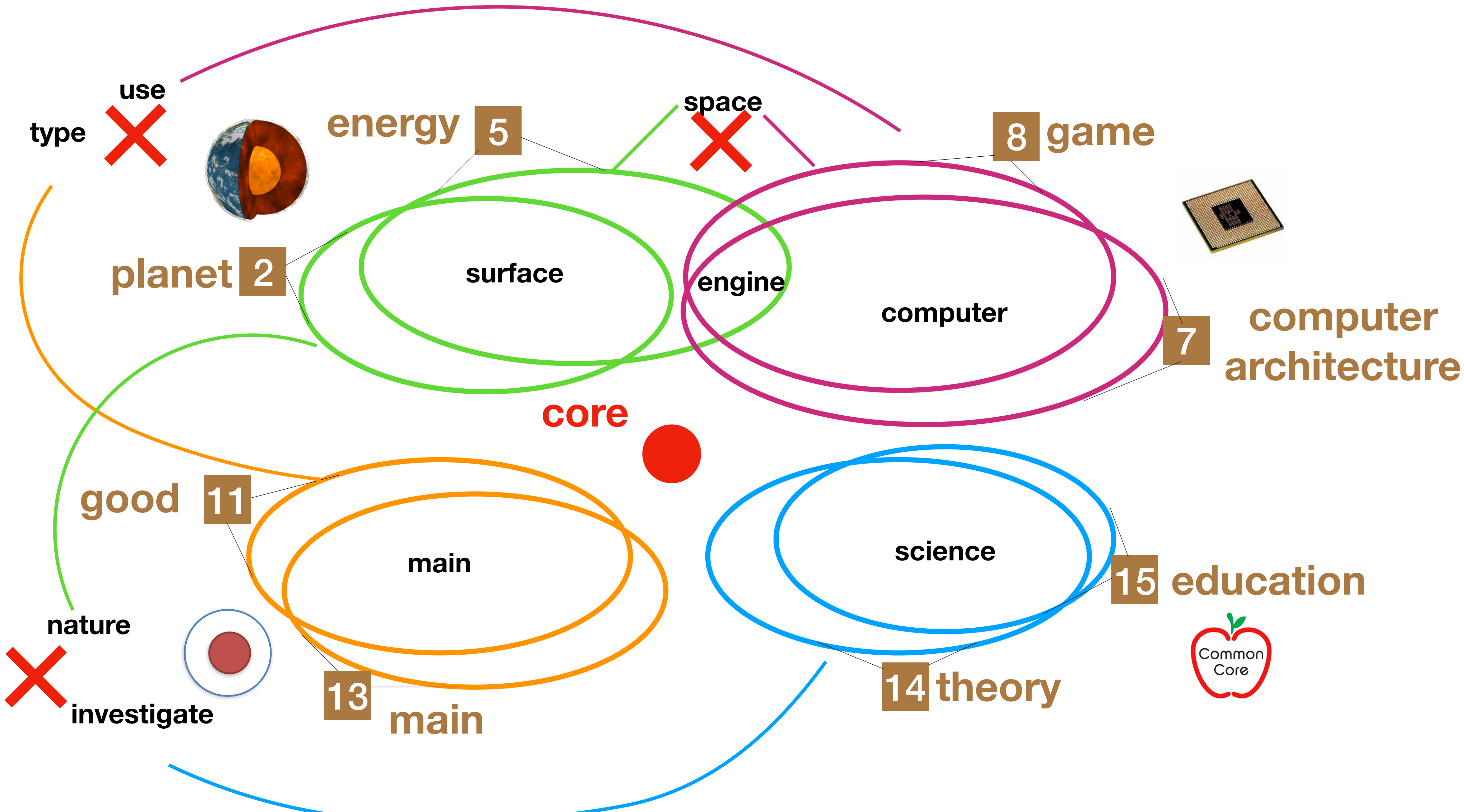
More efficient, and different senses usually appear in different topics

Issues

Similarity between topics depends on the target word

Our Solution:
Focus on Relevant Words

- Topic similarity measurement based only on words both 1) from relevant topics 2) representative in topics

