Institute for Web Science and Technologies
University of Koblenz–Landau, Germany

Linguistic Network Analysis
with the Koblenz Network Collection

Jérôme Kunegis

Modeling Linguistic Networks 2012

KONECT – The Koblenz Network Collection

KONECT currently holds
180 networks, of which

- 36 are undirected,
- 51 are directed,
- 93 are bipartite,
- 63 are unweighted,
- 100 allow multiple edges,
- 7 have signed edges,
- 9 have ratings as edges,
- 1 allows multiple weighted edges,
- and 87 have edge arrival times.

KONECT > **Networks**

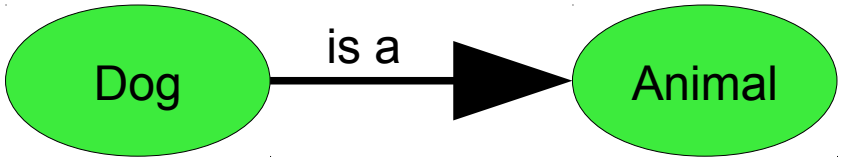
Networks

konect.uni-koblenz.de

Code	Name	Category	Format	Weights	Metadata	Size	Volume	Average degree	Download
AD	Advogato	Social	D	▬		6,535	51,397	7.86	
AM	Amazon	Interaction	U	▬		403,394	3,387,388	8.40	
AP	arXiv astro-ph	Interaction	U	▬		18,772	396,160	21.10	
AC	arXiv cond-mat	Authorship	B	▬		38,741	58,595	3.50	
PH	arXiv hep-ph	Interaction	U	▬	⊘	28,093	12,730,098	453.14	
PHc	arXiv hep-ph	Reference	D	▬		34,546	421,578	12.20	
THc	arXiv hep-th	Reference	D	▬		27,770	352,807	12.70	
TH	arXiv hep-th	Interaction	U	▬	⊘	22,908	11,209,368	489.32	
th	arXiv hep-th cit	Reference	D	▬		27,770	352,807	12.70	
BAI	Baidu	Reference	D	■		2,141,300	17,794,839	8.31	
BAr	Baidu	Reference	D	■		415,641	3,284,387	7.90	
BS	Berkeley/Stanford	Reference	D	▬		685,230	7,600,595	11.09	
Btl	BibSonomy ti	Folksonomy	B	▬	⊘	975,963	2,555,080	12.48	
Bul	BibSonomy ui	Folksonomy	B	▬	⊘	777,084	2,555,080	440.99	
But	BibSonomy ut	Folksonomy	B	▬	⊘	210,467	2,555,080	440.99	
PM	Caenorhabditis elegans	Interaction	U	■		453	4,596	10.15	
IN	CAIDA	Physical	U	▬		26,475	106,762	4.03	
RO	California	Physical	U	▬		1,965,206	5,533,214	2.82	
Sc	Catster	Social	U	▬		149,700	5,449,275	36.40	
Scd	Catster/Dogster	Social	U	▬		624,127	15,705,337	25.16	
CS	CiteSeer	Reference	D	▬		723,131	1,764,929	2.44	
Ctl	CiteULike ti	Folksonomy	B	■	⊘	885,046	2,411,819	15.74	
Cul	CiteULike ui	Folksonomy	B	■	⊘	754,484	2,411,819	106.16	
Cut	CiteULike ut	Folksonomy	B	■	⊘	175,992	2,411,819	106.16	
CN	Countries	Affiliation	B	■		512,781	557,587	1.09	
PI	DBLP	Reference	D	▬		12,591	49,793	3.95	
Pc	DBLP	Interaction	U	■	⊘	1,103,412	14,703,760	13.33	
Pa	DBLP	Authorship	B	▬		4,337,293	6,651,962	5.72	
DB	DBpedia	Semantic	D	■		2,152,642	7,494,124	3.48	
Dtl	Delicious ti	Folksonomy	B	■	⊘	33,792,636	301,254,171	8.92	
Dul	Delicious ui	Folksonomy	B	■	⊘	834,679	301,254,171	361.61	
Dut	Delicious ut	Folksonomy	B	■	⊘	5,368,293	301,254,171	361.61	
DG	Digg	Communication	D	■	⊘	30,398	87,627	2.88	
Sd	Dogster	Social	U	▬		426,820	8,546,581	20.02	
EA	Edinburgh Associative Thesaurus	Lexical	D	■		23,219	312,342	13.45	
EN	Enron	Communication	D	■	⊘	87,273	1,148,072	13.15	

