Multi-scale Semantic Network Analysis

Innovative Tools in
the Study of Language Workshop
Padua - 2021

Simon De Deyne

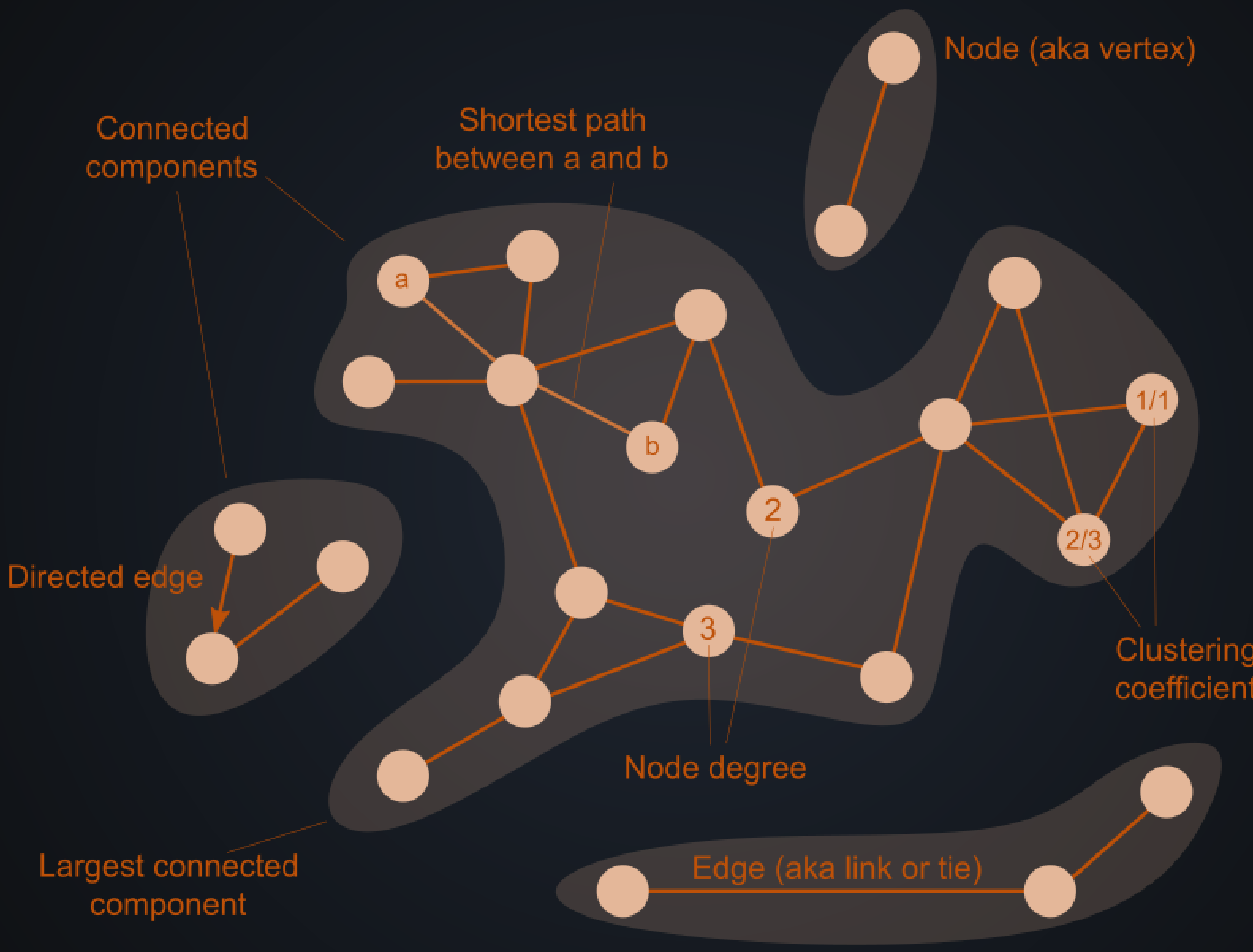
School of Psychological Sciences

simon.dedeyne@unimelb.edu.au

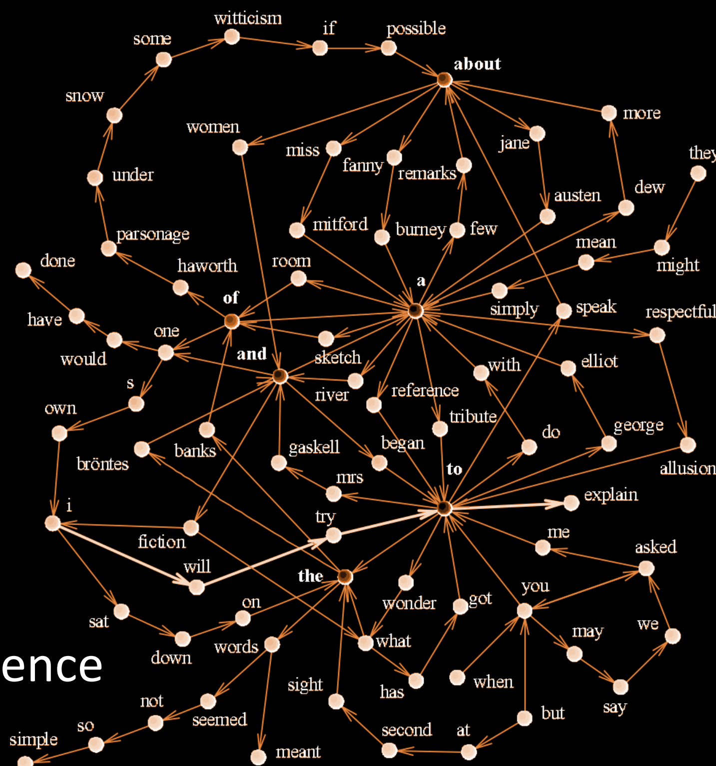
THE UNIVERSITY OF MELBOURNE

Australian Government
Australian Research Council

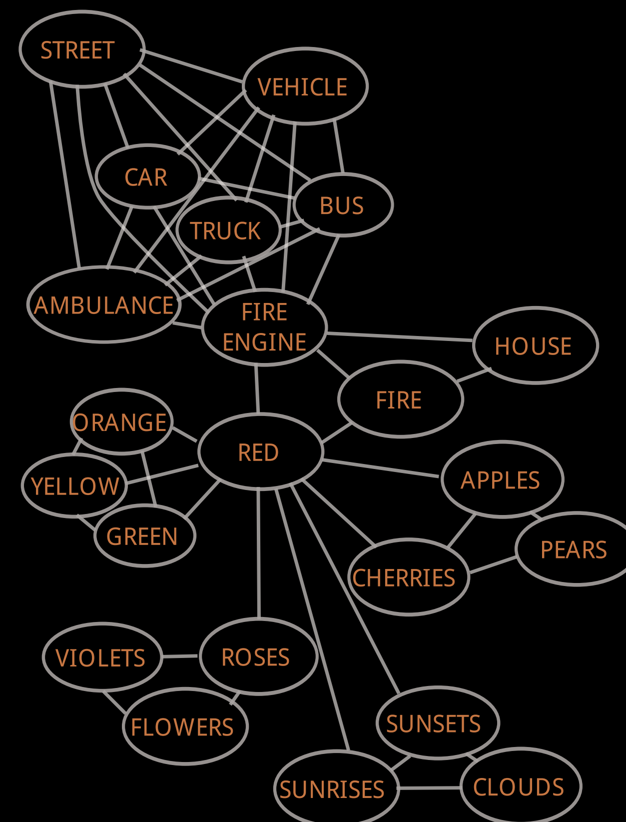
Network Measures



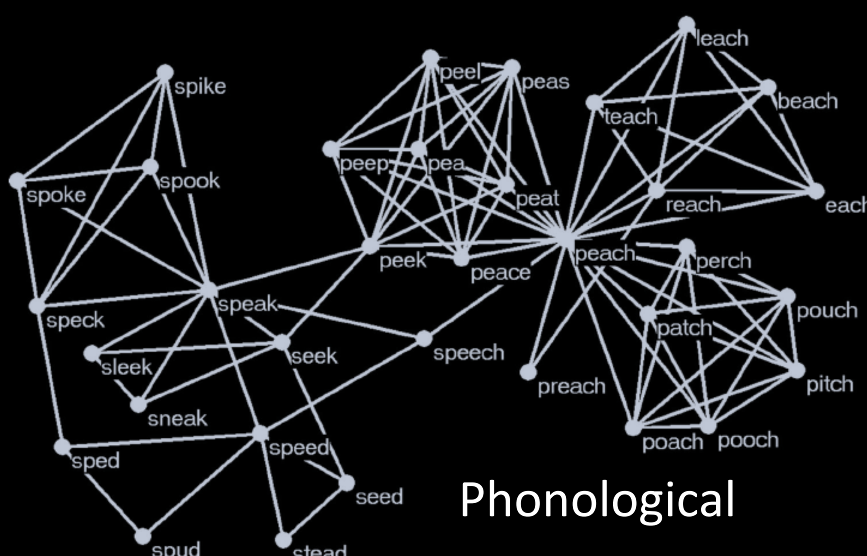
Language Networks



Co-occurrence



Semantic

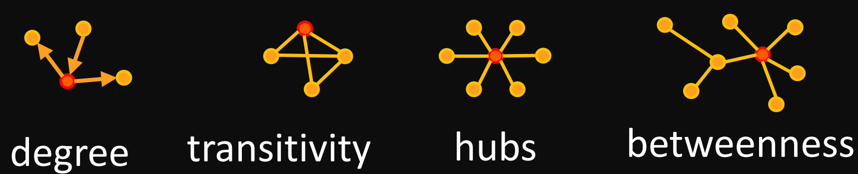


Phonological

Structure can be studied *simultaneously* at different levels

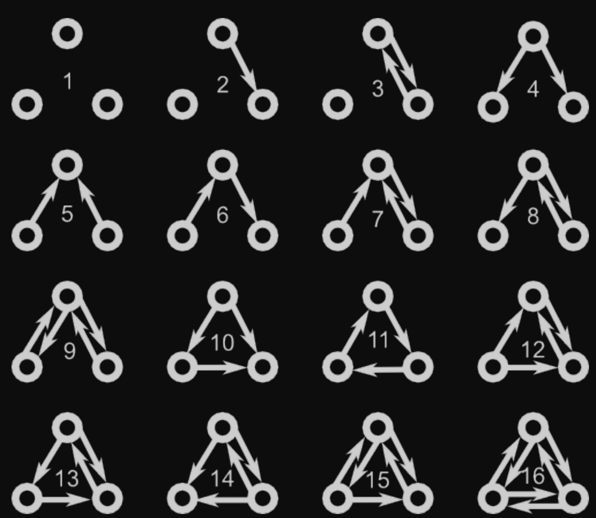
Microscopic structure

Centrality



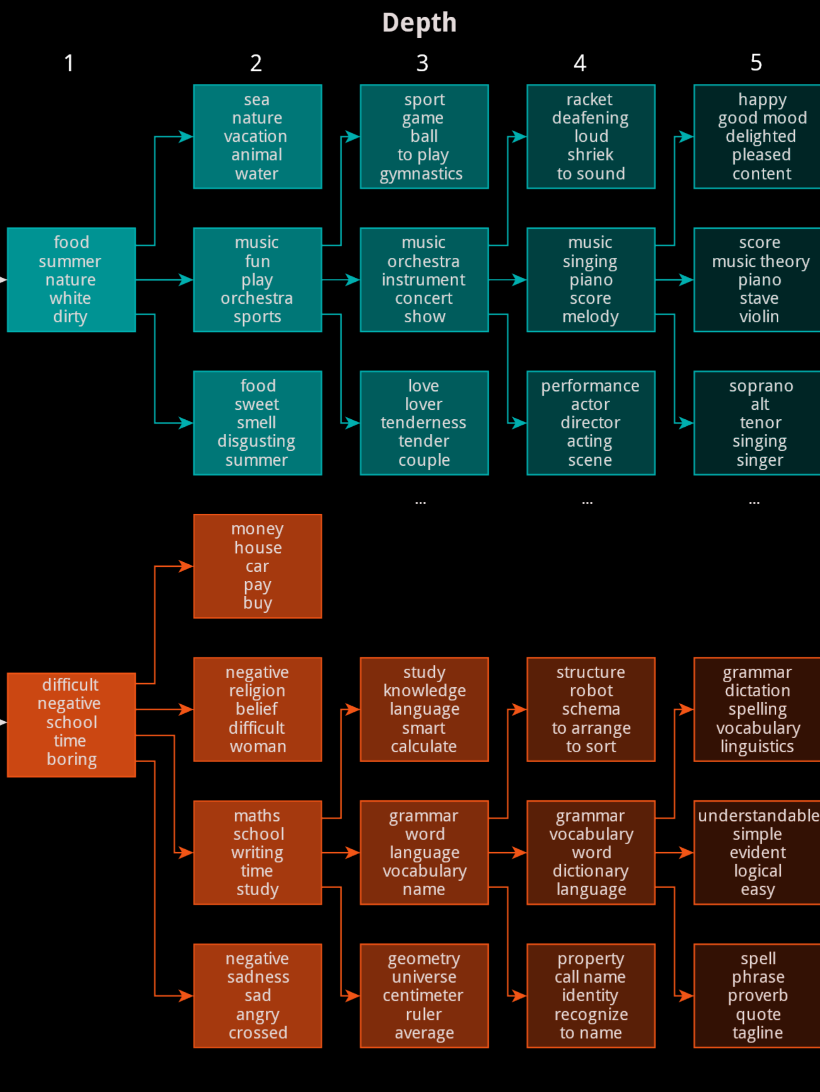
Concrete words have "lower degree" than abstract words
