

# Network Analysis Tools for Online Communities: The Koblenz Network Collection

Jérôme Kunegis

Institute for Web Science and Technologies (WeST), University of Koblenz-Landau

With acknowledgments to everyone who has made network datasets available to the public



People and Knowledge Networks



INFORMATIK 2013 – Informatik angepasst an Mensch, Organisation und Umwelt  
43. Jahrestagung der Gesellschaft für Informatik | 16. bis 20. September 2013 | Koblenz



# You Are Here



## On the Spectral Evolution of Large Networks

Jérôme Kunegis

Institute for Web Science and Technologies  
University of Koblenz-Landau  
kunegis@uni-koblenz.de

November 2011

Vom Promotionsausschuss des Fachbereichs 4: Informatik der Universität  
Koblenz-Landau zur Verleihung des akademischen Grades

Doktor der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation.

PhD thesis at the University of Koblenz-Landau.

Datum der wissenschaftlichen Aussprache:	9. November 2011
Vorsitz des Promotionsausschusses:	Prof. Dr. Karin Harbusch
Berichterstatter:	Prof. Dr. Steffen Staab
Berichterstatter:	Prof. Dr. Christian Bauckhage
Berichterstatter:	Prof. Dr. Klaus Obermayer

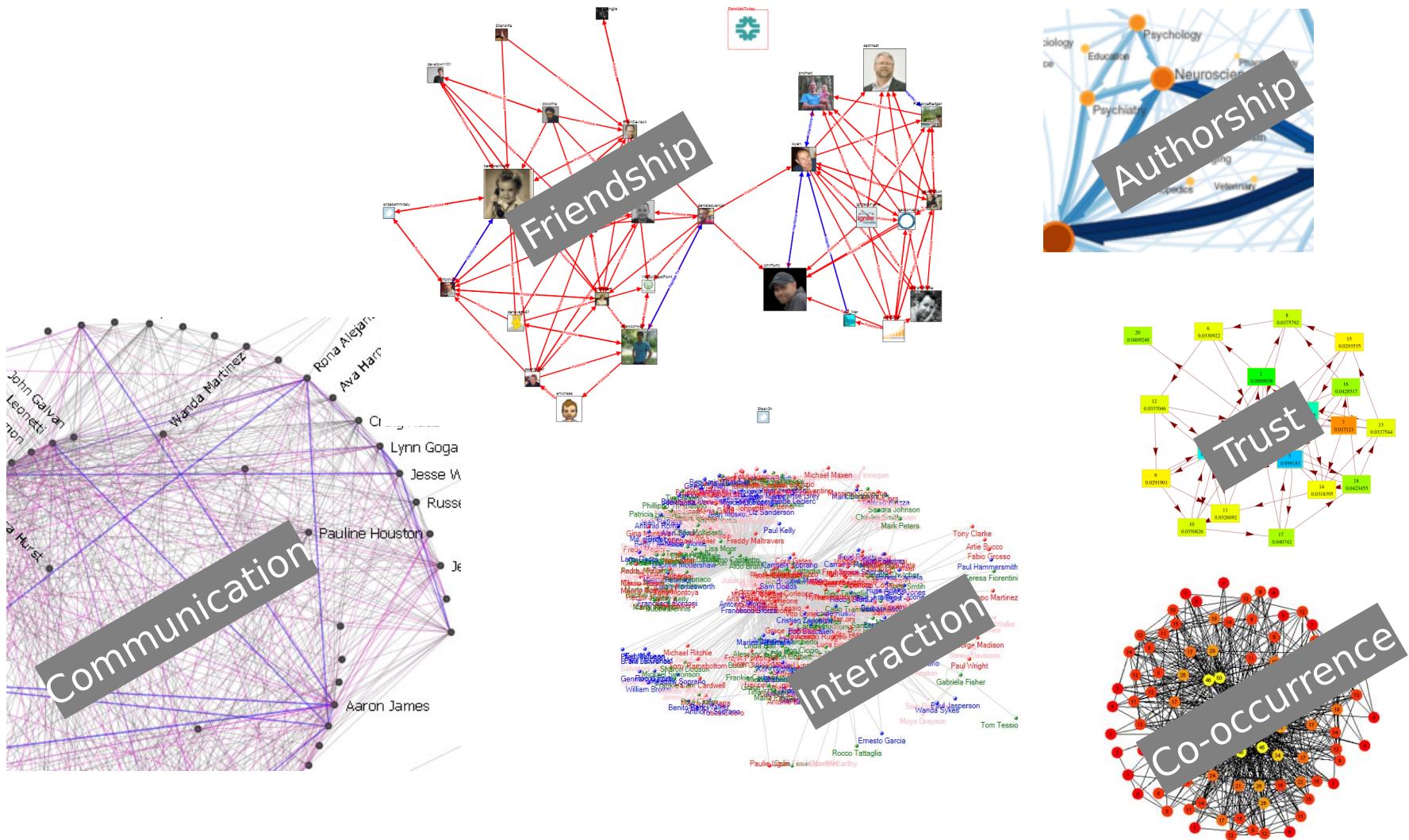
# The Trick Is...