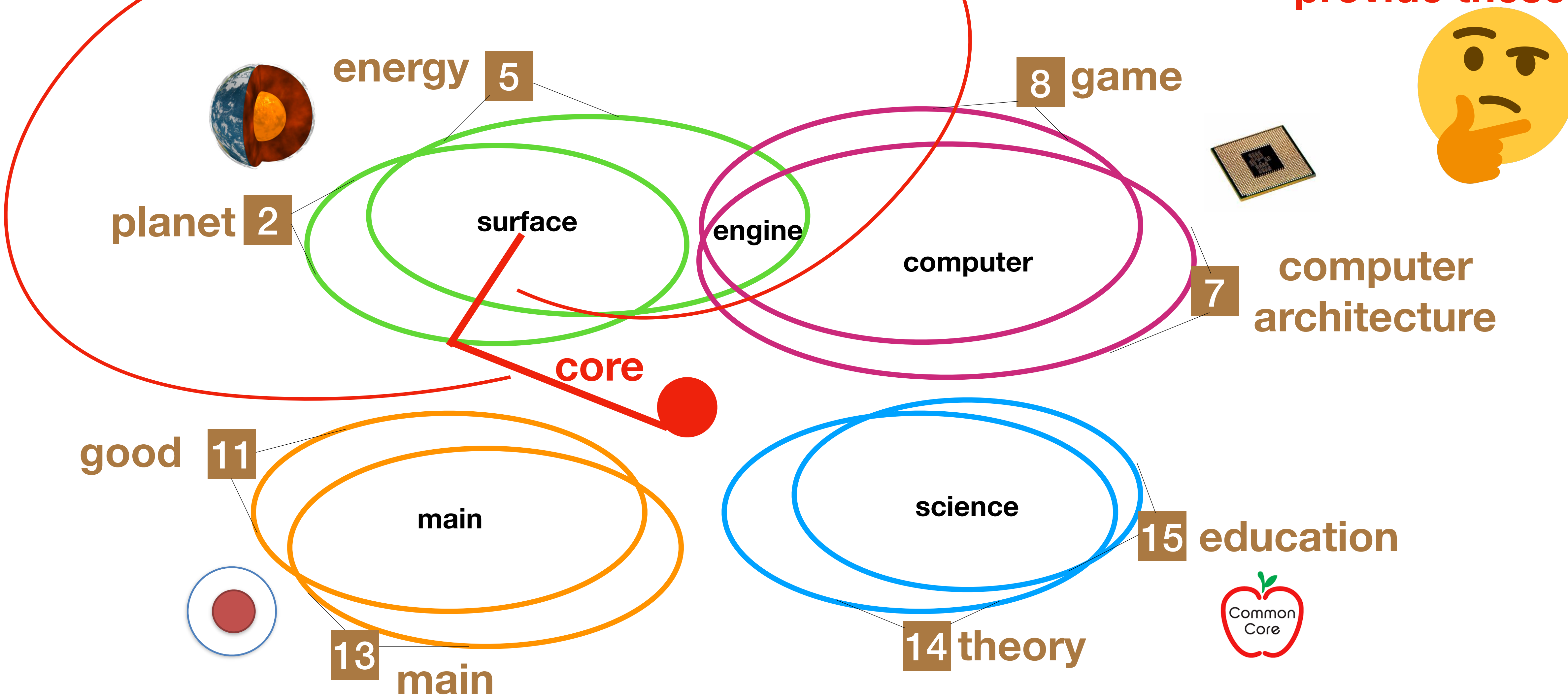


Our Solution: Focus on Relevant Words

- Topic similarity measurement based only on words both 1) from relevant topics 2) representative in topics

$P(\text{topic}_j \mid \text{target word}) \uparrow$ **$P(\text{word} \mid \text{topic}_j) \uparrow$**

Which topic modeling could provide these two?



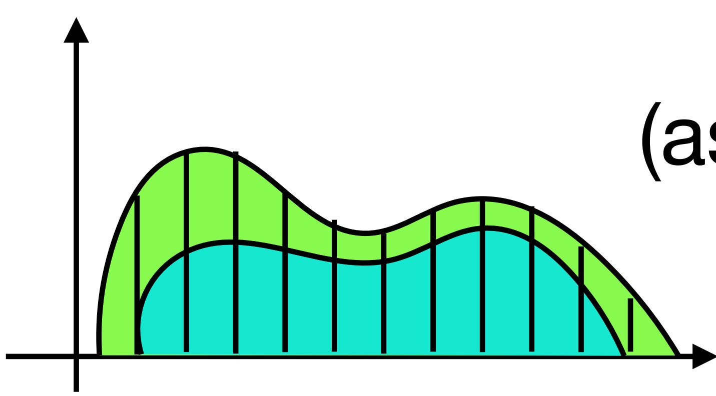
Distributional Inclusion Vector Embedding (DIVE)

context word count → context topic count

DIVE

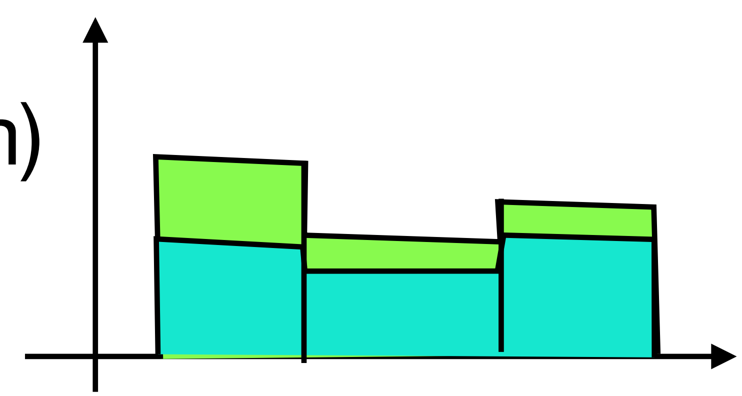
A general way to compress sparse bag of words

(as efficient as skip-gram)