Dense phonological neighbourhoods, affect lexical retrieval times

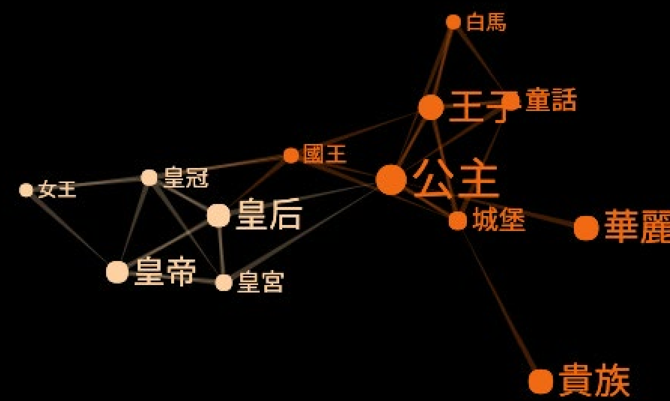
Motives



Classify different language networks based on their *motive signatures* (Biemann, et al. 2016)

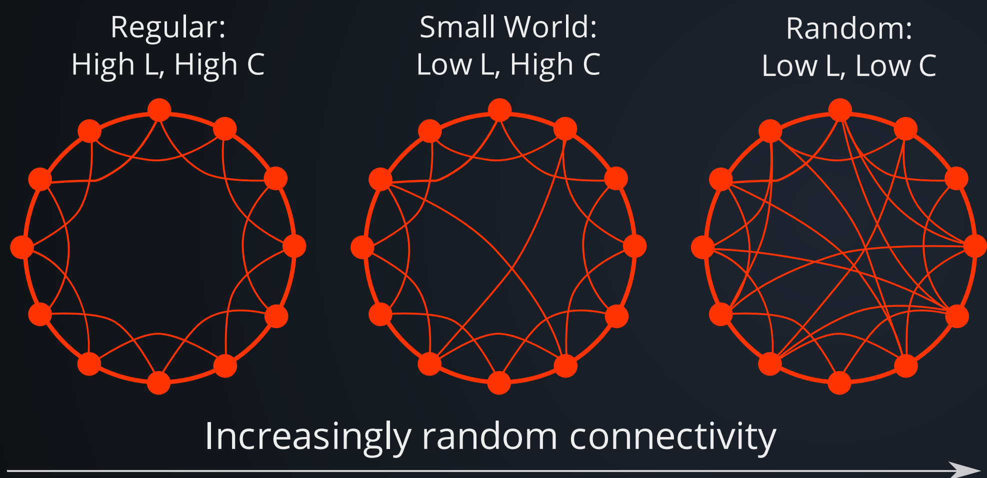
Mesoscopic structure





Macroscopic structure

Small World Scale-Free Networks

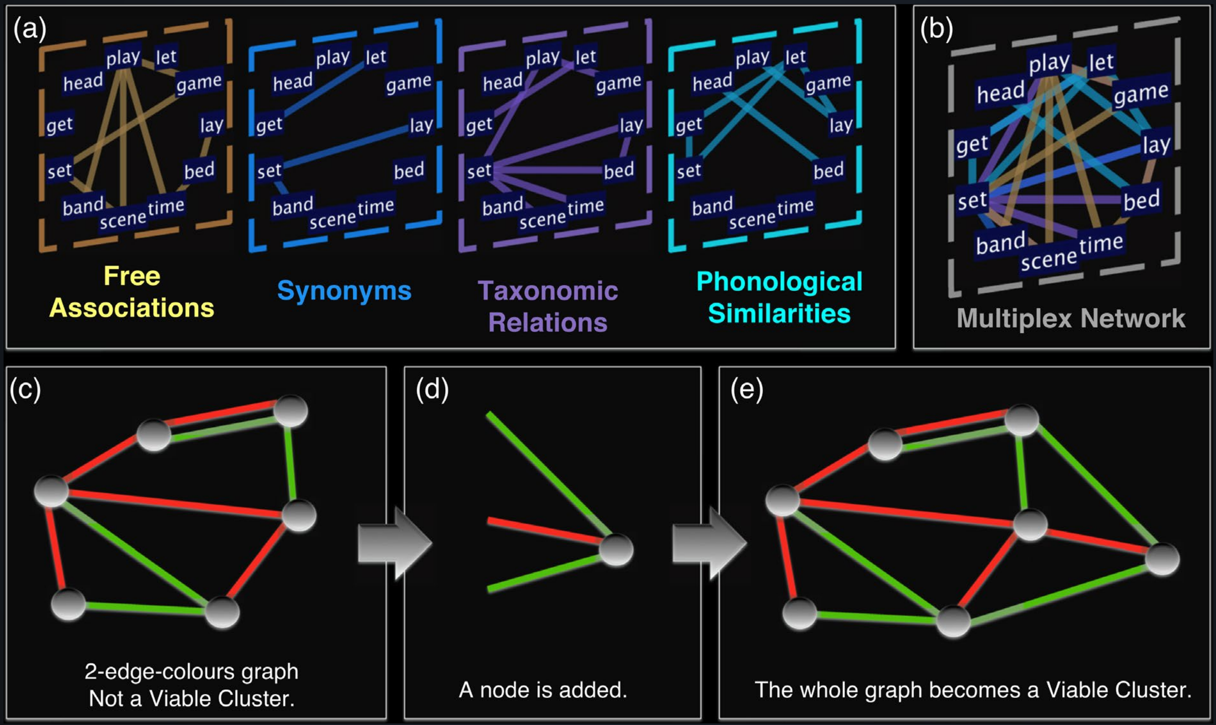


- Short Average Path length (L)
- High Clustering (C)
- Hubs: (good, bad, woman, old)
- Scale-free: degree distribution = power law

- Highly navigable structure, robustness to random failure
- Common in many naturally evolving systems
- Possible mechanism: Preferential Attachment (Steyvers & Tenenbaum , 2014)

Multilevel analysis – Multiplex graphs

Explosive language learning in humans (Stella et al., Scientific Reports, 2018).