# Everything Is a NETWORK !



# In Fact, Communities Are Everywhere



# A Community Is Like a Gummi Bear



# A Community Is Like a Gummi Bear



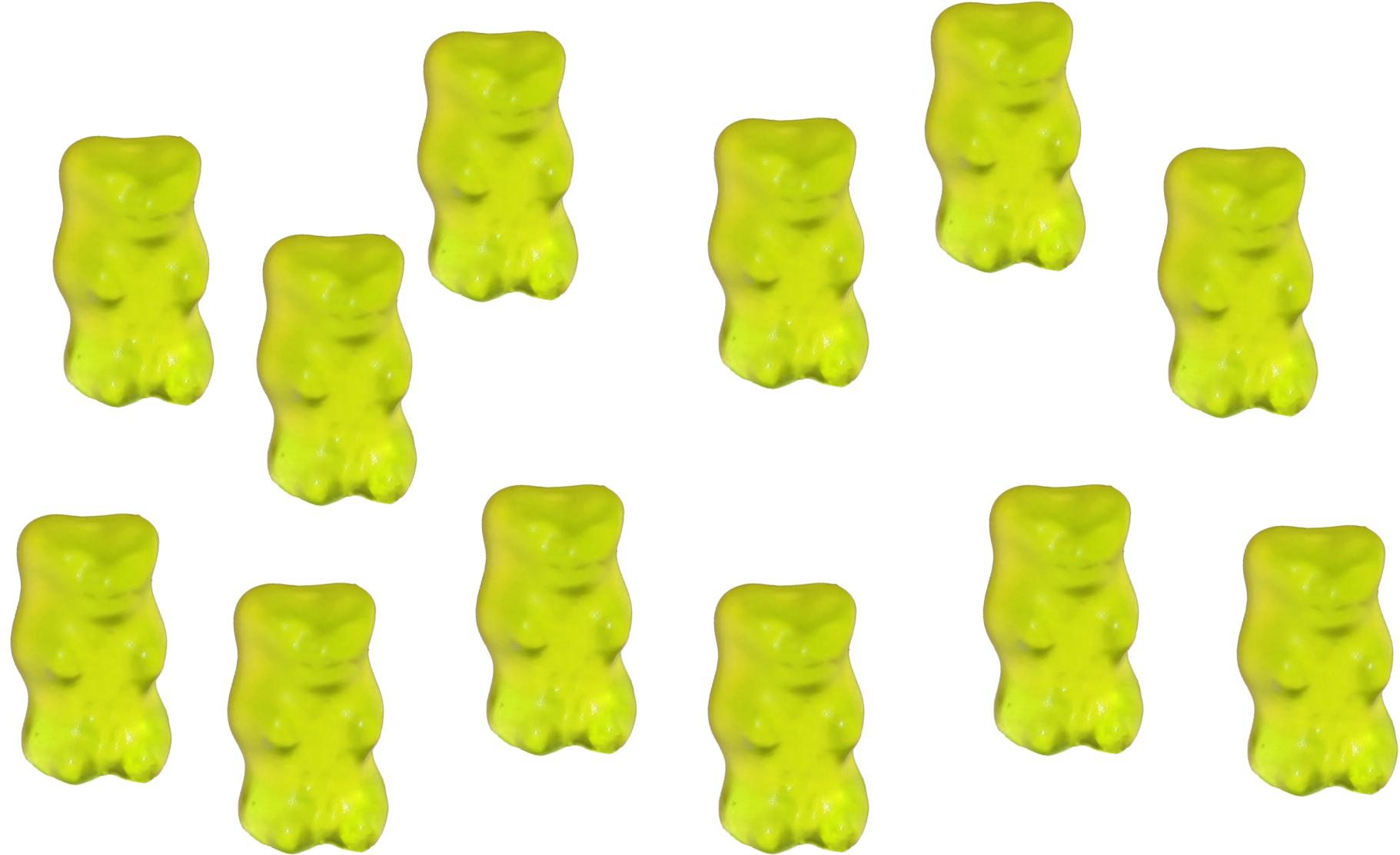
Lots of content  
to analyse

Test network  
models

Evaluate  
prediction  
algorithms

Test search and  
recommender  
systems

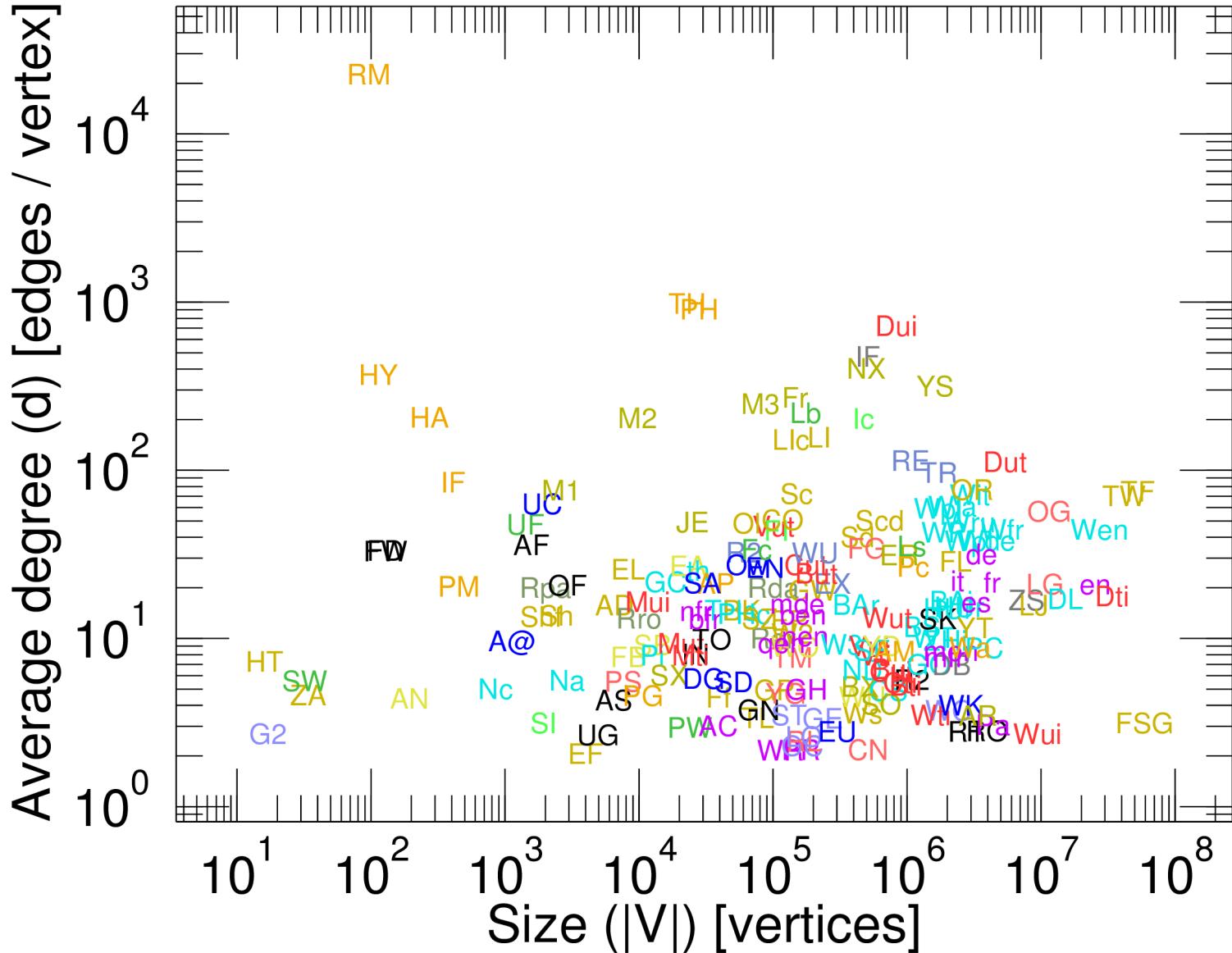
When You Have Tested One, You Have  
Tested All ?!



# Or Do You?



# Diversity of Web Network Datasets



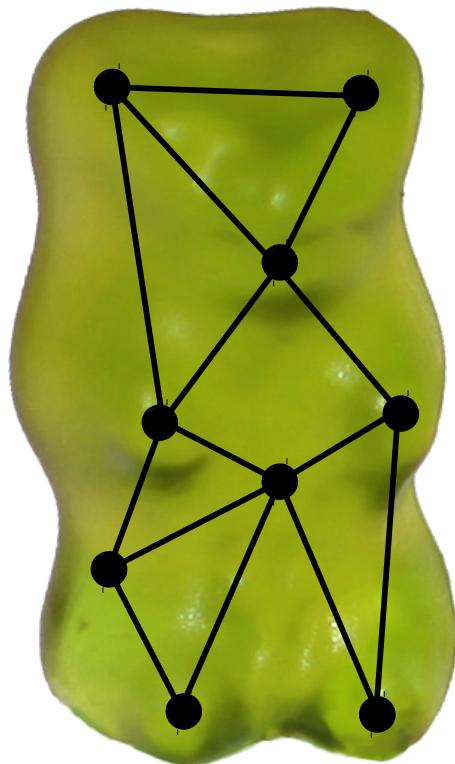
# Overview

Category	Vertices	Edges	Properties	Count
Affiliation	Actors, groups	Memberships	B – =	8
Authorship	Authors, works	Authorships	B – =	18
Co-occurrence	Items	Co-occurrences	U D –	2
Communication	Persons	Messages	U D – =	8
Contact	Persons	Interactions	U D =	4
Features	Items, features	Properties	B –	5
Folksonomy	Users, tags, items	Tag assignments	B =	17
Interaction	Persons, items	Interactions	B – =	14
Lexical	Words	Lexical relationships	U D B – =	5
Physical	Various	Physical connections	U D – =	13
Ratings	Users, items	Ratings	B – * * *	11
Reference	Documents	References	D – =	28
Semantic	Entities	Relationships	D =	1
Social	Persons	Ties	U D – = + ± *	29
Text	Documents, words	Occurrences	B =	5