A <= B

~

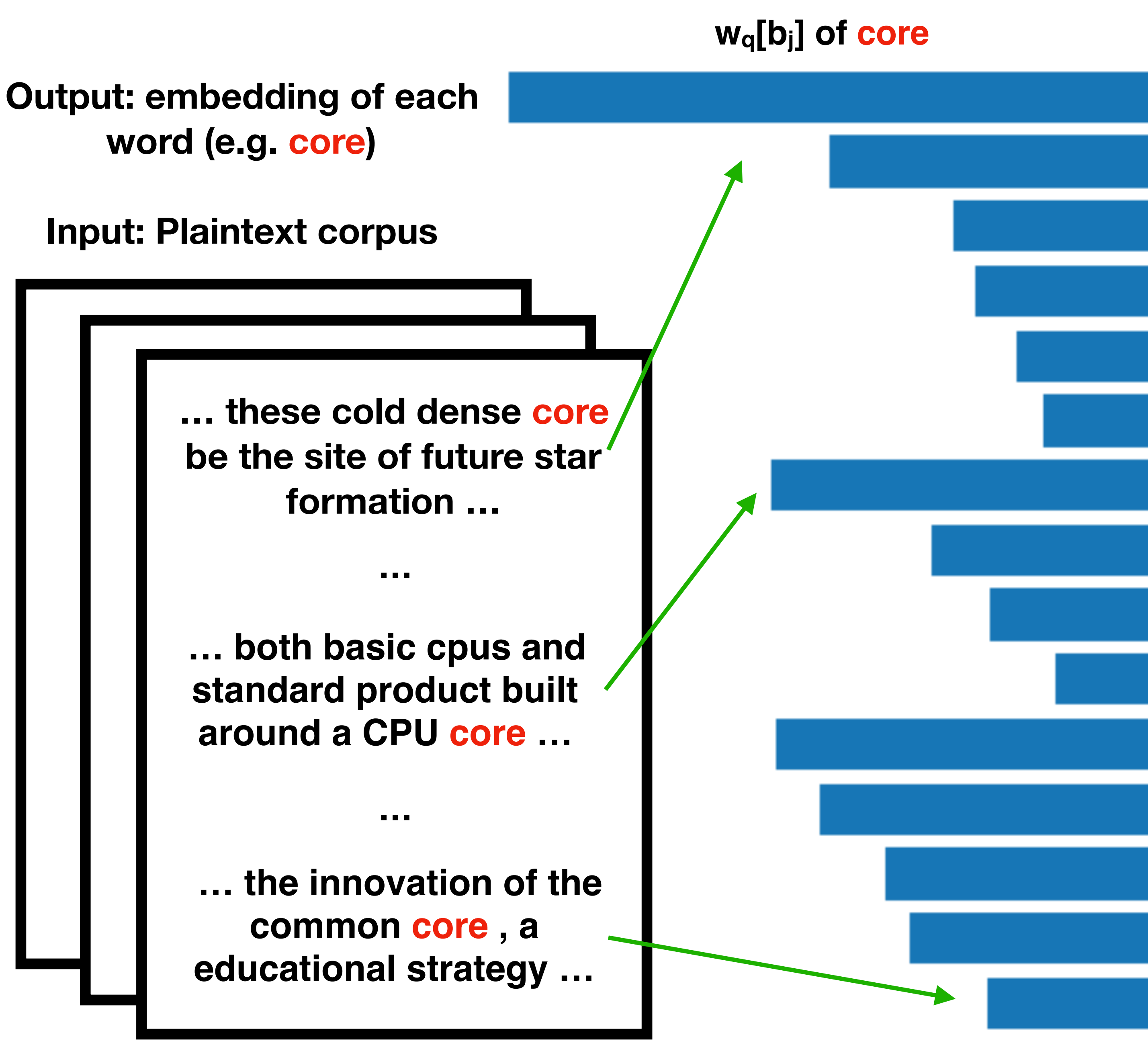


A <= B

DIVE also achieves state-of-the-art performances in unsupervised hypernym detection [1]

[1] Haw-Shiuan Chang, ZiYun Wang, Luke Vilnis, and Andrew McCallum. 2018. Distributional inclusion vector embedding for unsupervised hypernymy detection. In HLT/NAACL.