Idea:

- Manipulate presence of links according to age-of-acquisition
- What determines tipping points in connectivity ?

Small World of Words project

Build a realistic approximation of the mental lexicon
in major world languages

De Deyne & Storms, 2008, De Deyne et al., 2013;
De Deyne et al. (2019)

small world of words

Discover what **words** mean
for **people** worldwide

Small World of Words project

+ 500,000 participants

https://smallworldofwords.org/project

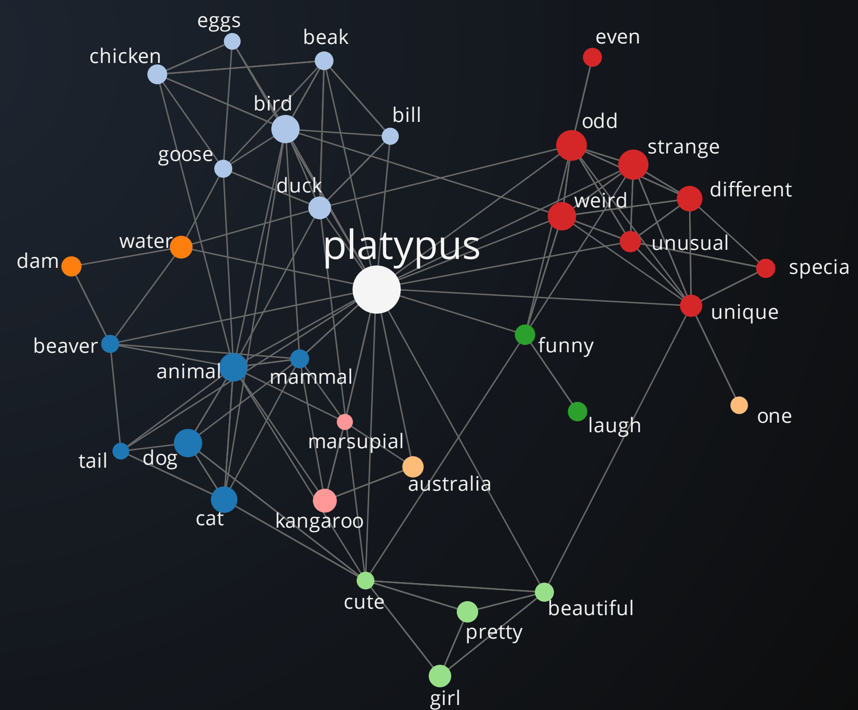
Empirical lexico-semantic network

MAIN CHARACTERISTICS

- Free association
- Edges, nodes and weights observed
- Meaning = spreading activation over weighted edges
- Sizeable implementation covering most words

WORD ASSOCIATION TASK

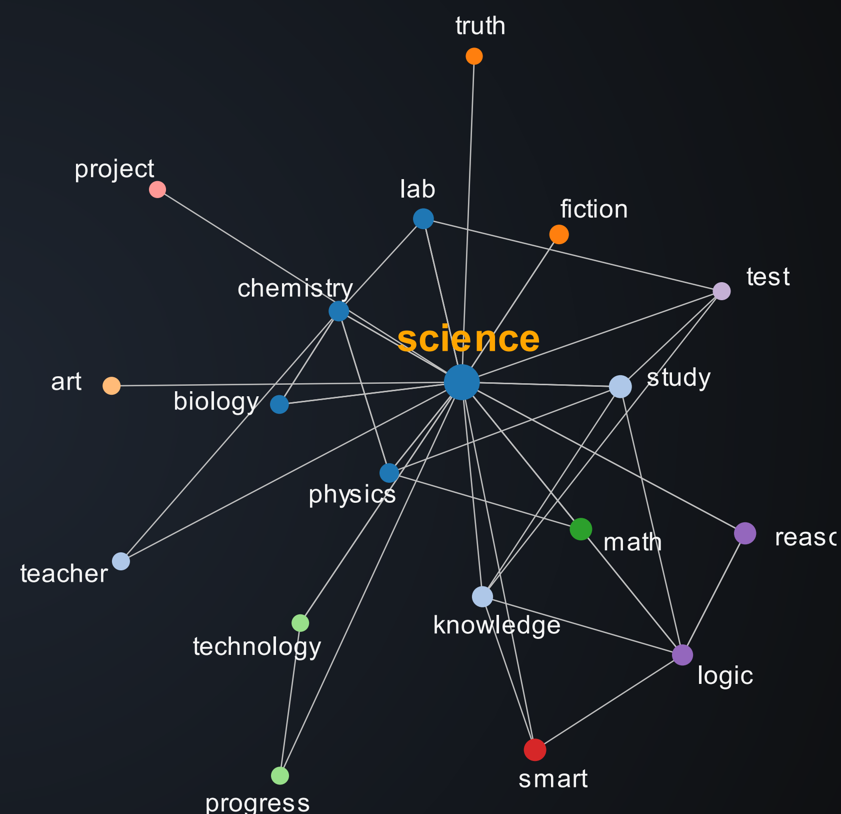
- 12,000 English cue words, +80,000 pps
- Continued Responses: R1, R2, R3
- 3M association responses



Why continued associations?

Single Responses

- General drawback of word associations (Aitchison, 2003)
- Lack of weak responses problematic for various tasks including mediated priming (Chwilla et al., 2000)



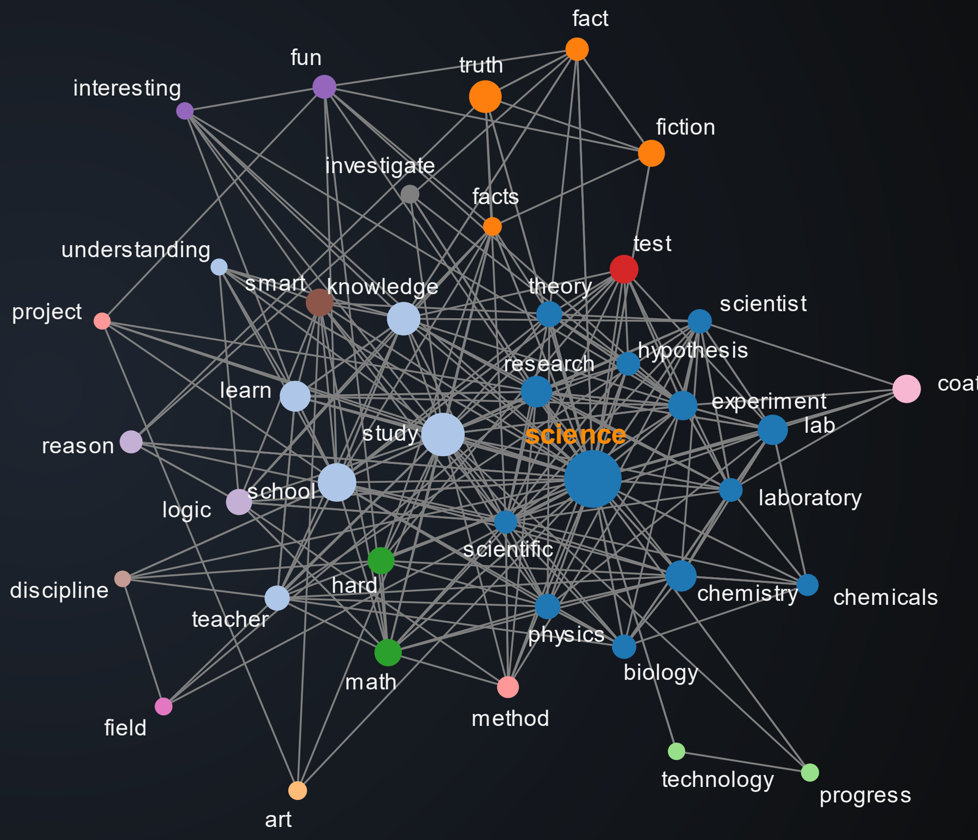
Why continued associations?

Single Responses

- General drawback of word associations (Aitchison, 2003)
- Lack of weak responses problematic for various tasks including mediated priming (Chwilla et al., 2000)

Dominance Effects

- umbrella: 90% of people provide rain
- Possible other associates?



The structure of associations in language and thought

James Deese (1965):

"The interest of psychologists in associations has always been misguided… the whole classical analysis of associations centered around the circumscribed and uninteresting problem of stimulus – response, of what follows what."