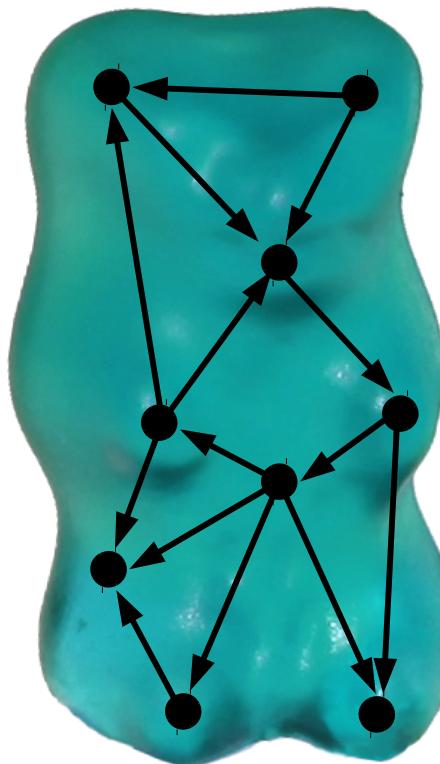
+ 18 new networks

# Network Formats

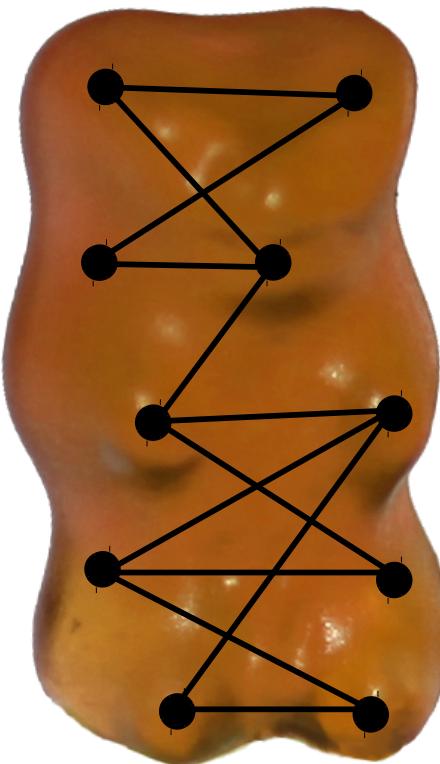
U • Undirected



D • Directed



B • Bipartite

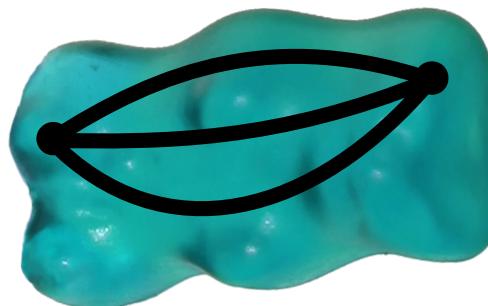


# Edge Weight and Multiplicity Types

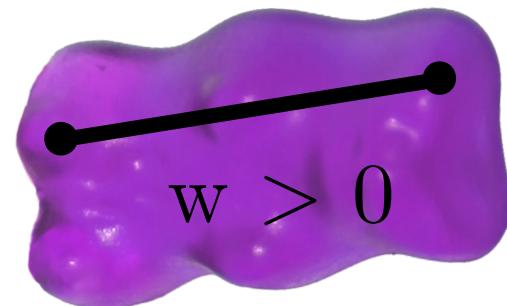
- • Unweighted



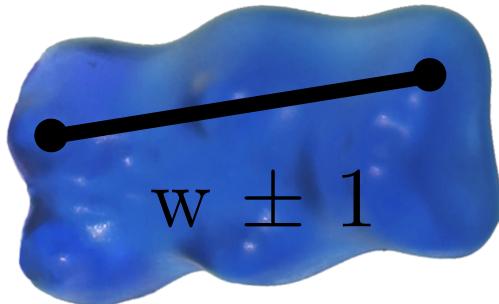
= • Multiple



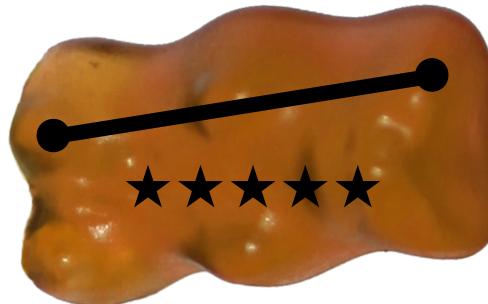
+ • Positive



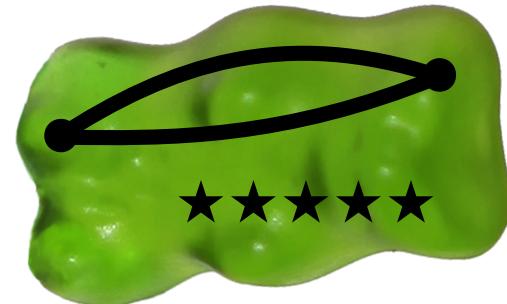
± • Signed



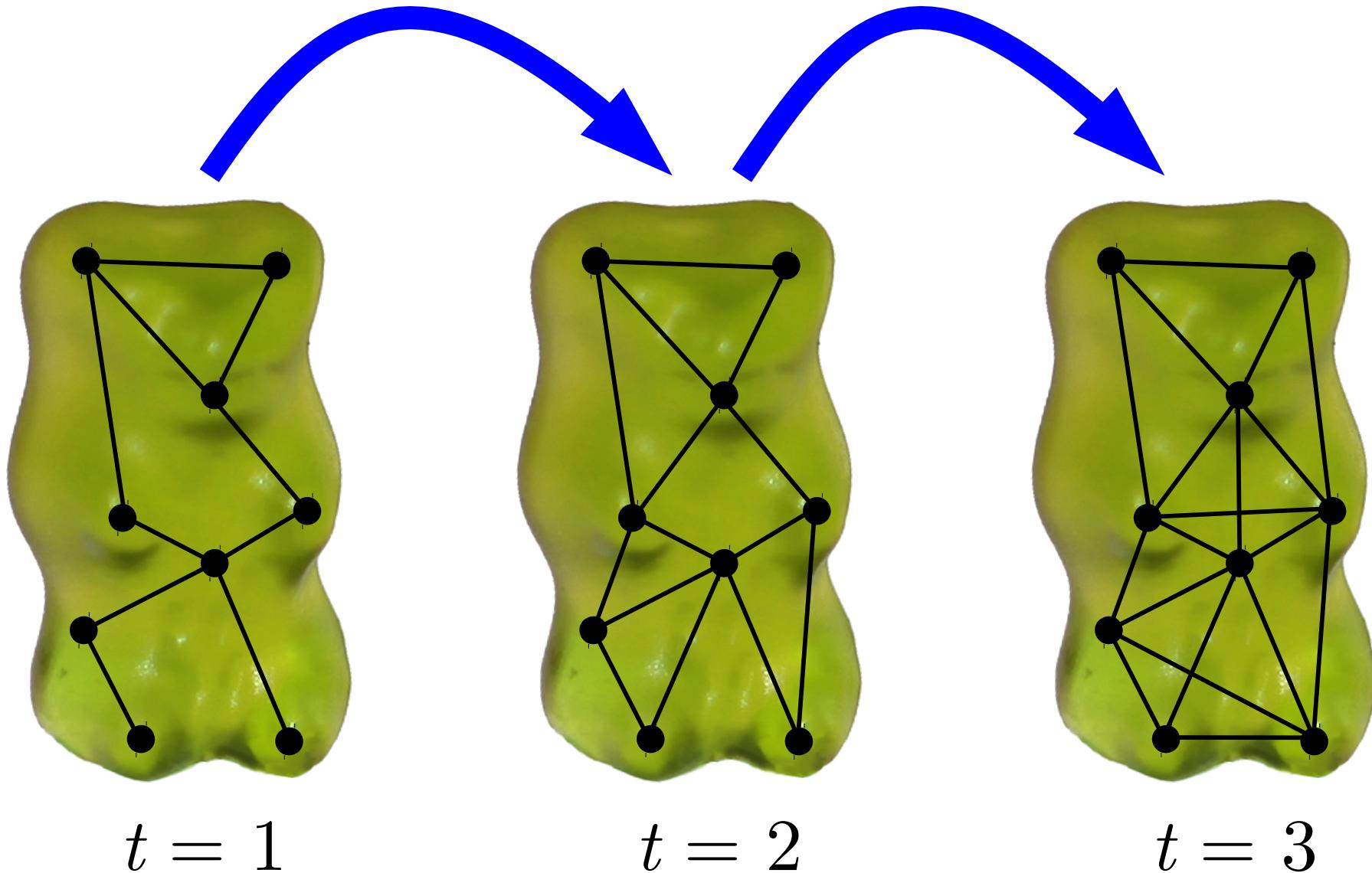
\* • Rating



\* • Multiple Rating



# Timestamps



# Example Dataset

KONECT > Networks > **Slashdot Zoo**