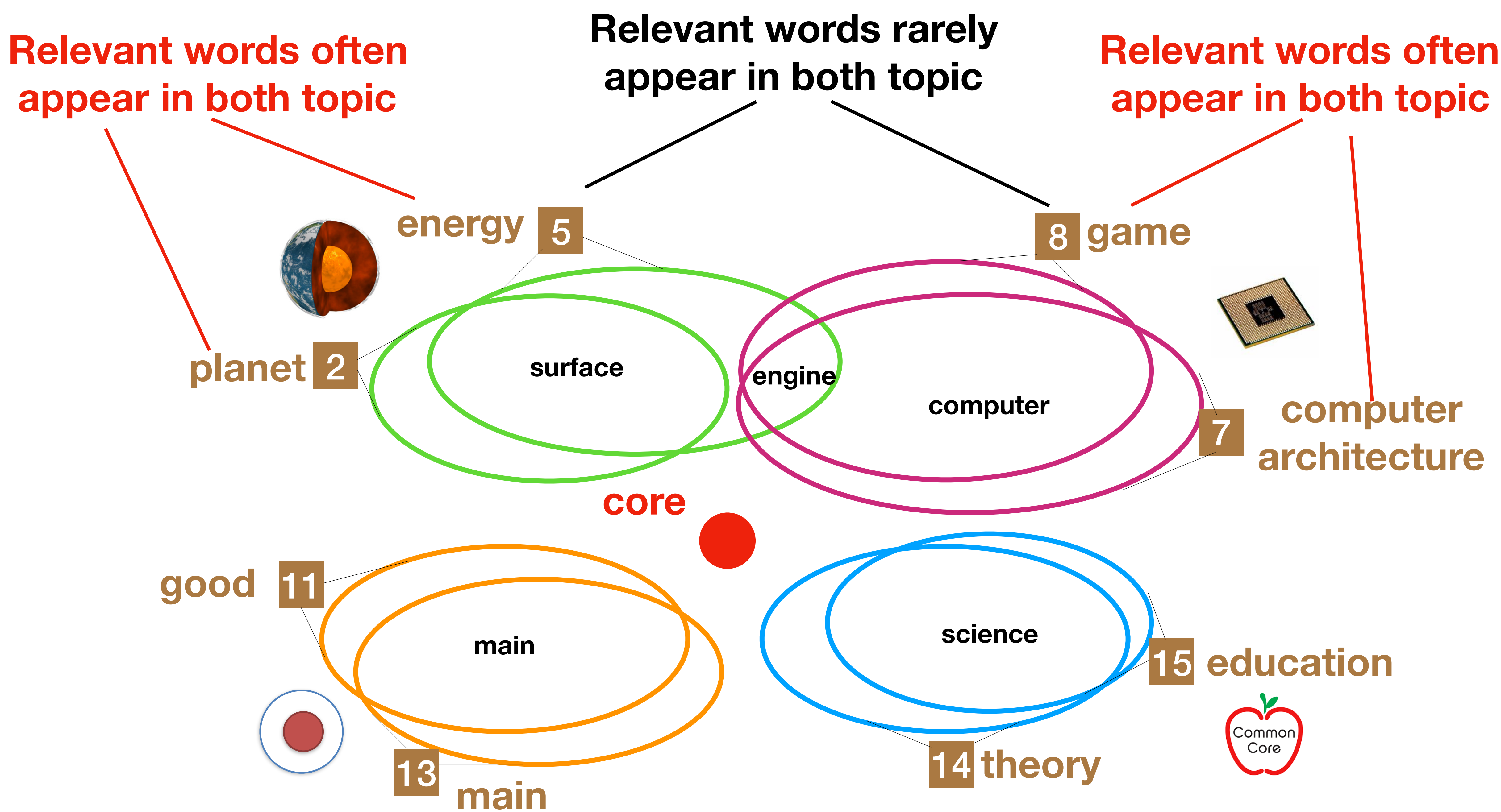
Distributional Inclusion Vector Embedding (DIVE)



b_j	Top 1-5 words
1	element, gas, atom, rock, carbon
2	star, orbit, sun, orbital, planet
3	electron, current, electric, circuit, voltage
4	tank, cylinder, wheel, engine, steel
5	high, low, temperature, energy, speed
6	acid, carbon, product, use, zinc
7	system, architecture, develop, base, language
8	version, game, release, original, file
9	network, user, server, datum, protocol
10	access, need, require, allow, program
11	also, well, several, early, see
12	part, almost, see, addition, except
13	several, main, province, include, consist
14	science, philosophy, theory, philosopher, term
15	school, university, student, education, college

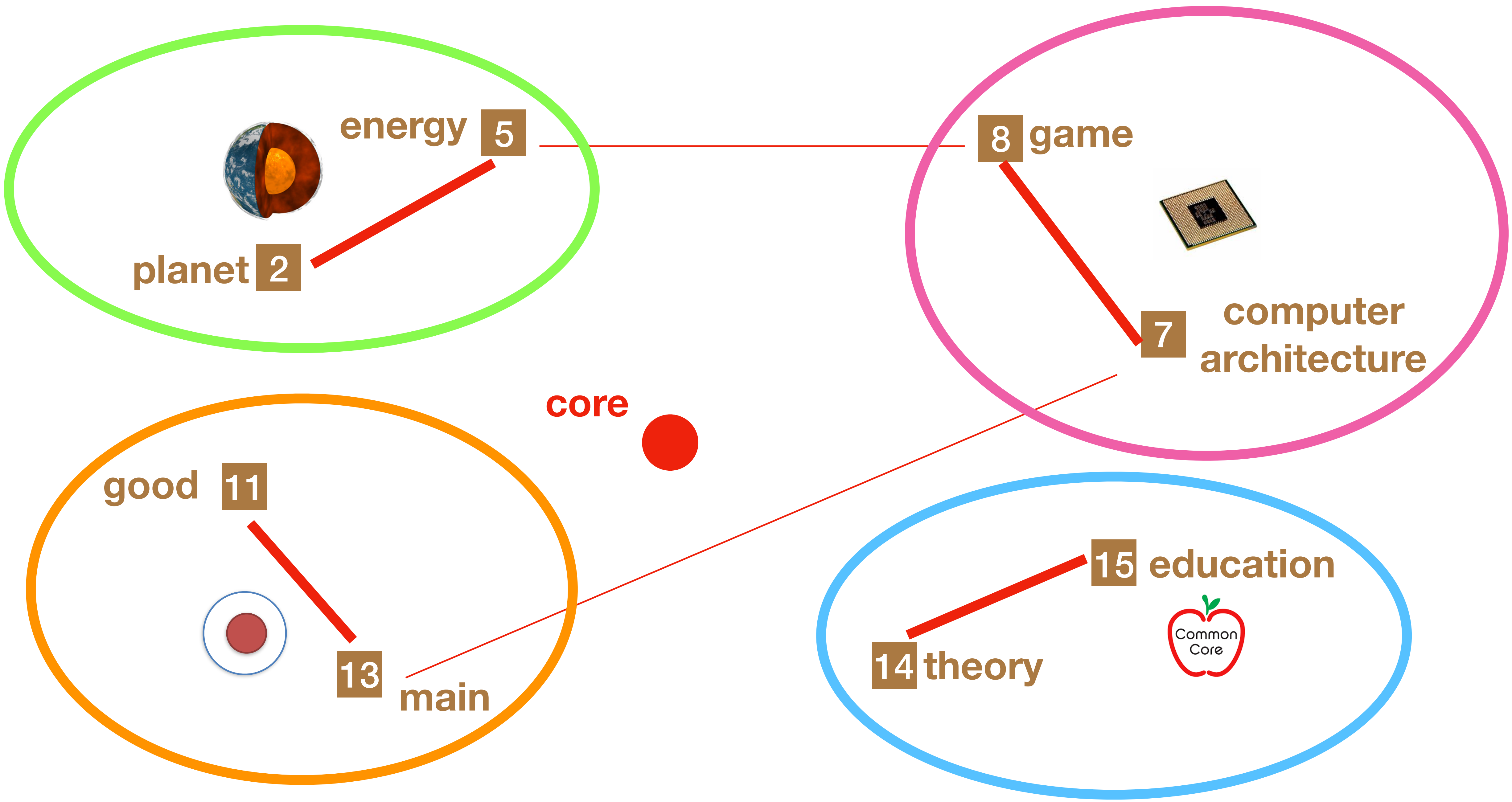
[1] Haw-Shiuan Chang, ZiYun Wang, Luke Vilnis, and Andrew McCallum. 2018. Distributional inclusion vector embedding for unsupervised hypernymy detection. In HLT/NAACL.