Microscopic Level

Microscopic Structure

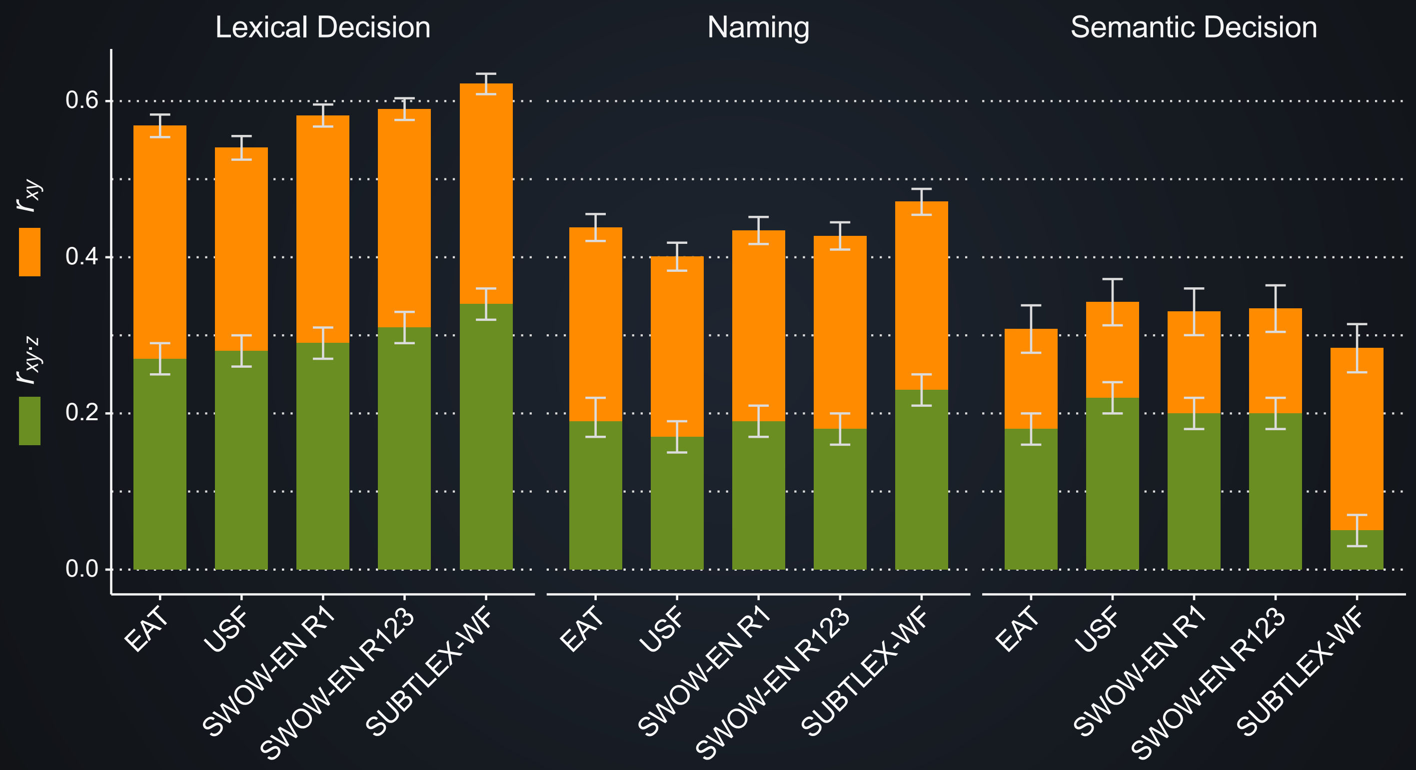
Word association hubs

Dataset	1	2	3	4	5	6	7	8	9	10
SWOW-NL	money	water	food	car	music	pain	tasty	school	child	pretty
SWOW-EN	money	food	water	car	love	music	work	happy	old	bird
EAT	money	water	food	man	car	me	sex	sea	good	home
USF	money	car	water	food	bird	dog	tree	child	good	book

Natural language hubs

Dataset	1	2	3	4	5	6	7	8	9	10
SUBTLEX-NL	go	do	know	come	good	say	see	think	love	sit
SUBTLEX-US	make	do	time	little	people	know	think	way	see	new
BNC	say	see	know	time	year	think	come	last	give	people
KF	say	can	time	new	do	man	make	year	people	little
CELEX-NL	say	see	stand	know	find	time	sit	take	two	woman
CELEX-EN	go	come	say	see	make	stand	good	do	think	big

Centrality in psycholinguistic tasks



Correlation between node in-strength and RT

De Deyne et al. (2019)

Mesoscopic Level

Explore semantic structure

Method

- Network Community Detection
- Modularity optimisation
- Huge directed weighted graphs
- Statistical approach:
  significance of hierarchical overlapping communities

Local Algorithm

Instantiate cluster from a random vector

Iteratively add or remove node to this cluster depending on fitness score calculated for all nodes

Check clusters to be grouped together (fitness of the subgraph)

Fitness score: probability of finding cluster in null model

Results

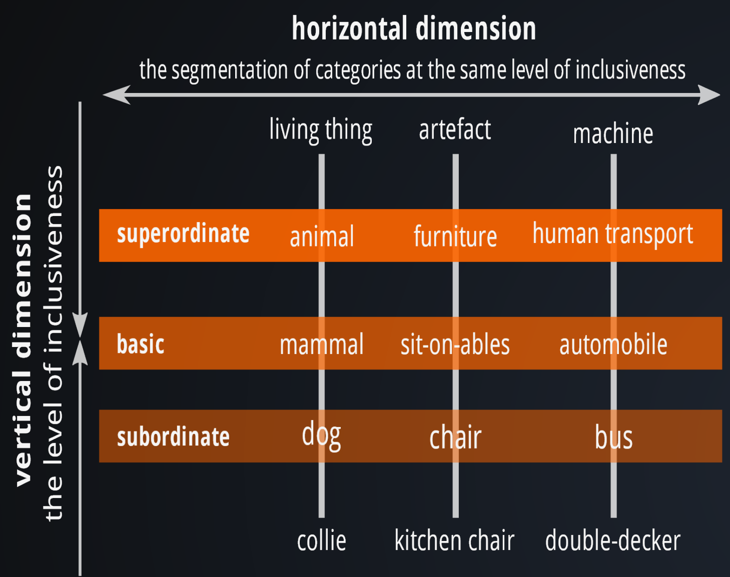
Measure	1	2	3	4	5
# Communities	2	7	37	161	506
Average size	8588	3049	515	112	25
# Homeless nodes	18	18	39	86	380
# Overlapping nodes	5943	6956	5263	4717	1676
Max(overlapping nodes)	2	7	8	10	6
Mean(p)	0	0.062	0.04	0.035	.051

infreq. words: *codex, agalev,…*

homonyms: *bank*
polysemes: *language* (nationality, speech, education, communication)

De Deyne, Verheyen, Perfors & Navarro (2015)

Mesoscopic Structure



- Evidence for hierarchical thematic structure across different POS
- Clusters: partly overlapping with central category labels
- Taxonomic structure present when category exemplars (concrete nouns) are isolated

Widespread Thematic structure

***p* < 0.01**	p = 0.03	p = 0.07	p = 0.08	p = 0.14	p = 0.14
French fries	arrogant	get	beach	mystery	coffee
greasy	haughty	understanding	sea	solve	tea
fryer	haughtily	insight	sun	puzzle	bag
mayonnaise	conceited	understand	holidays	question	cup
fat	snob	awareness	summer	reflect	chocolate milk
unhealthy	unsympathetic	understanding	sand	answer	warm
ketchup	pride	realize	sunbathing	solution	milk
hamburger	vain	see	coast	ask	cocoa
roll	pretension	realize	parasol	problem*	mug
sauce	patronizing	notion	wind	think	tea time
fastfood	look down	know	warm	problem	teapot
fryhouse	superior	conscious	sandcastle	difficult	tea-bag
snack	arrogance	clue	tan	question mark	tableware
croquette	inferior	sink in	bikini	sudoku	cocoa
fries	egocentric	know	sunbathe	quiz	filter
stall	pedantic	click	shell	quiz*	thermos
snack	contempt		terrace	challenge	cappucino
ground beef	spoiled		surf	crossword	coffee
mustard	elitist		dune	guess	visit
sausage	pleased with		Australia	assignment	
kebab	mundane		sea breeze	questionnaire	
...	...		...	...	

Macroscopic Level

Macroscopic Structure

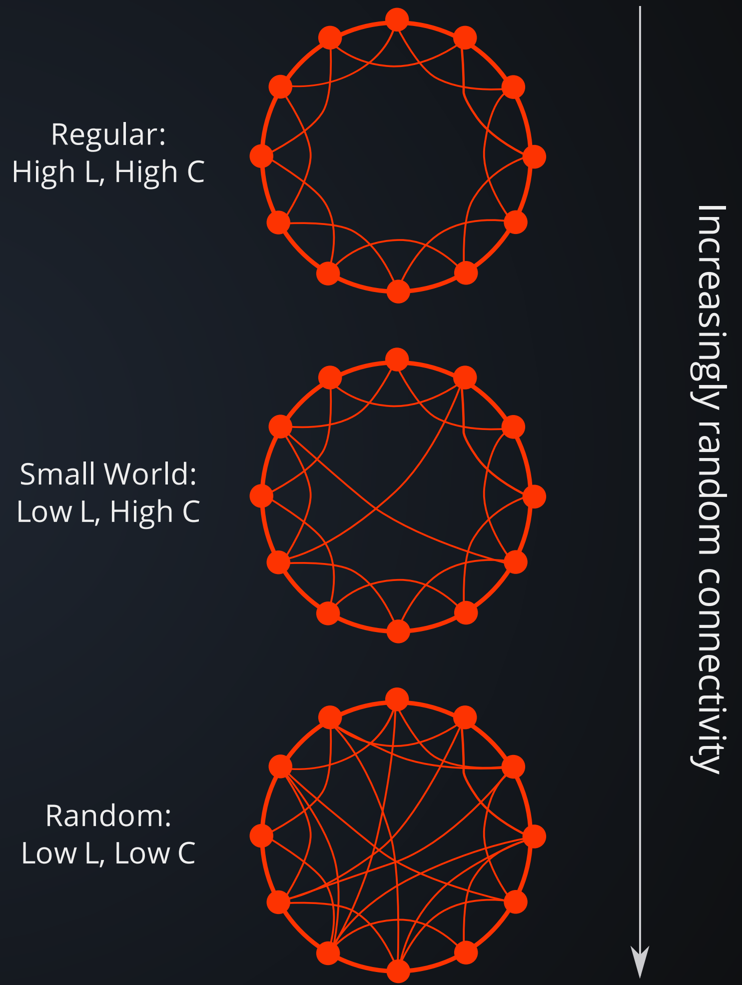
Density and Average Path Length

- G_1 : d = .22, L = 4.14 steps
- G_{12}: d = .43, L = 3.38 steps
- G_{123}: d = .64, L = 3.06 steps