## Slashdot Zoo

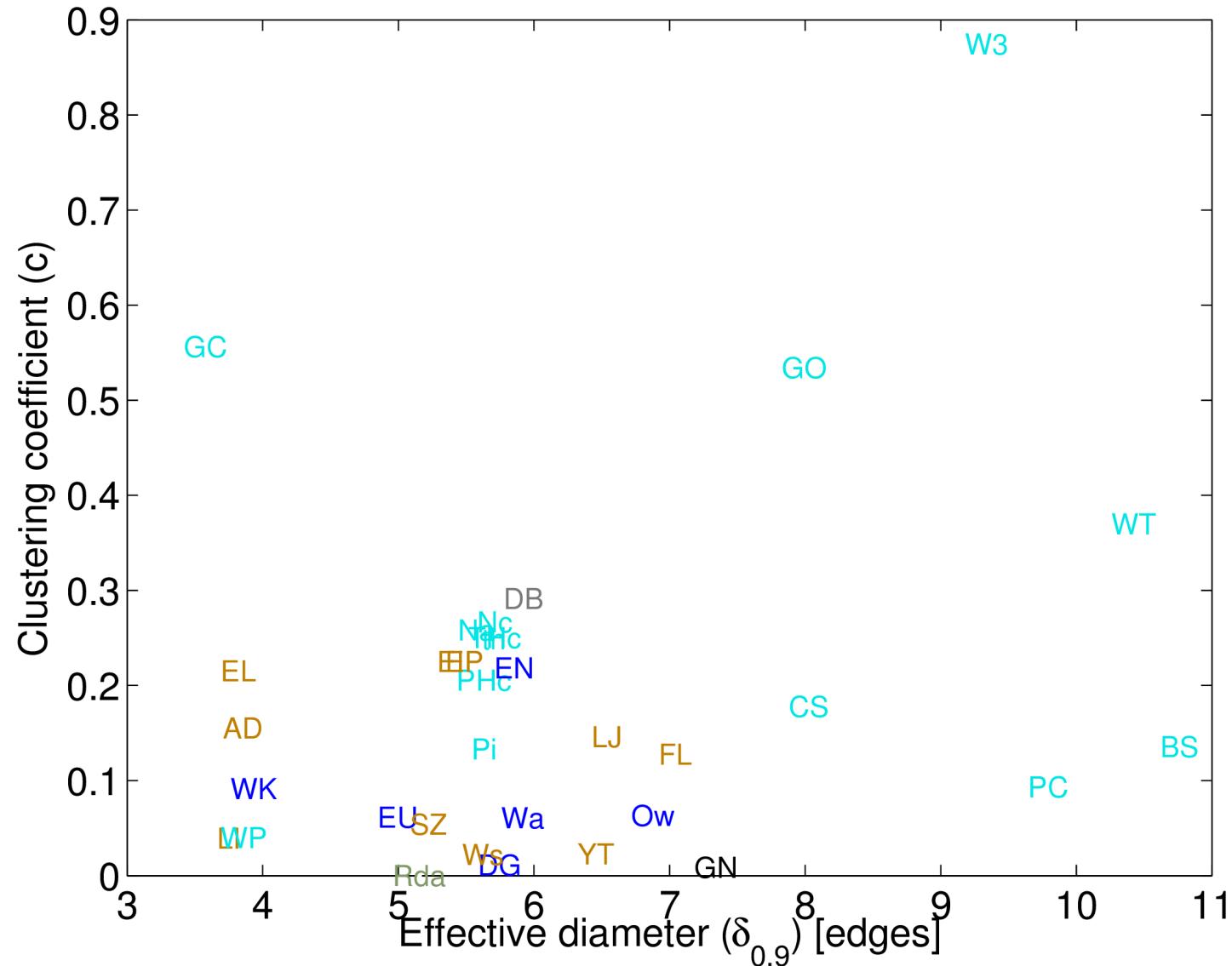
### About this network

This is the signed social network of users of the technology news site Slashdot (slashdot.org), connected by directed "friend" and "foe" relations. The "friend" and "foe" labels are used on Slashdot to mark users, and influence the scores as seen by each user. For instance, If user A marks user B as a foe, the score of user B's posts will be decreased as shown to user A.

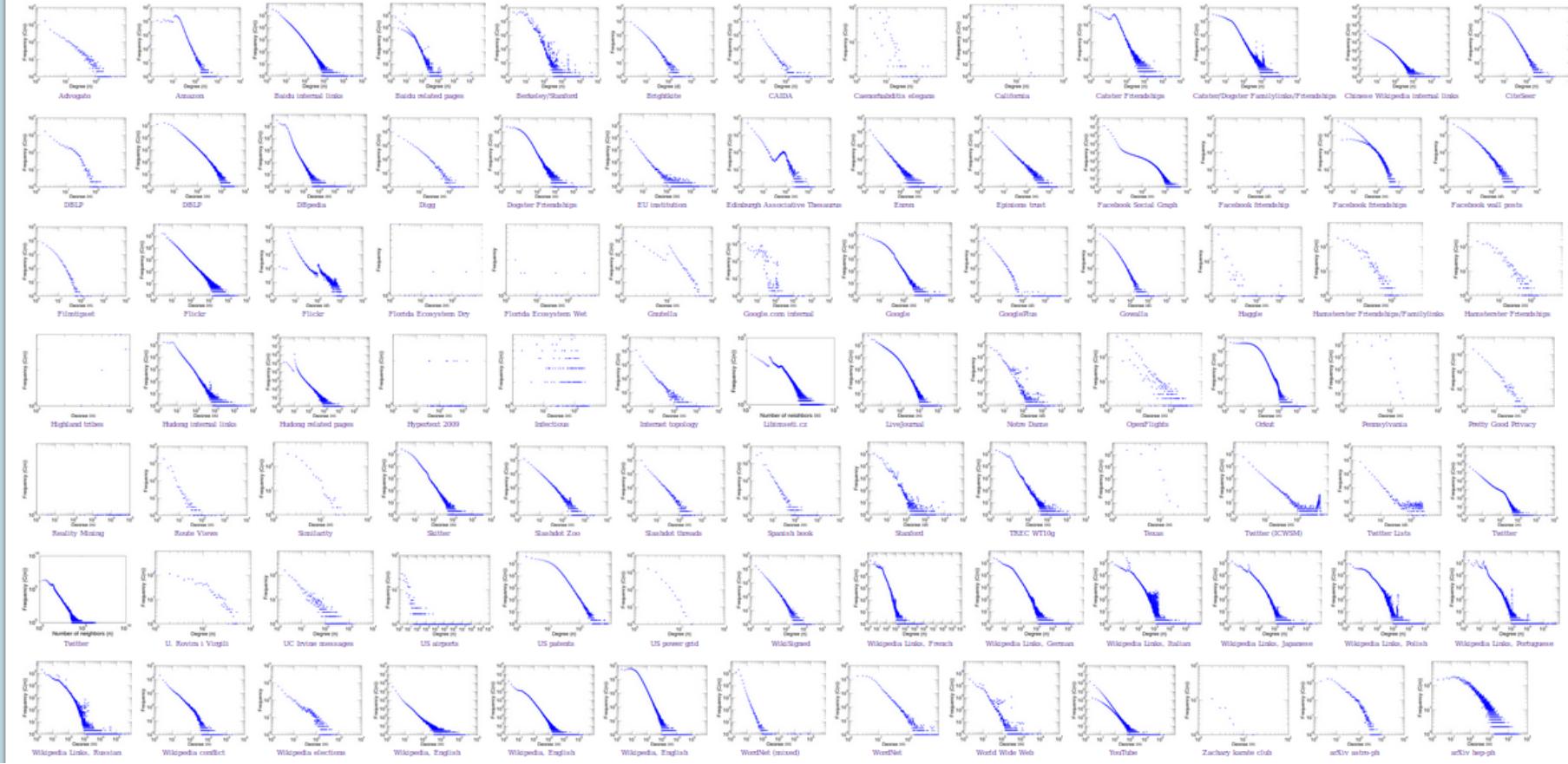
### Network info

<b>Code</b>	<b>sz</b>
<b>Category</b>	Social
<b>Data source</b>	<a href="http://dai-labor.de/IRML/datasets">http://dai-labor.de/IRML/datasets</a>
<b>Description</b>	User-user friend/foe
<b>Format</b>	Directed
<b>Edge weights</b>	Signed
<b>Size</b>	79,120 vertices
<b>Volume</b>	515,581 edges
<b>Average degree</b>	13.03 edges / vertex
<b>Maximum degree</b>	2,546 edges
<b>Fill</b>	$1.733523 \times 10^{-4}$ edges / vertex <sup>2</sup>
<b>Largest connected component</b>	79,120 vertices
<b>Largest strongly connected component</b>	26,997 vertices
<b>Clustering coefficient</b>	3.18%
<b>Algebraic connectivity</b>	$9.516676 \times 10^{-3}$
<b>Spectral norm</b>	52.24
<b>Effective diameter</b>	5.22 edges (estimated)
<b>Gini coefficient</b>	77.43%