Similarity Estimation



Graph-based Clustering

- For simplicity, we use spectral clustering

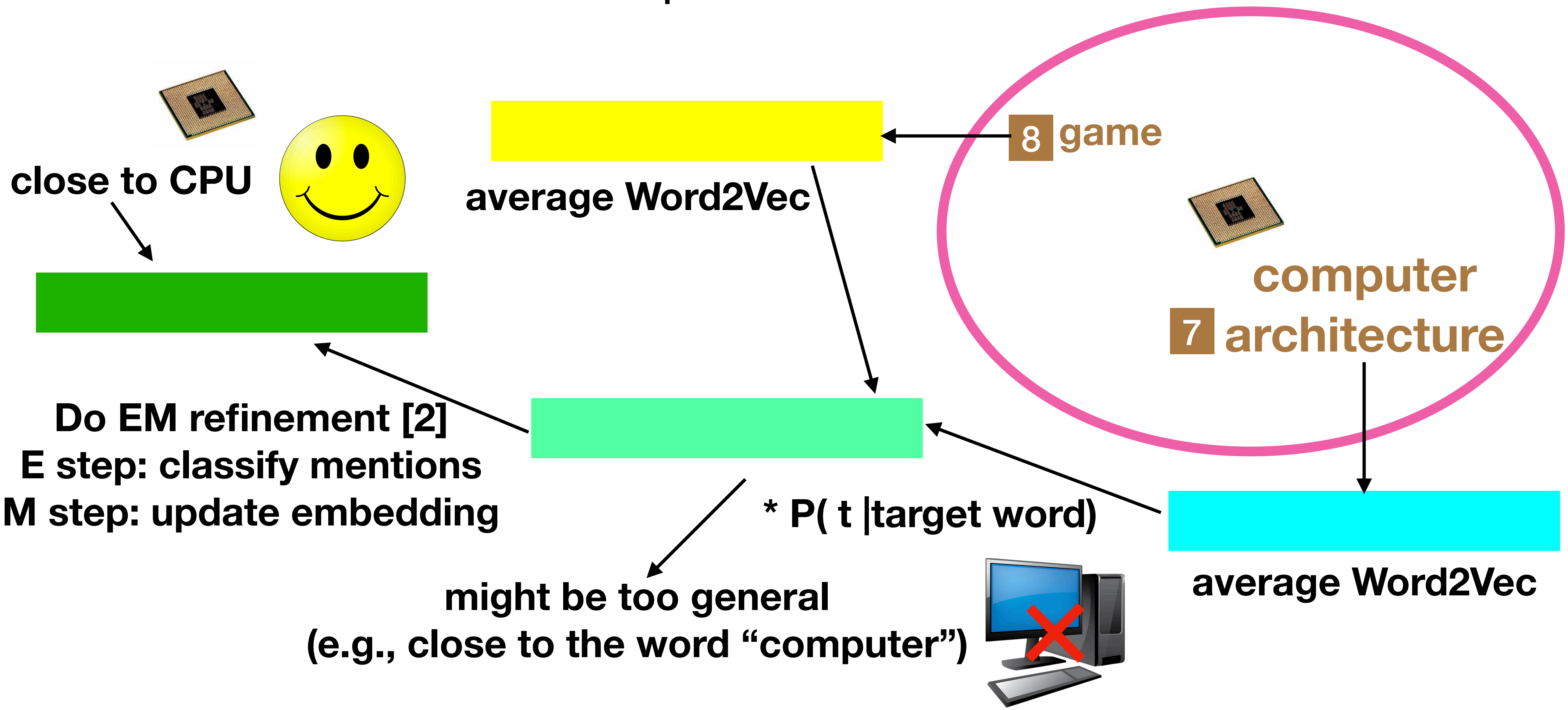


Some Examples

Query	CID	Top 5 words in the top dimensions	
rock	1	element, gas, atom, rock, carbon	sea, lake, river, area, water
		find, specie, species, animal, bird	point, side, line, front, circle
	2	band, song, album, music, rock	write, john, guitar, band, author
		early, work, century, late, begin	include, several, show, television, film
bank	1	county, area, city, town, west	several, main, province, include, consist
		building, build, house, palace, site	sea, lake, river, area, water
	2	money, tax, price, pay, income	company, corporation, system, agency, service
		united, states, country, world, europe	state, palestinian, israel, right, palestine
apple	1	food, fruit, vegetable, meat, potato	goddess, zeus, god, hero, sauron
		war, german, ii, germany, world	write, john, guitar, band, author
	2	version, game, release, original, file	car, company, sell, manufacturer, model
		system, architecture, develop, base, language	include, several, show, television, film