Small-World properties of G_{123}

Strong interconnected regions

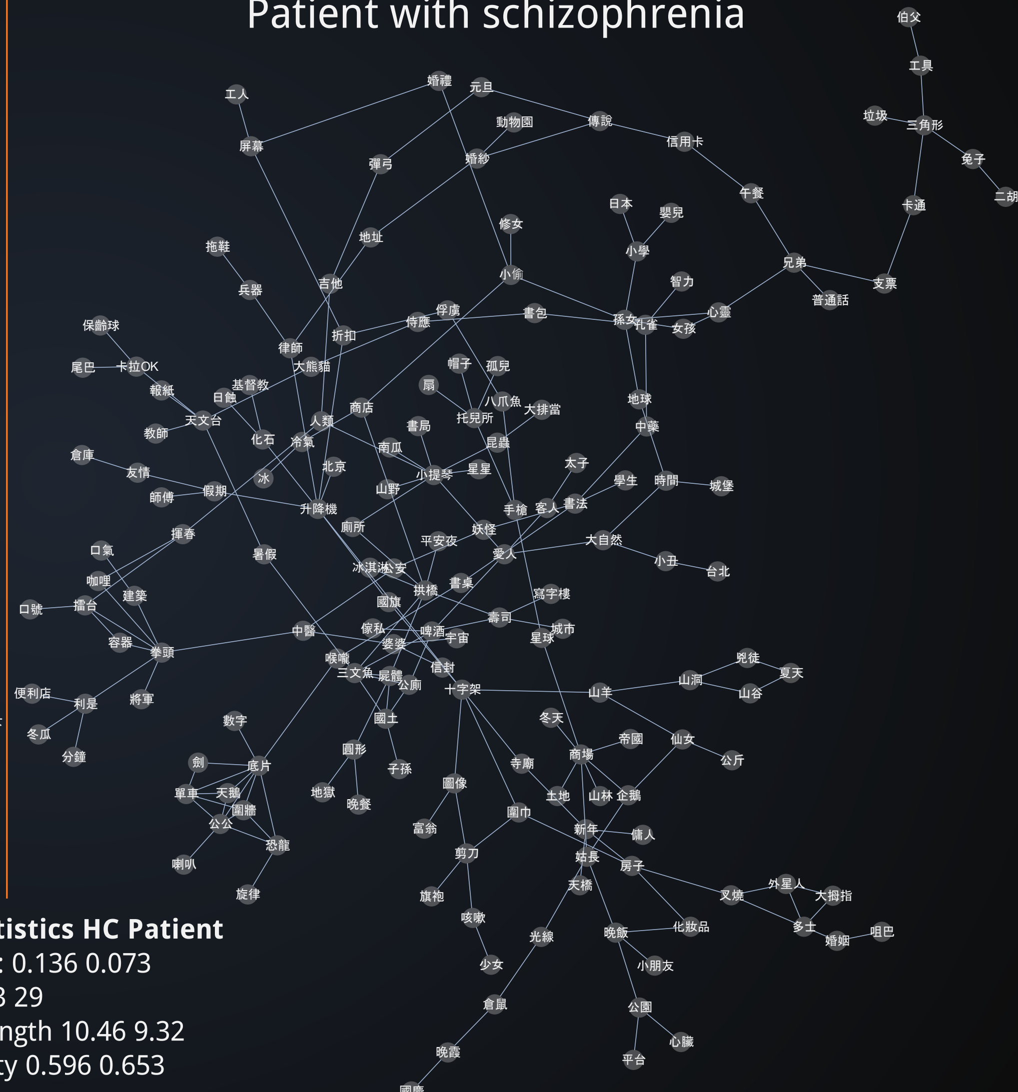
- (CC = .331 vs CC_{rand} =.045)



Applications to language disorders

Healthy control

Patient with schizophrenia



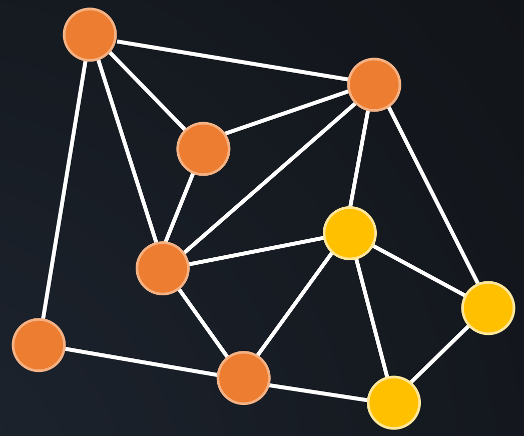
Global Network Statistics	HC	Patient
Clustering coefficient:	0.136	0.073
Network Diameter:	23	29
Characteristic path length	10.46	9.32
Network heterogeneity	0.596	0.653

De Deyne, Elvevåg, Hui, Poon, & Chen (2017)

Connecting Micro, Meso & Macro

1. The case of assortative mixing

Node Assortativity



Assortative mixing

"Likes like like"

- Age, race in social networks
- Measure ≈ correlation coef.
- Different from bias

What word properties determine the presence of an associative link?

Discrete lexical covariates

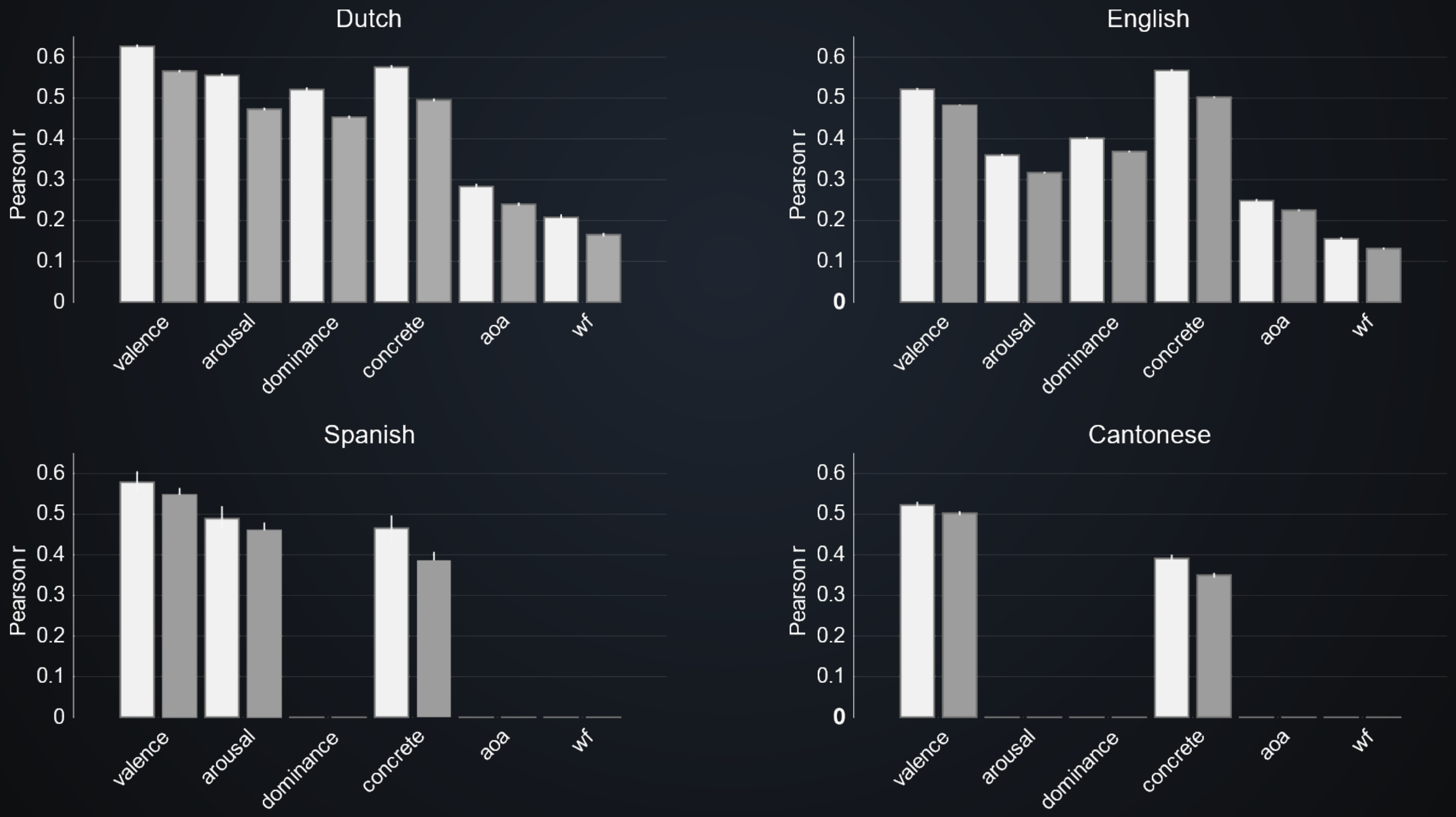
POS: Noun, Verb, Adj
Assortative: paradigmatic
Disassortative: syntagmatic