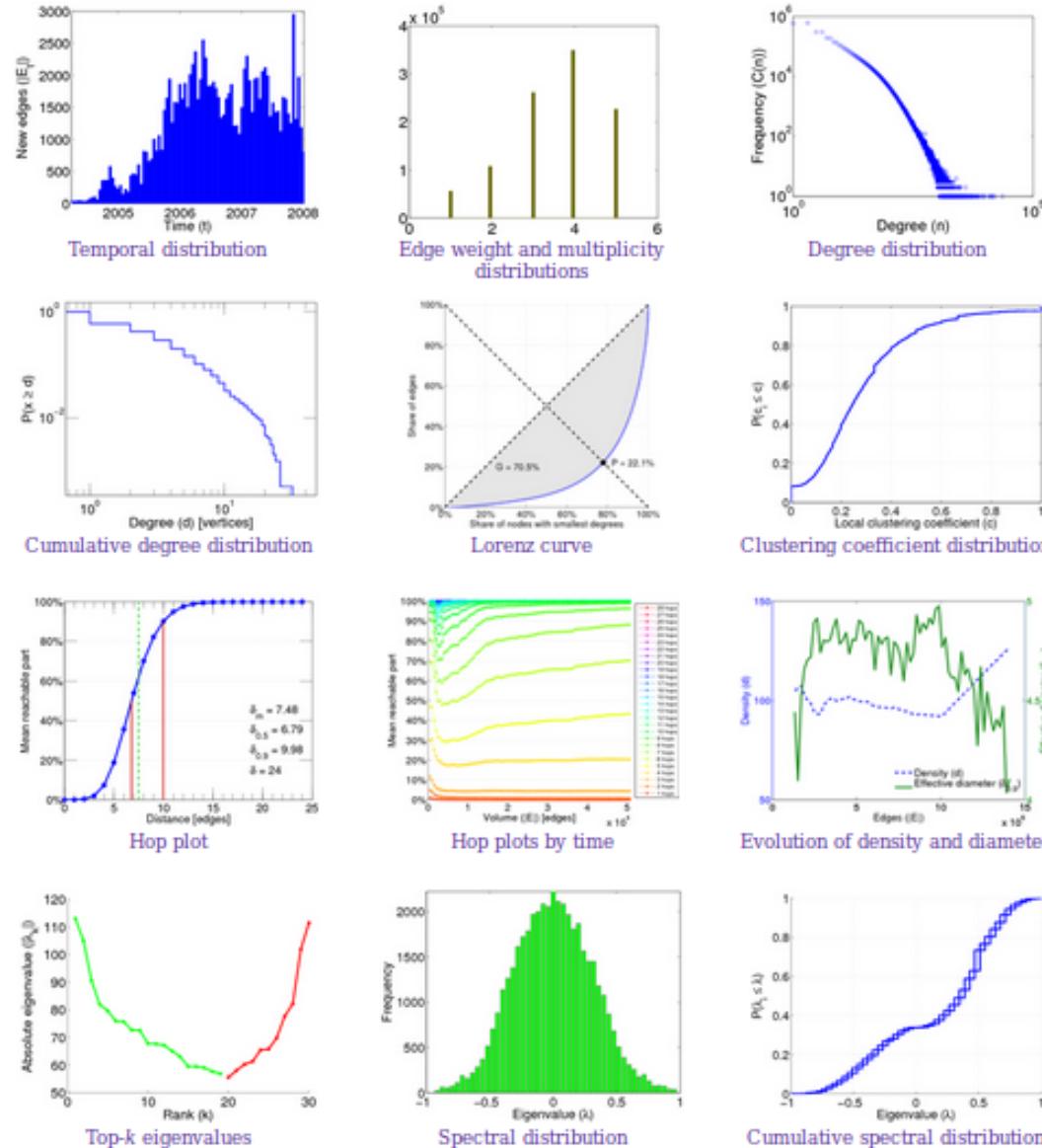
# Network Comparison



# Network Comparison: Plots



# More Plots



# Download

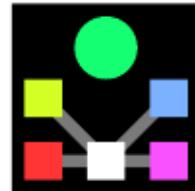
Code	Name	Category	E	W	M	Size	Volume	Avg. degree	Download
Ds	discogs_1style	Features	B	—		244,147	5,235,950	21.56	
AN	Adjective-noun relationships	Lexical	B	—		194	425	3.34	
AD	Advogato	Social	D	+		6,531	51,332	7.84	
AM	Amazon	Contact	U	—		605,731	3,387,366	4.20	
AR	Amazon ratings	Ratings	B	+	⊗	3,376,972	5,638,041	2.72	
AP	arXiv astro-ph	Contact	U	—		37,544	396,160	10.35	
AC	arXiv cond-mat	Authorship	B	—		38,741	58,985	3.30	
PH	arXiv hep-ph	Contact	U	—		26,093	12,730,096	453.14	
PHc	arXiv hep-ph	Reference	D	—		60,386	421,376	6.98	
THc	arXiv hep-th	Reference	D	—		48,239	352,807	7.31	
TH	arXiv hep-th	Contact	U	—		22,906	11,209,366	469.32	
th	arXiv hep-th (KDD Cup)	Reference	D	—		27,770	352,807	12.70	
BAI	Baidu	Reference	D	—		2,141,300	17,794,639	8.31	
BAr	Baidu	Reference	D	—		415,641	3,284,387	7.90	
BS	Berkeley/Stanford	Reference	D	—		1,297,580	7,600,595	5.66	
Bti	BibSonomy ti	Polksonomy	B	—	⊗	975,963	2,535,060	12.48	
Bui	BibSonomy ui	Polksonomy	B	—	⊗	777,084	2,535,060	440.99	
But	BibSonomy ut	Polksonomy	B	—	⊗	210,467	2,535,060	440.99	
BK	Brightkite	Social	U	—		58,228	214,078	3.68	
PM	Caenorhabditis elegans	Contact	U	—		433	4,396	10.15	
IN	CAIDA	Physical	U	—		26,473	106,762	4.03	
RO	California	Physical	U	—		3,930,412	5,533,214	1.41	
Sc	Catster	Social	U	—		149,700	5,449,275	36.40	
Scd	Catster/Dogster	Social	U	—		624,127	15,705,337	23.16	
CS	CiteSeer	Reference	D	—		723,131	1,764,929	2.44	
Cti	CiteULike ti	Polksonomy	B	—	⊗	885,046	2,411,619	15.74	
Cul	CiteULike ul	Polksonomy	B	—	⊗	754,484	2,411,619	106.18	
Cut	CiteULike ut	Polksonomy	B	—	⊗	175,992	2,411,619	106.18	
CN	Countries	Affiliation	B	—		512,781	557,367	1.09	
PI	DBLP	Reference	D	—		12,591	49,793	3.95	

```
@prefix void: <http://rdfs.org/ns/void#> .  
@prefix konect_networks: <http://konect.uni-koblenz.de/networks/> .  
@prefix dcterms: <http://purl.org/dc/terms/> .  
@prefix foaf: <http://xmlns.com/foaf/0.1/> .  
@prefix konect_udata_zachary: <http://konect.uni-koblenz.de/networks/ucidata-zachary/Member/> .  
konect_networks:ucidata-zachary a void:Dataset .  
konect_networks:ucidata-zachary dcterms:title "Zachary karate club" .  
konect_networks:ucidata-zachary dcterms:description "This is the well-known Zachary karate club network. The  
data was collected from the members of a university karate club by Wayne Zachary in 1977. Each node represents a  
member of the club, and each edge represents a tie between two members of the club. This is a classical social  
network dataset from the literature. The network is very small: it has 34 vertices and 78 undirected edges." .  