star	1	film, role, production, play, stage	character, series, game, novel, fantasy
		wear, blue, color, instrument, red	write, john, guitar, band, author
	2	element, gas, atom, rock, carbon	star, orbit, sun, orbital, planet
		give, term, vector, mass, momentum	light, image, lens, telescope, camera

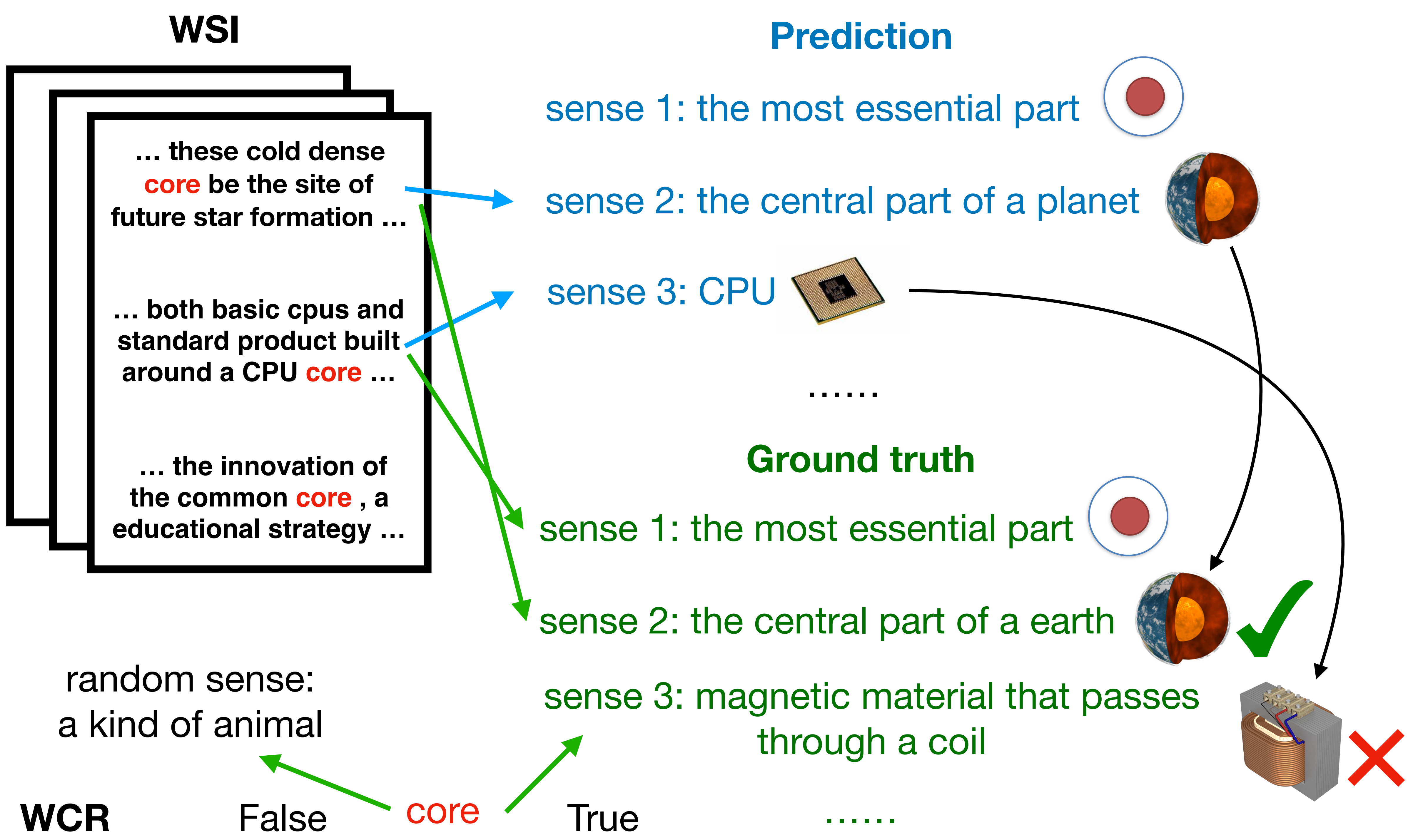
Topic Clustering to Sense Embedding

- Any word embedding could be used, we use Word2Vec in experiments



[2] Arvind Neelakantan, Jeevan Shankar, Alexandre Passos, and Andrew McCallum. 2014. Efficient nonparametric estimation of multiple embeddings per word in vector space. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP

Evaluation



WSI
... these cold dense core be the site of future star formation ...
... both basic cpus and standard product built around a CPU core ...
... the innovation of the common core , a educational strategy ...
Prediction
sense 1: the most essential part
sense 2: the central part of a planet
sense 3: CPU
......
Ground truth
sense 1: the most essential part
sense 2: the central part of a earth
sense 3: magnetic material that passes through a coil
random sense: a kind of animal
WCR
False
core
True
......