Scalar lexical covariates

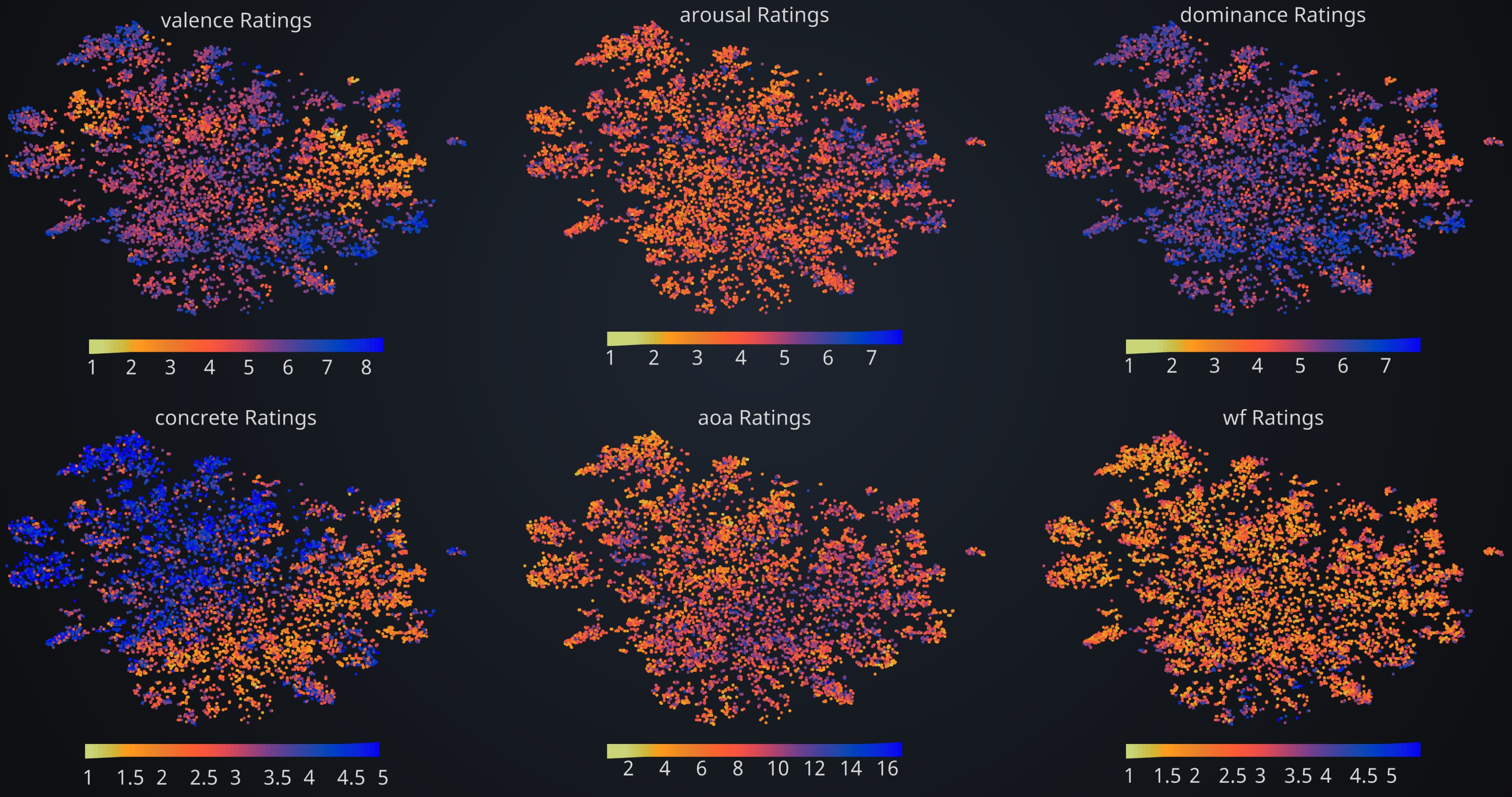
Psycho-affective: Valence, Arousal, Dominance
Semantic: Concreteness, Age-of-Acquisition
Distributional: Word Frequency, Contextual Diversity

Local Assortativity

To what degree do cues and responses exhibit assortativity?



Macroscopic Assortative structure



Implications

1. Local assortativity determines a graph's global structure

2. Global emergent pattern suggests two main organising factors:
   - Valence
   - Concreteness (cf PCA-weights)

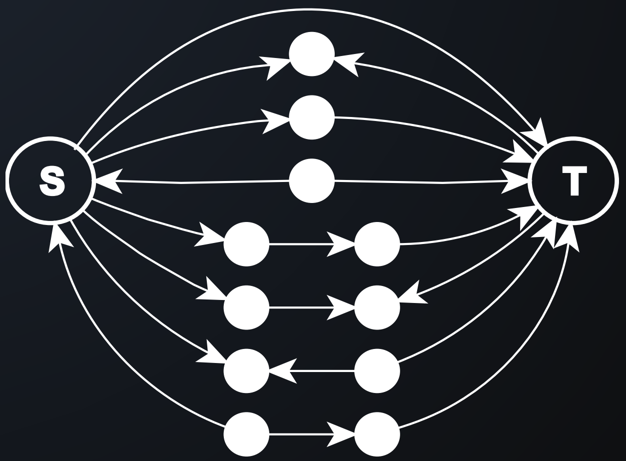
3. Extends Osgoods' findings for nouns to entire lexicon

4. Issue of scale: including all words (not just emotive ones) shows novel contribution of concreteness

Van Rensbergen, Storms & De Deyne, 2016
De Deyne, Navarro & Perfors, in preparation

Connecting Micro, Meso & Macro

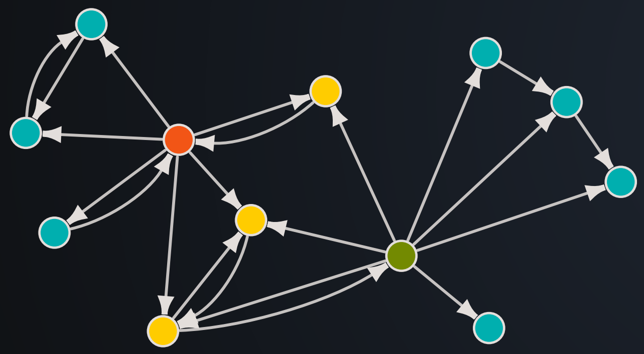
2. Local and Global Similarity



Spreading activation model

Local Similarity Measure

Neighbour distributional overlap
Cosine: Degree of sharing immediate neighbours



Weights

$$\max\left(0, \log_2 \frac{P(resp|cue)n}{P(resp)}\right)$$

Graph-based global similarity

Random walk based measure similar to **Katz Index**
Similar nodes share neighbours that are also similar

Procedure

G converted to a row-normalized transition matrix

$$G_{rw(1)} = I$$

$$G_{rw(2)} = I + \alpha G$$

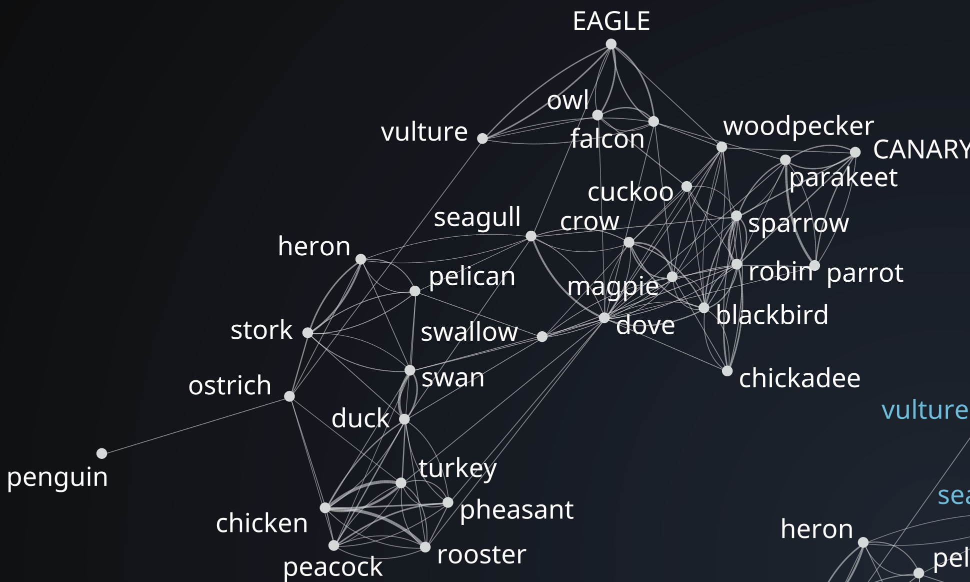
$$G_{rw(3)} = I + \alpha G + \alpha^2 G$$

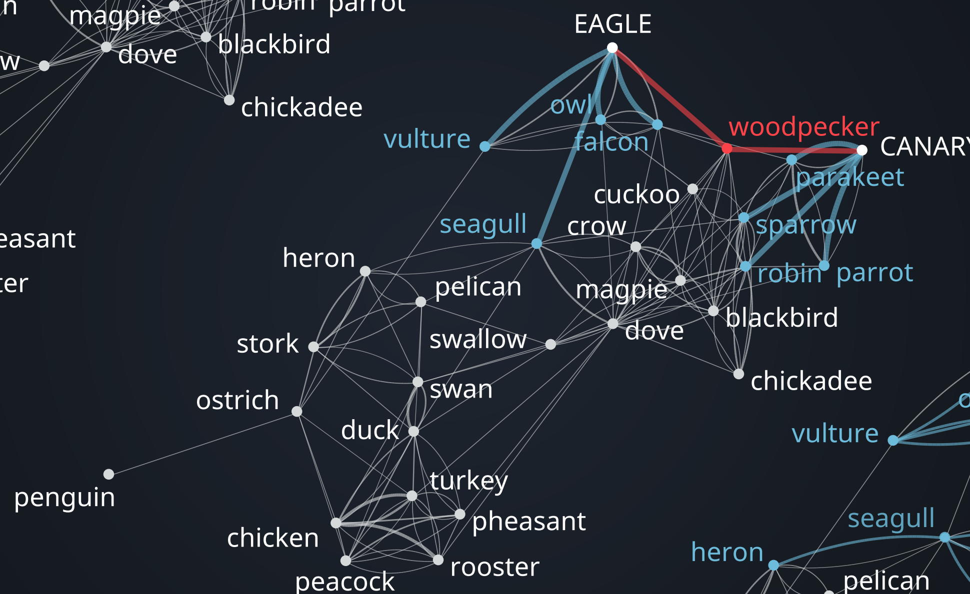
$$in\ the\ limit\ \mathrm{G}_{\infty} = \sum_{r=0}^{\infty} (\alpha G)^r$$

α determines contribution of short over longer paths
➔ limits spread of activation

Paradigmatic NN example

How similar is EAGLE to CANARY?