konect_networks:ucidata-zachary dcterms:source <http://vlado.fmf.uni-lj.si/pub/networks/data/ucinet/  
ucidata.htm#zachary> .  
konect_udata_zachary:i1 a foaf:Person .  
konect_udata_zachary:i2 a foaf:Person .  
konect_udata_zachary:i1 foaf:knows konect_udata_zachary:i2 .  
konect_udata_zachary:i2 foaf:knows konect_udata_zachary:i1 .  
konect_udata_zachary:i3 a foaf:Person .  
konect_udata_zachary:i1 foaf:knows konect_udata_zachary:i3 .  
konect_udata_zachary:i3 foaf:knows konect_udata_zachary:i1 .  
konect_udata_zachary:i2 foaf:knows konect_udata_zachary:i3 .  
konect_udata_zachary:i3 foaf:knows konect_udata_zachary:i2 .  
konect_udata_zachary:i4 a foaf:Person .  
konect_udata_zachary:i1 foaf:knows konect_udata_zachary:i4 .  
konect_udata_zachary:i4 foaf:knows konect_udata_zachary:i1 .
```

# Matlab Toolbox

```
konect_dentropy.m
konect_diameff.m
konect_diammean.m
konect_effective_diameter.m
konect_eigl.m
konect_eign.m
konect_eigskew.m
konect_first_index.m
konect_fromto.m
konect_gini_direct.m
konect_gini.m
konect_hopdistr_ex.m
konect_hopdistr.m
konect_imageubu_complex.m
konect_imageubu.m
konect_jain.m
konect_join.m
konect_label_statistic.m
konect_map.m
konect_matrix.m
konect_mauc.m
konect_network_rank_abs.m
konect_normalize_additively.m
konect_normalized_entropy.m
konect_normalize_matrix.m
konect_normalize_rows.m
konect_order_dedicom.m
konect_own.m
```

# Handbook of Network Analysis



## Handbook of Network Analysis KONECT – the Koblenz Network Collection

Jérôme Kunegis

May 10, 2013

### 1 Introduction

Everything is a network. Whenever we look at the interactions between things, a network is formed implicitly. In the areas of data mining, machine learning, information retrieval, etc., networks are modeled as *graphs*. Many, if not most problem types can be applied to graphs: clustering, classification, prediction, pattern recognition, and others. Networks arise in almost all areas of research, commerce and daily life in the form of social networks, road networks, communication networks, trust networks, hyperlink networks, chemical interaction networks, neural networks, collaboration networks and lexical networks. The content of text documents is routinely modeled as a network where words are nodes and taste as

<http://konect.uni-koblenz.de/publications>

different properties, e.g., word frequency or co-occurrence. In order to provide a unified view on such network datasets, and to allow the application of network analysis methods across disciplines, the KONECT project defines a comprehensive network taxonomy and provides a consistent access to network datasets. To validate this approach on real-world data from the Web, KONECT also provides a large number (160+) of network datasets of

<http://konect.uni-koblenz.de/>

@KONECTproject



Give Me  
Datasets!