Types of Linguistic Networks

(1) Semantic network



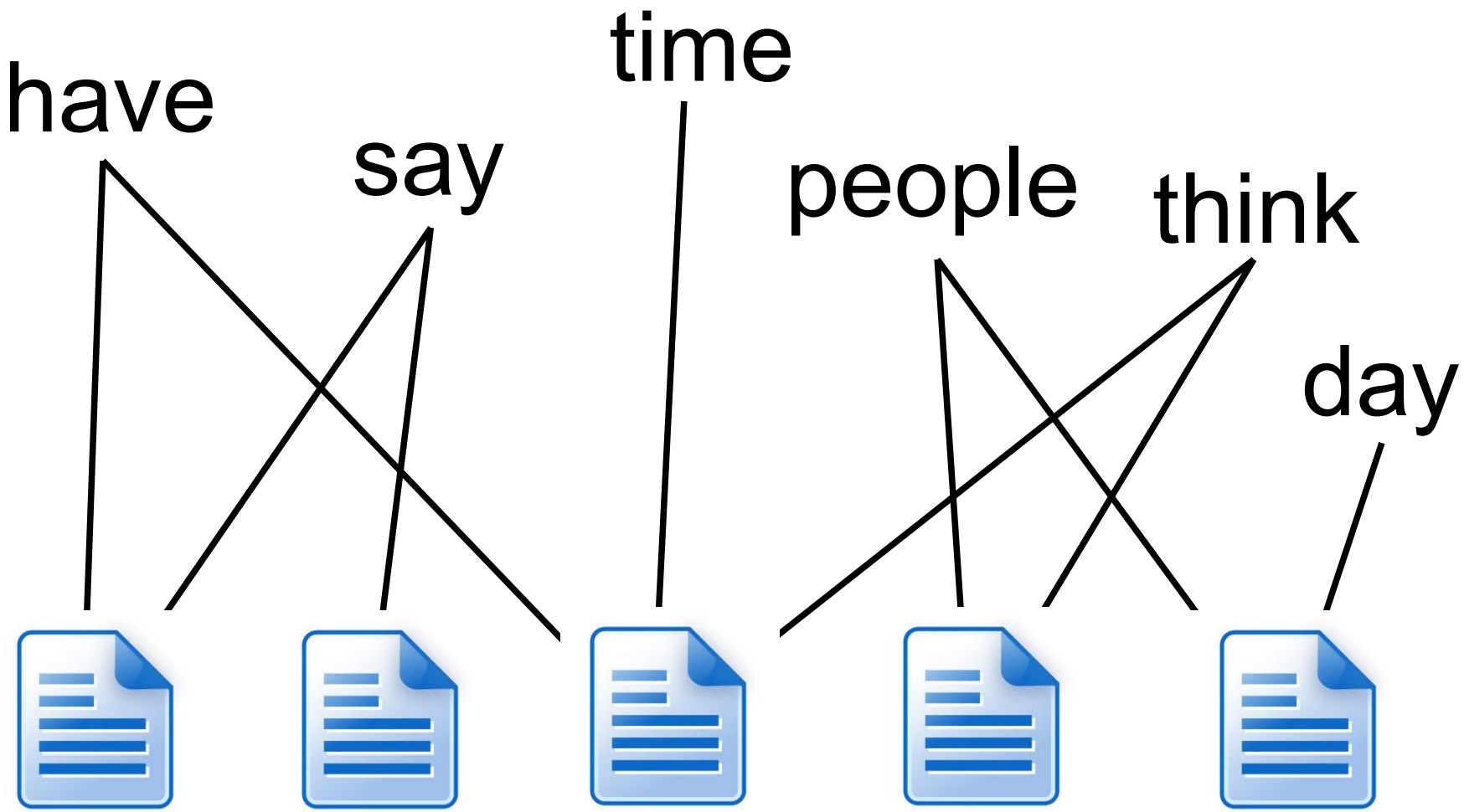
(2) Adjacency network

"Whether I shall **turn out** to be the hero of my own life . . ."

(3) Association network

"Turn" → "Revolve"