Local similarity: only one shared neighbor

Indirect paths of length 2: 8 shared neighbors

De Deyne, S., Navarro, D., & Perfors, A. (2018)

Spreading activation example II

Example of new paths for *tiger*

α = 0.25	**α = 0.50**	**α = 0.75**	**α = 0.95**
leopard	leopard	leopard	lioness
safari	safari	hyena	hyena
bear	hyena	lioness	cougar
zebra	lioness	safari	leopard
giraffe	zebra	zebra	devour
fox	giraffe	devour	jungle
	devour	pheasant	Jerry can
	pheasant	carnivore	carbine
	carnivore	cougar	pheasant
	jaguar	bird of prey	bird of prey
$D = 0.01$	$D = 0.03$	$D = 0.08$	$D = 0.10$

Relatedness across the lexicon

Common categories
fruit, insects, tools, …

Domains
artifacts, animals

Arbitrary concepts
nouns

Behavioural

High similarity / inter-subject agr.

Low similarity / inter-subject agr.

Psychologically small, finite number of features shared among individuals (Medin & Ortony, 1989)

Infinite number of ways two concepts can be similar (Goodman, '72, Medin, Goldstone & Gentner, '93)

Explanations

Lexico-semantic view

Humans continuously track weak linguistic contingencies lead to stable structure shared among individuals across a range of concept relations

Empirically no clear evidence, and issue of sparsity of the input

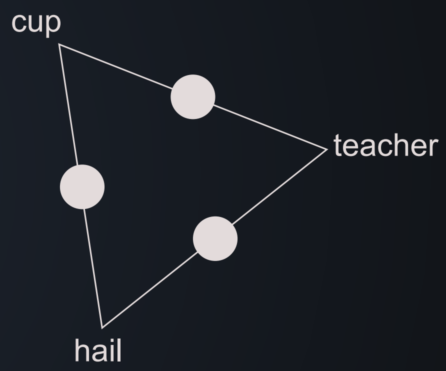
Structure in remote concepts

Stimuli

- 100 randomly generated triads
- Examples:

  Sunday - vitamin - idiot

  dark - heat - shot
- Same frequency band, same concreteness range
- Mostly weakly related pairs



Method

"Choose strongest related pair"

Structure in domains and categories

DOMAIN TRIADS

Stimuli

- Animals, Artifacts, Food, Activities
- Words occur only once

pigeon - sardine - butterfly

jeep - gun - sweater

deer - peacock - salmon

BASIC LEVEL TRIADS

Stimuli

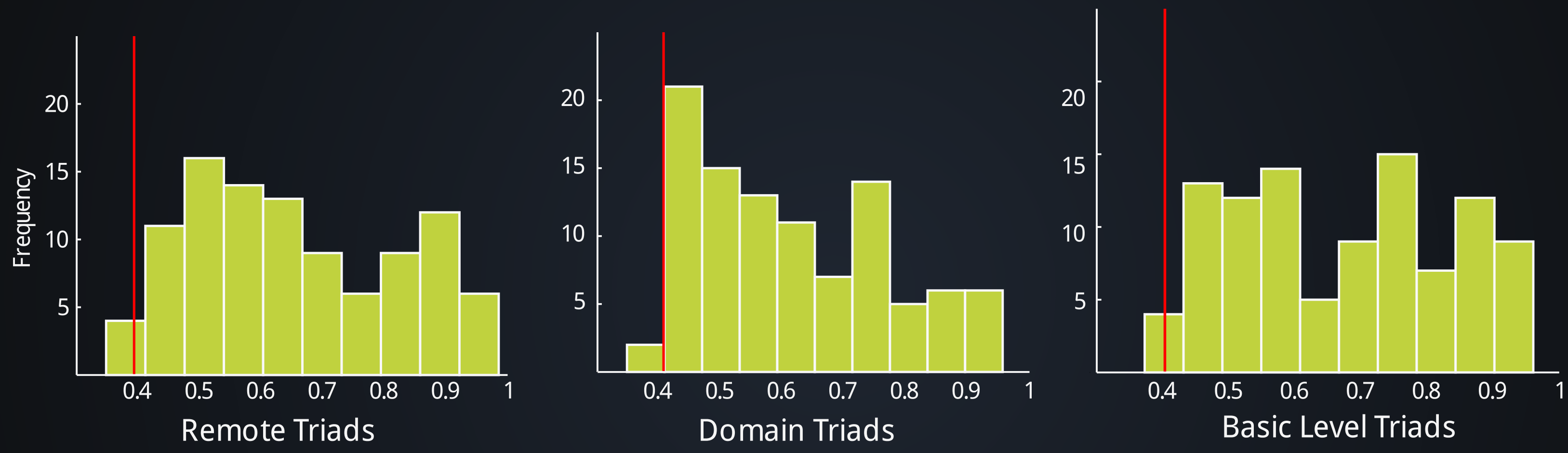
- taken from De Deyne et al. (08)
- 4 x Animals, 6 x Artifacts
  2 x Food, 2 x Activities

goldfish - shark - pike

donkey - horse - zebra

onion - cauliflower - parsley

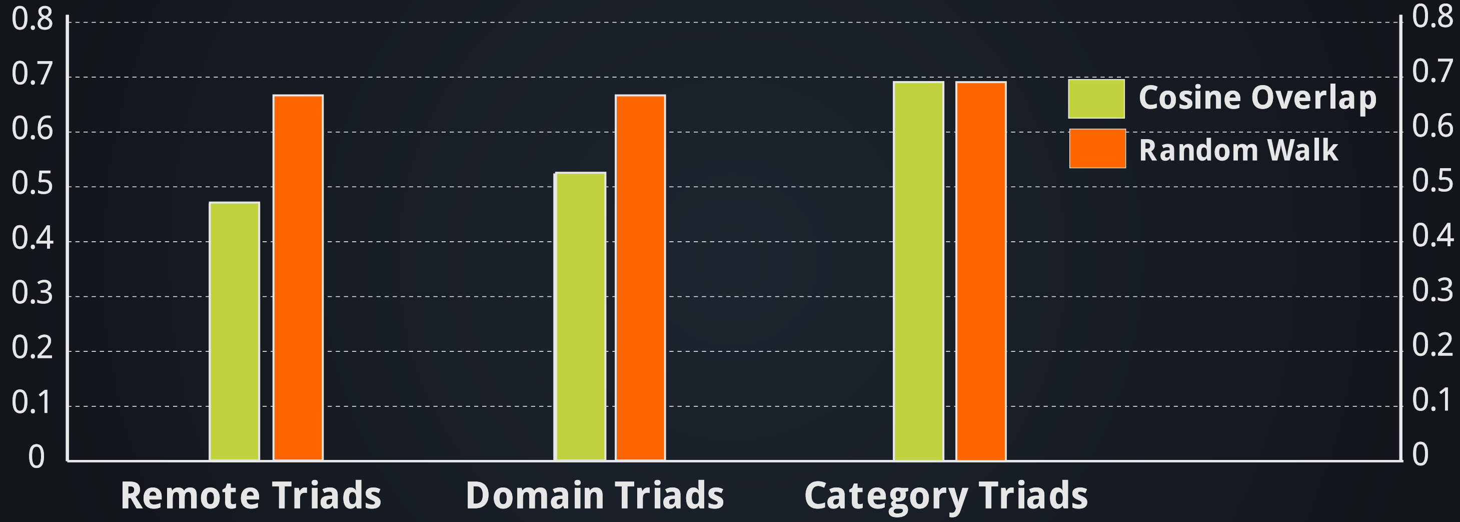
Do participants agree about remote, domain and basic triads?