Experiments

- Train on Wikipedia
- Test on R1 (WCR), TWSI (WSI), SemEval-2013 task 13 (WSI)
- We fix number of senses to be 2 for each word
- Compare with
  - Random,
  - Single sense (with Word2Vec),
  - MSSG (only doing EM refinement) [2],
  - WG (clustering related words) [3],
  - WG+EM

[3] Maria Pelevina, Nikolay Arefiev, Chris Biemann, and Alexander Panchenko. 2016. Making sense of word embeddings. In Proceedings of the 1st Workshop on Representation Learning for NLP, Rep4NLP@ACL 2016, Berlin, Germany, August 11, 2016 .

Experiment Results

- Our method performs similarly compared with STOA[2], while capturing less frequent senses better
  - Using global topics won't hurt performance due to bad resolution

Skip-gram	WG	WG+EM
52.7	42.1	59.1
MSSG	DIVE (100)	DIVE (300)
60	**63.2**	62.6

Table 2: Precision@1 on the WCR R1 (%).

Model	TWSI			balanced TWSI		
	P	R	F1	P	R	F1
MSSG rnd	66.1	65.7	65.9	33.9	33.7	33.8
MSSG	66.2	65.8	66.0	34.3	34.2	34.2
WG	**68.6**	**68.1**	**68.4**	38.7	38.5	38.6
WG+EM	68.3	67.8	68.0	38.4	38.2	38.3
DIVE rnd	63.4	63.0	63.2	33.4	33.2	33.3
DIVE (100)	67.6	67.2	67.4	**39.7**	**39.5**	**39.6**
DIVE (300)	67.4	66.9	67.2	39.0	38.8	38.9

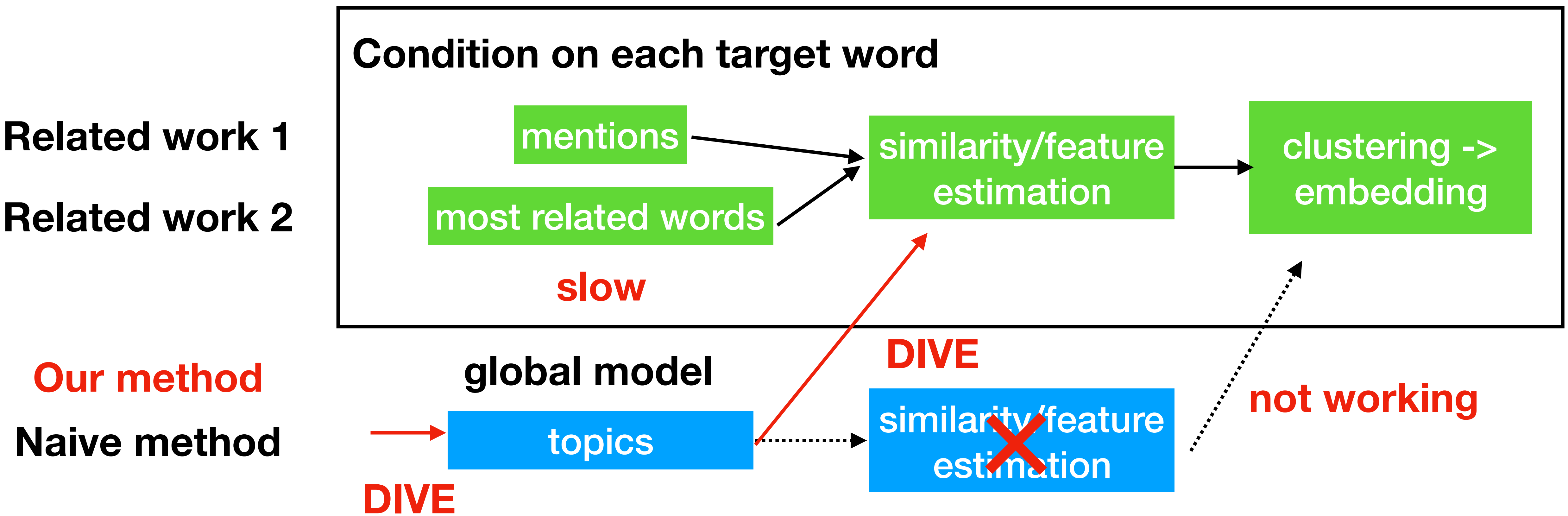
Table 3: Results obtained on the TWSI task (%), where P is precision and R is recall. MSSG rnd and DIVE rnd are baselines which randomly assign sense given inventory built by MSSG and DIVE, respectively.

Model	JI	Tau	WNDCG	FNMI	FB-C
All-1	19.2	60.9	28.8	0	**62.3**
Rnd	21.8	62.8	28.7	2.8	47.4
MSSG	**22.2**	**62.9**	29.0	3.2	48.9
WG	21.2	61.2	29.0	1.6	58.1
WG+EM	21.0	61.5	29.0	1.3	57.8
DIVE (100)	21.9	61.9	**29.3**	3.1	50.6
DIVE (300)	22.1	62.8	29.1	**3.5**	49.9

Table 4: Results obtained on the SemEval 2013 task (%), where JI is Jaccard Index, FNMI is Fuzzy NMI, and FB-C is Fuzzy B-Cubed. All-1 is to assign all senses to be the same and Rnd is to randomly assign all senses to 2 groups.