Not in This Talk: Word–Document Networks

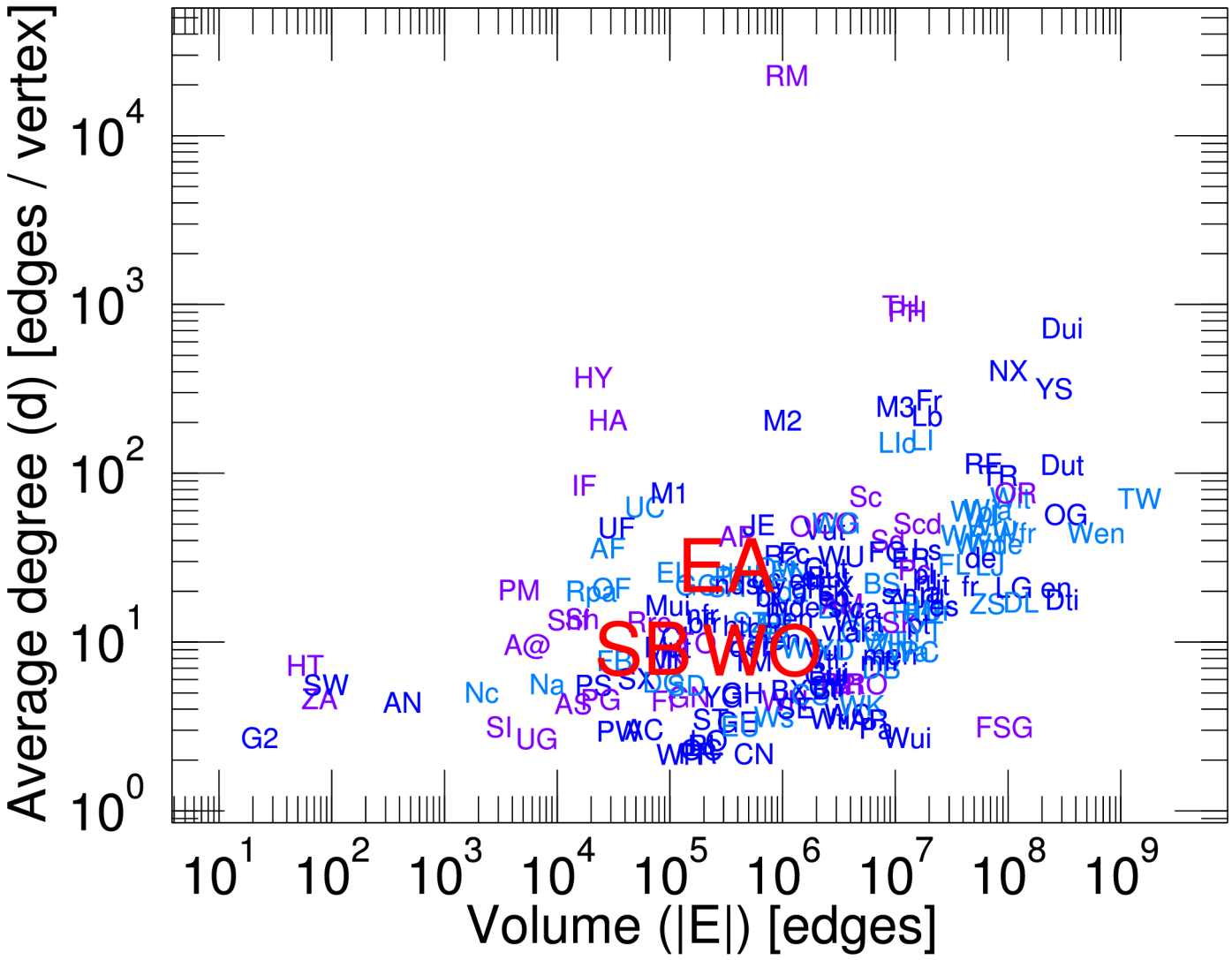


Datasets

Type	Network	#Nodes	#Edges	Avg. degree	Max. degree
Semantic	WordNet (WO)	144,511	643,863	8.91	991
Adjacency	Spanish book (SB)	12,643	57,772	9.14	5,542
Association	Edinburgh Associative Thesaurus (EA)	23,219	312,342	26.90	1,898

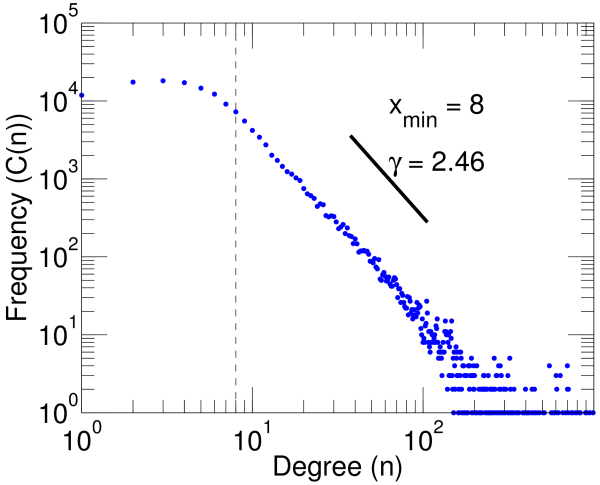
http://konect.uni-koblenz.de/networks/wordnet
http://konect.uni-koblenz.de/networks/lasagne-spanishbook
http://konect.uni-koblenz.de/networks/eat