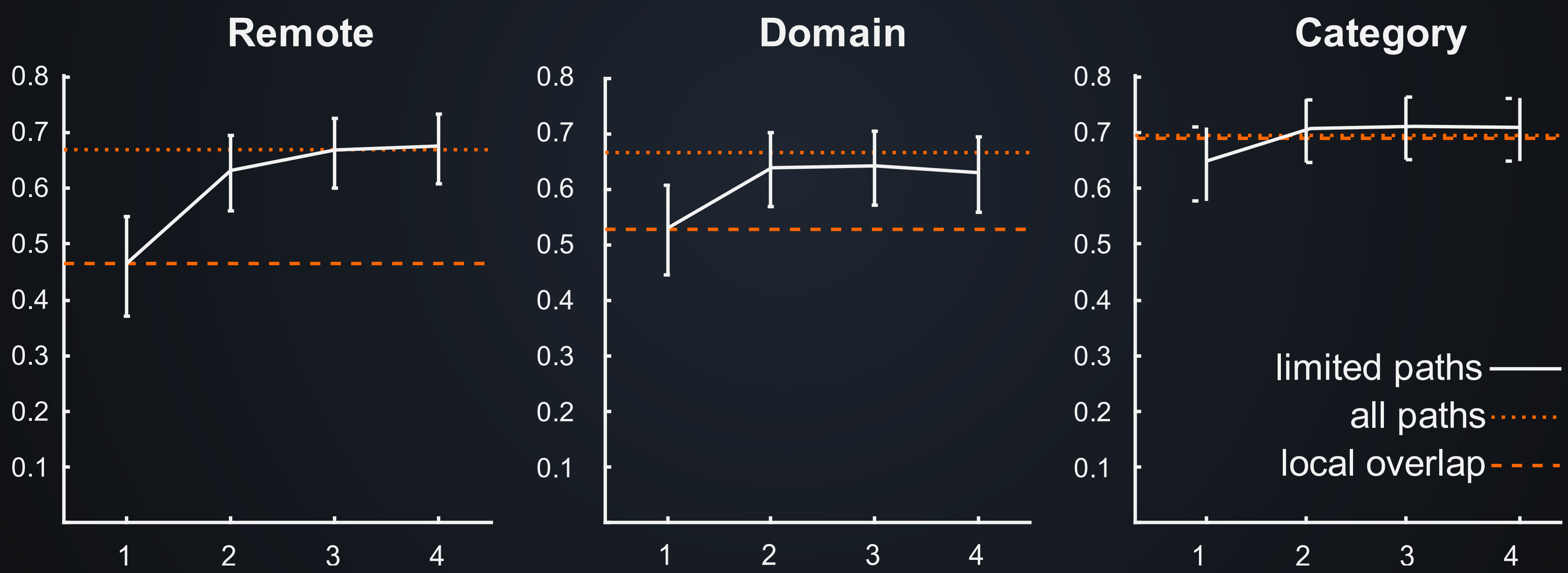
Dominant choice probability and 95% confidence boundary for random choice (N=100)

→ Agreement among participants preferences for majority of triads

Overview model findings



The role of path length



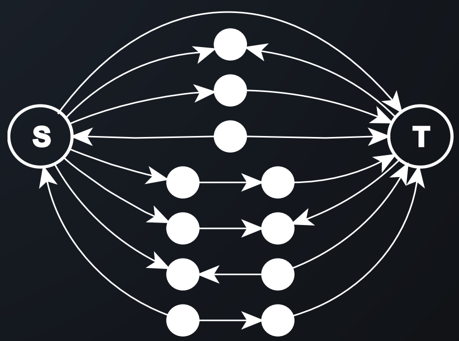
Interim conclusions

Methodological

- Random walk: interesting way to model spreading activation, using little heuristics
- Improvements across a range of relations (remote, domains, basic)
- Limited walks address concerns by extent of spreading activation throughout network
- Strong test for inference in lexico-semantic models

Theoretical

- Similarity:
  Sufficiently constrained between remote pairs
- A connected homogeneous view of lexicon bridging associative and semantic meaning



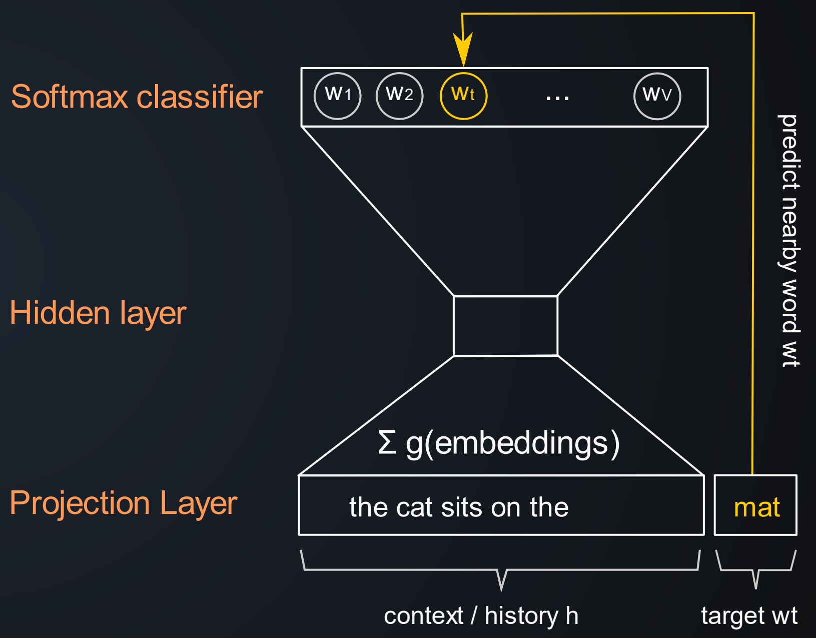
Evaluating distributional internal models

Problem: Sparsity of the input

- Extreme for single responses
- Moderate for continued responses?

External Language model benchmark

- Count model: PPMI word co-occurrences
- Prediction model: continuous bag of words (word2vec, Mikolov et al)



Evaluation

Model parameters

External Language Corpus models

- PPMI: window-size: 3
- w2v: window-size: 7
- w2v: 400 dimensions

Internal Word Association models

- RW: α=.75

WordSim-353A (Agirre et al, 2009)
MTURK-771 (Halawi et al, 2012)
Radinsky2011 (Radinsky et al., 2011)
RG1965 (Rubenstein & Goodenough, 1965)
MEN (Bruni et al., 2012)

WordSim-353A (Agirre et al, 2009) *
SimLex-999 (Hill et al, 2016) **

Similarity and relatedness

Relatedness

car - petrol

frequently occur together in space and language (functional relationship)

Similarity

car - bike

common physical features (e.g. wheels)
common function (transport)
share clearly definable category (modes of transport).

** similarity, ** strict similarity*

Basic Level Triads

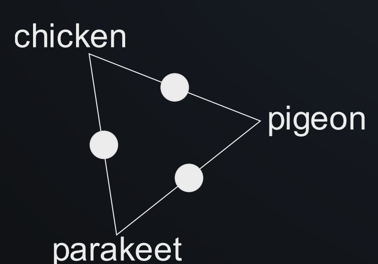
Two ideas

1. Basic level concrete categories: most informative taxonomic level, predominantly perceptual and early acquired (Rosch et al. 1976)
2. Basic level abstract categories: no perceptual info, but meaning acquired through grounding internal affective states (Vigliocco et al, 2009)

Concrete basic triads

100 unique items

categories: fruit, tools, insects, ...



Abstract basic triads

100 unique items

categories: feelings, states, ideas, sciences, ...

Results

	n	n(overlap)	Linguistic Model		Word Association Model	
			Count	w2v	Count	RW
Similarity Judgements						
WordSim-353B	203	175	.74	.79	.84	.87
SimLex-999	998	927	.37	.43	.70	.68
Relatedness Judgements						
WordSim-353A	252	207	.67	.70	.77	.82
MTURK-711	778	771	.67	.71	.81	.83
Radinsky2011	287	137	.75	.78	.74	.79
RG1965	65	52	.78	.83	.93	.95
MEN	3000	2611	.75	.79	.85	.87
Remote relatedness judgments						
Remote triads	300	300	.65	.52	.62	.74
Average						

Discussion

Overall performance

- Word associations:
  - Excellent predictions considering attenuation by unreliability
  - Only exception: “pure” similarity judgments (SimLex)
  - (Poor predictions for direct association)

- Linguistic models:
  - Consistent improvement for prediction models, except for remote pairs

Published vectors

- SNAUT (Mandera et al., 2016), r = .69
- GloVe (Pennington et al, 2014)
  - 6B tokens, r = .64
  - 64B tokens, r = .67
  - 840B tokens, r = .70

De Deyne, Perfors, & Navarro (2016)

Why language networks…

1. Multiple levels of analysis: structure at every scale (local – global)

2. Complex networks: more is different: levels tell a distinct story

3. Multiple relations: interactions between syntax, morphology, phonology and semantics

4. Networks as a <u>tool</u> in quantitative linguistics: often simplification, but scale + transparency

My collaborators

- Belgium: Gert Storms, Steven Verheyen, Guillem Collell, Marc Brysbaert
- Australia: Andy Perfors, Danielle Navarro
- SWOW-team: Francesca! Maria!
- Australian Research Council

Questions - simon.dedeyne@adelaide.edu.au

Data - https://smallworldofwords.org/project/

Online study - https://smallworldofwords.org/

small world of words

Discover what **words** mean for **people** worldwide

Selected readings

- Baronchelli, A., Ferrer-i-Cancho, R., Pastor-Satorras, R., Chater, N., & Christiansen, M. H. (2013). Networks in cognitive science. *Trends in cognitive sciences*, *17*(7), 348-360.
- Biemann, C., Krumov, L., Roos, S., & Weihe, K. (2016). Network motifs are a powerful tool for semantic distinction. *In Towards a Theoretical Framework for Analyzing Complex Linguistic Networks* (pp. 83-105). Springer.
- Stella, M., Beckage, N. M., Brede, M., & De Domenico, M. (2018). Multiplex model of mental lexicon reveals explosive learning in humans. *Scientific Reports*.
- Steyvers, M., & Tenenbaum, J. B. (2005). The Large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive science*, *29*(1), 41-78.
- Watts, Duncan J.; Strogatz, Steven H. (1998). "Collective dynamics of 'small-world' networks". *Nature*. 393 (6684): 440–442.

Tutorial

Entirely optional: You can follow along in RStudio.

1. install.packages('devtools')
2. library(devtools)
3. install_github('SimonDeDeyne/workshop')

Caveat: This is work in development and might not work!