[2]Maybe slightly worse than AdaGram, which determines number of senses dynamically, which we haven't did

Summary



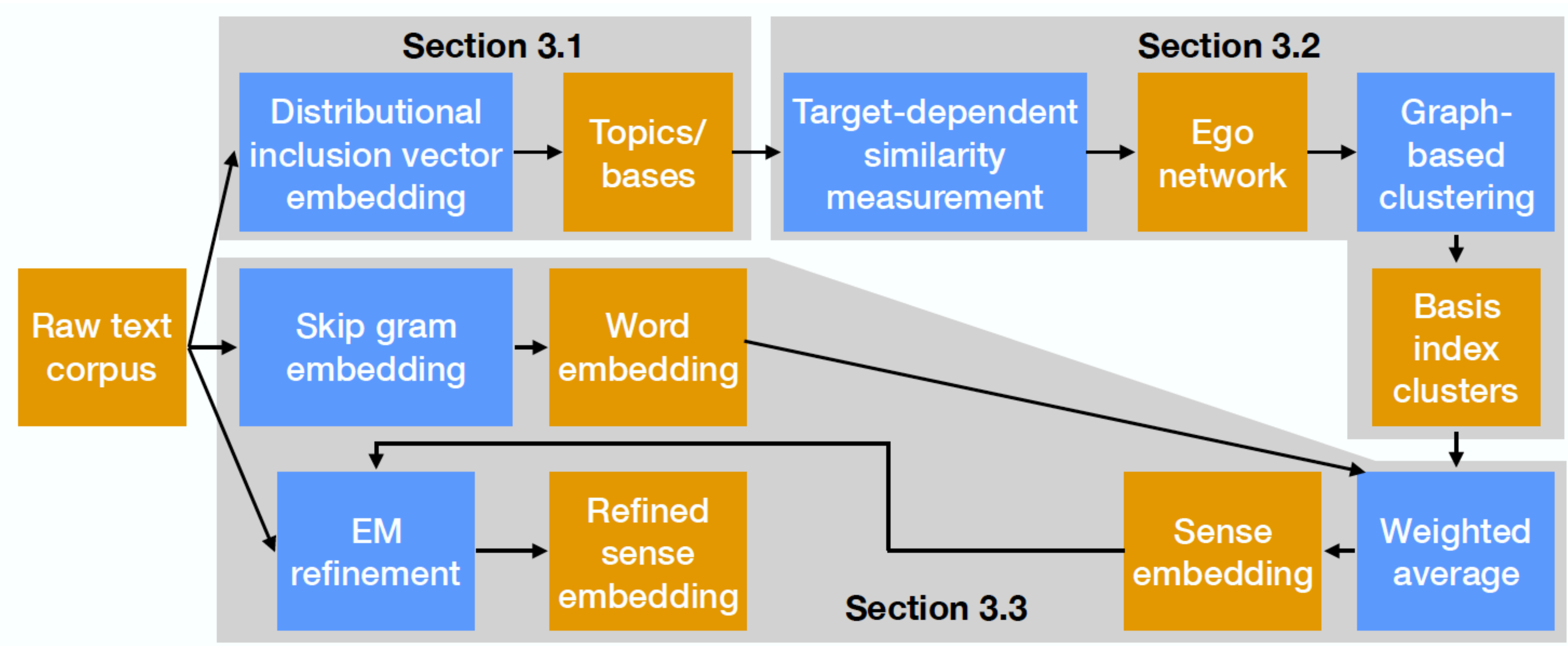
- Clustering mentions or most related words is expansive
- By the help of DIVE, similarity measurement can depend on the target word, which makes clustering topics practical

Future Work

- Make our implementation more efficient
- Dynamically determine the number of clusters
- Use downstream task (e.g., sentiment classification) to guide clustering process

Appendix

Flow Chart


Section 3.1
Distributional inclusion vector embedding
Topics/ bases
Section 3.2
Target-dependent similarity measurement
Ego network
Graph-based clustering
Basis index clusters
Raw text corpus
Skip gram embedding
Word embedding
EM refinement
Refined sense embedding
Sense embedding
Weighted average
Section 3.3

More Examples

tank	1	tank, cylinder, wheel, engine, steel acid, carbon, product, use, zinc	industry, export, industrial, economy, company network, user, server, datum, protocol
	2	army, force, infantry, military, battle however, attempt, result, despite, fail	aircraft, navy, missile, ship, flight war, german, ii, germany, world
race	1	win, world, cup, play, championship	two, one, three, four, another
	2	railway, line, train, road, rail	car, company, sell, manufacturer, model
	3	population, language, ethnic, native, people	female, age, woman, male, household
run	1	system, architecture, develop, base, language	access, need, require, allow, program
	2	railway, line, train, road, rail	also, well, several, early, see
	3	game, team, season, win, league	game, player, run, deal, baseball
tablet	1	bc, source, greek, ancient, date	book, publish, write, work, edition
	2	use, system, design, term, method	version, game, release, original, file
	3	system, blood, vessel, artery, intestine	patient, symptom, treatment, disorder, may