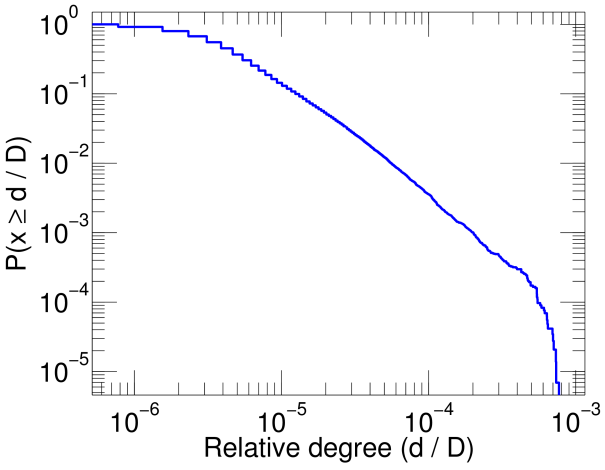
Overview


Average degree (d) [edges / vertex]
Volume (|E|) [edges]
10^0
10^1
10^2
10^3
10^4
10^5
10^6
10^7
10^8
10^9
RM
HY
HA
M2
M3
Fr
Lb
Dui
NX
YS
IF
M1
UC
UF
AF
Sc
JE
Scd
TW
OG
Wen
EA
SBWO
PM
HT
SW
ZA
AN
Nc
Na
A@
SI
UG
G2
CN
Wui
FSG
ZSDL
Dti
LG
en

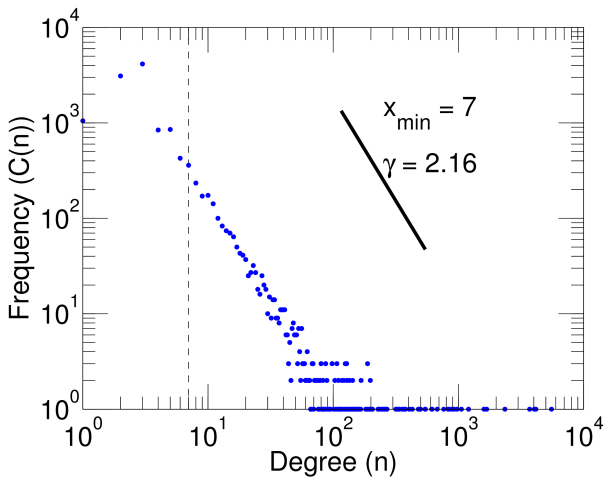
Degree Distributions: Are They Power Laws?

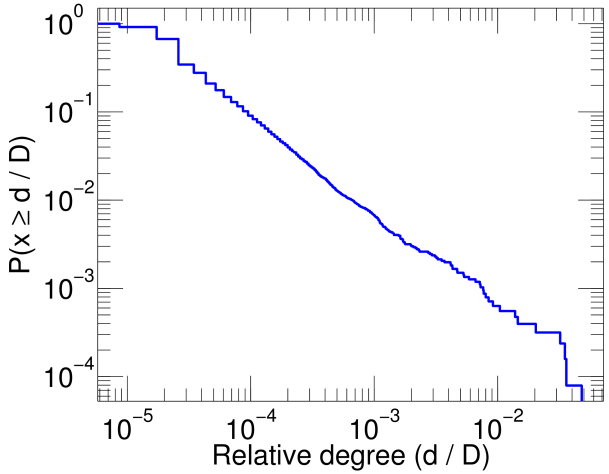
WordNet



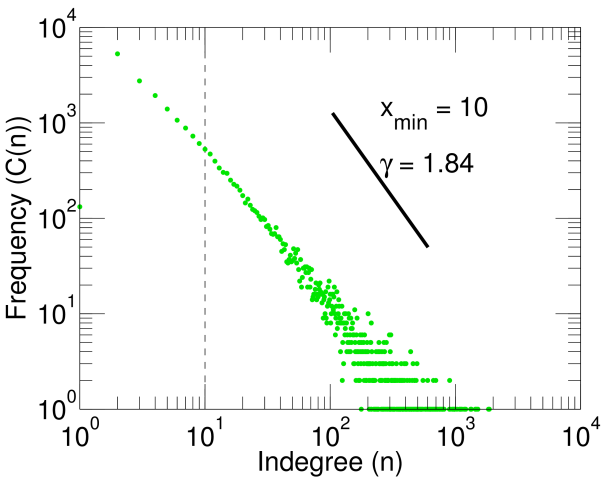


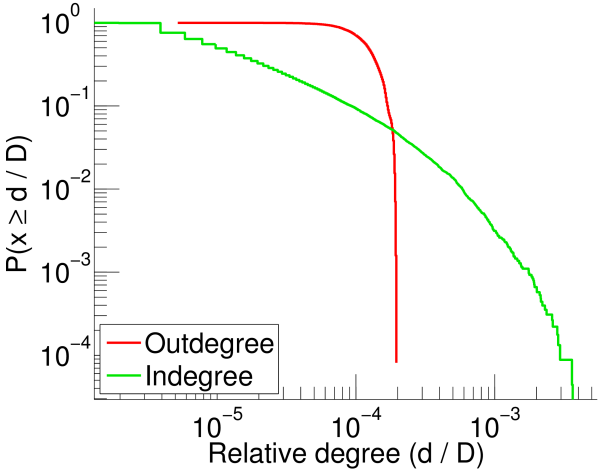
Spanish Book





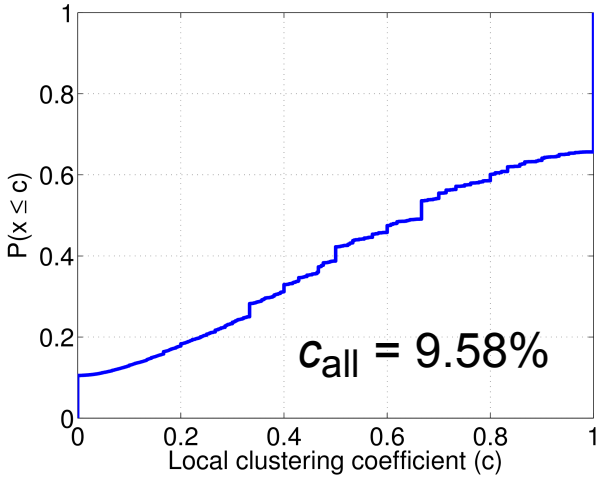
E. Associative Thesaurus

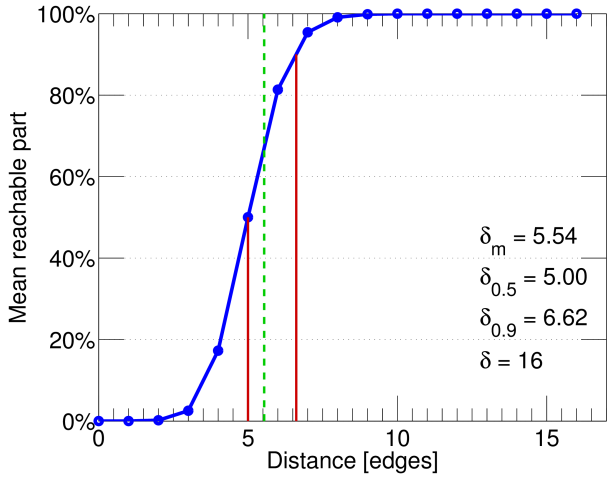




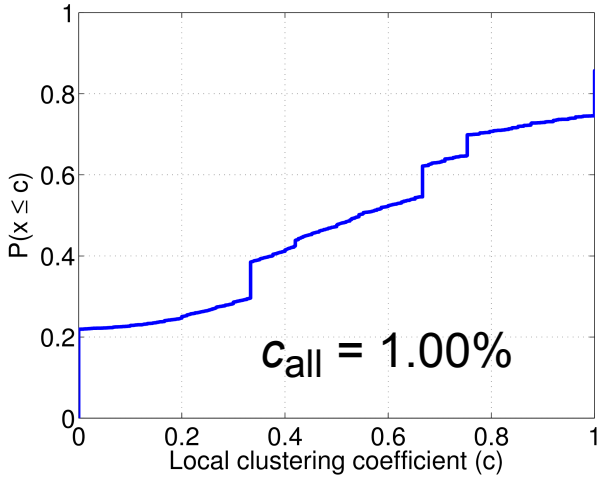
Is It a Small World?

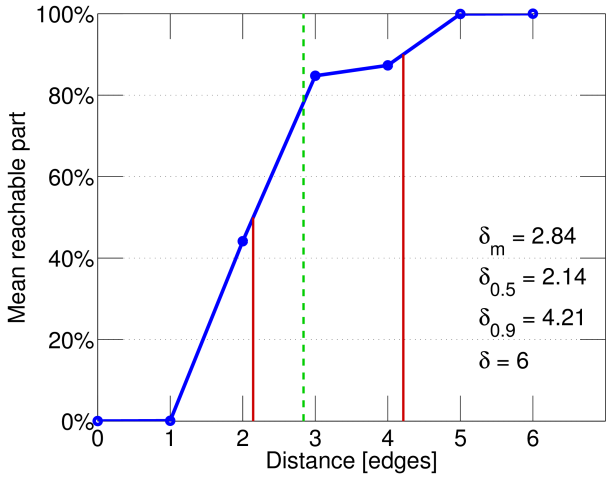
WordNet



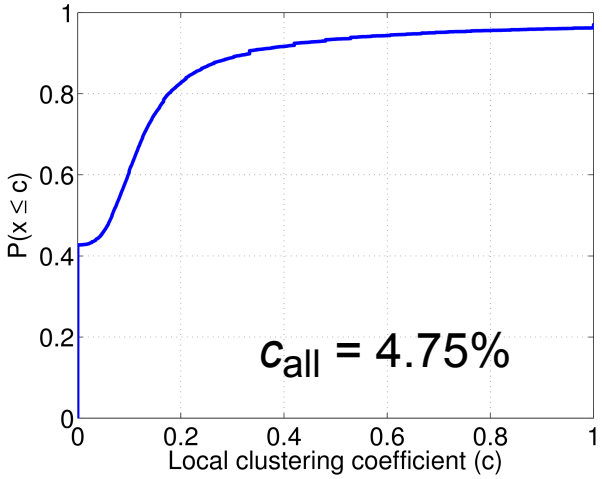


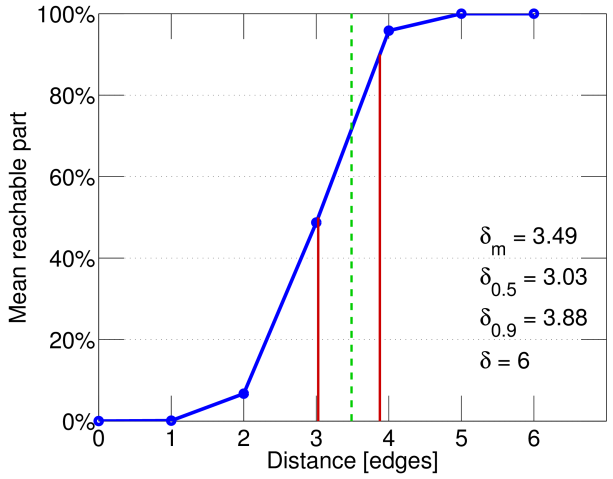
Spanish Book



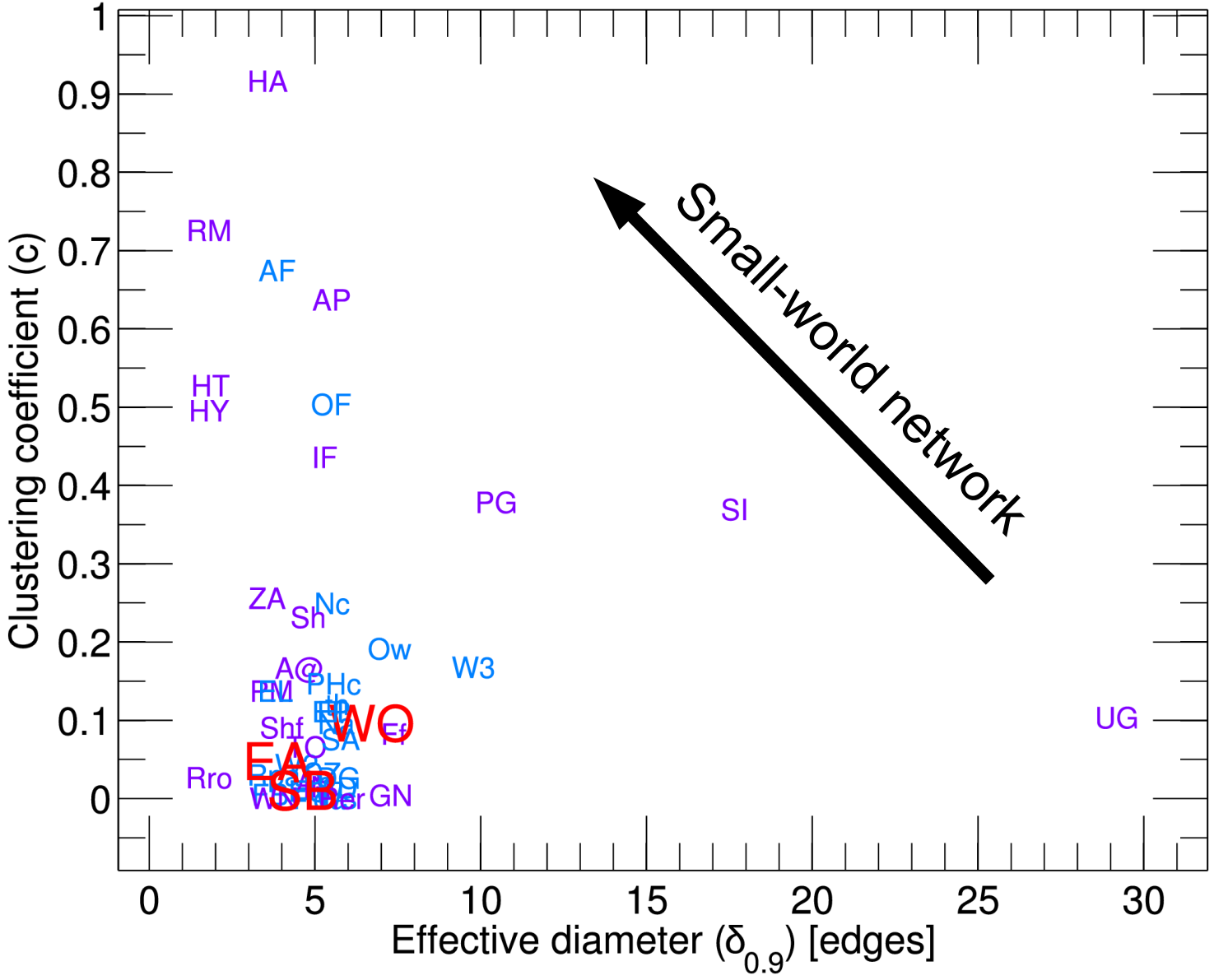


E. Associative Thesaurus



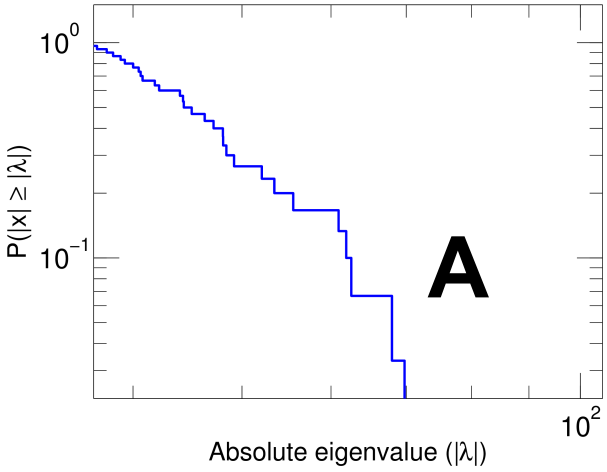


Is It a Small World? (2)

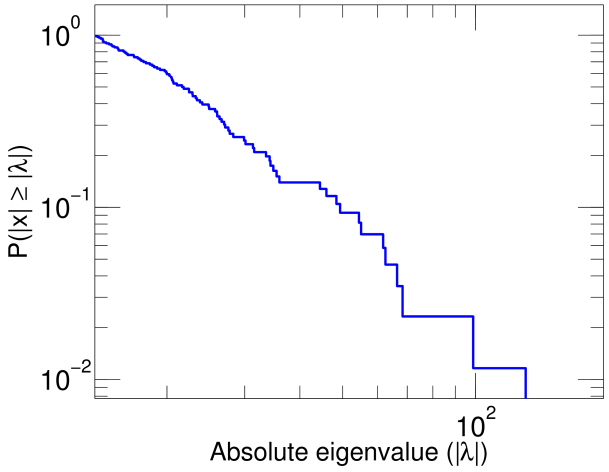


Spectral Analysis

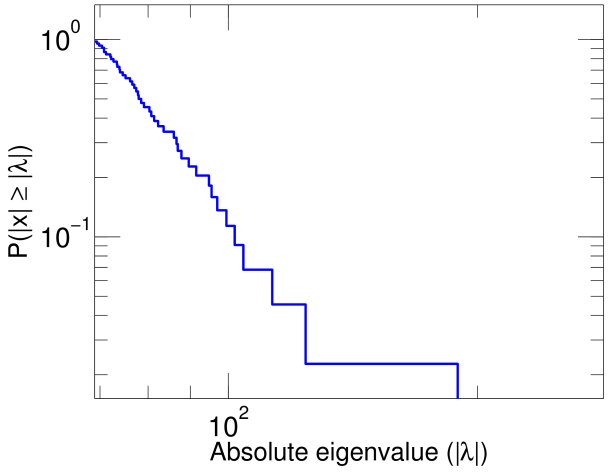
WordNet

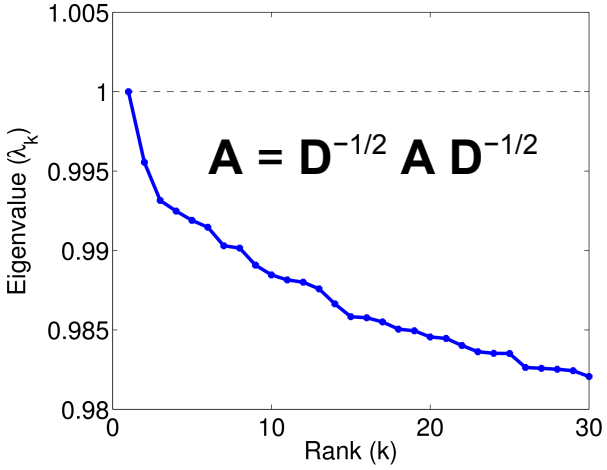


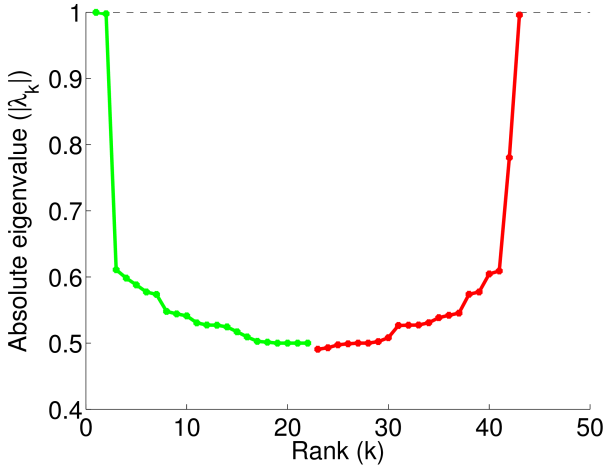
Spanish Book

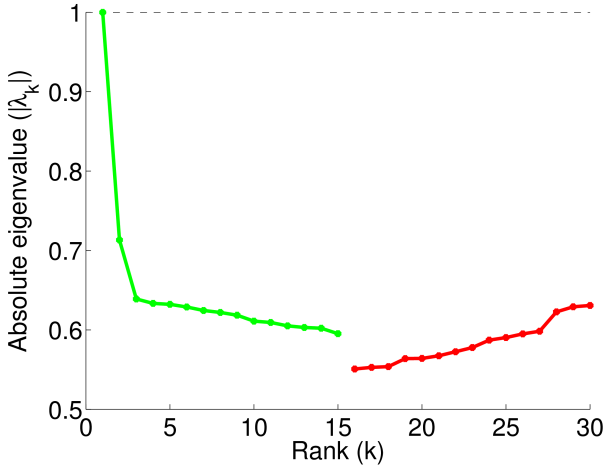


E. Associative Thesaurus









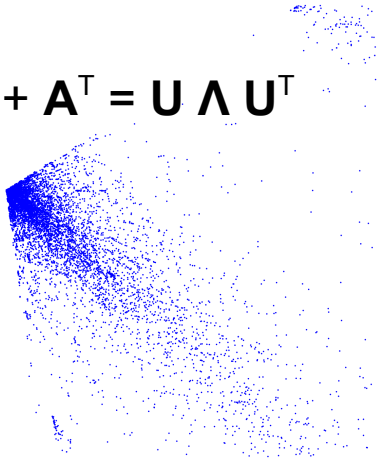
Visualization

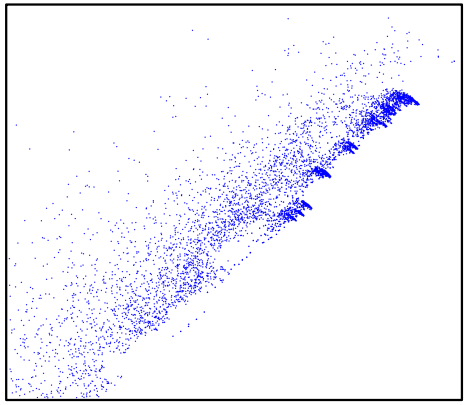
WordNet

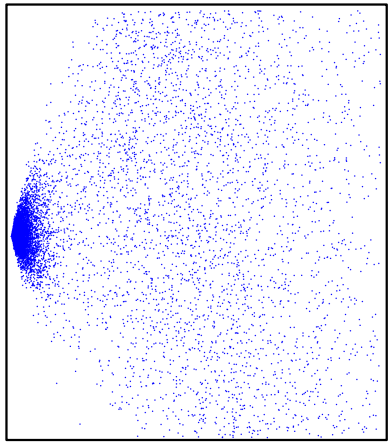
Spanish Book

E. Associative Thesaurus

$\mathbf{A} + \mathbf{A}^T = \mathbf{U}\,\mathbf{\Lambda}\,\mathbf{U}^T$

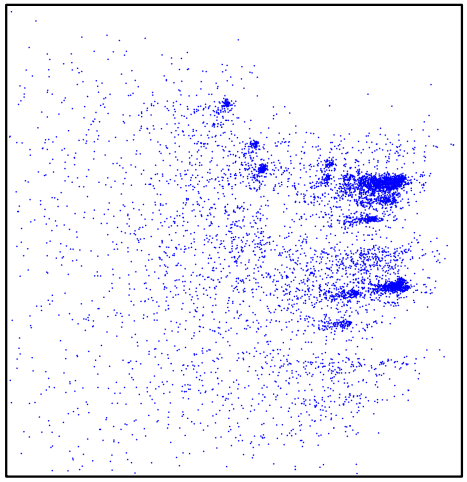


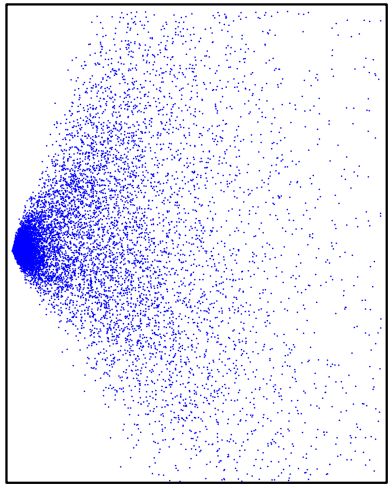




$\mathbf{A} + \mathbf{A}^T + i(\mathbf{A} - \mathbf{A})^T = \mathbf{U}\,\mathbf{\Lambda}\,\mathbf{U}^T$

(only directed networks)





Jérôme Kunegis
@kunegis

Contribute to KONECT
– Propose/donate datasets
– Propose/donate analyses
– Propose/donate plots

Tools
– Matlab

Thank You

konect.uni-koblenz.de