By *Jure Leskovec*

STANFORD UNIVERSITY

SNAP

SNAP for C++
SNAP for Python
SNAP Datasets
BIOSNAP Datasets
What's new
People
Papers
Projects
Citing SNAP
Links
About
Contact us

Open positions

Open research positions in **SNAP** group are available at **undergraduate, graduate** and **postdoctoral** levels.

Stanford Large Network Dataset Collection

- **Social networks** : online social networks, edges represent interactions between people
- **Networks with ground-truth communities** : ground-truth network communities in social and information networks
- **Communication networks** : email communication networks with edges representing communication
- **Citation networks** : nodes represent papers, edges represent citations
- **Collaboration networks** : nodes represent scientists, edges represent collaborations (co-authoring a paper)
- **Web graphs** : nodes represent webpages and edges are hyperlinks
- **Amazon networks** : nodes represent products and edges link commonly co-purchased products
- **Internet networks** : nodes represent computers and edges communication
- **Road networks** : nodes represent intersections and edges roads connecting the intersections
- **Autonomous systems** : graphs of the internet
- **Signed networks** : networks with positive and negative edges (friend/foe, trust/distrust)
- **Location-based online social networks** : social networks with geographic check-ins
- **Wikipedia networks, articles, and metadata** : talk, editing, voting, and article data from Wikipedia
- **Temporal networks** : networks where edges have timestamps
- **Twitter and Memetracker** : memetracker phrases, links and 467 million Tweets
- **Online communities** : data from online communities such as Reddit and Flickr
- **Online reviews** : data from online review systems such as BeerAdvocate and Amazon
- **User actions** : actions of users on social platforms.
- **Face-to-face communication networks** : networks of face-to-face (non-online) interactions
- **Graph classification datasets** : disjoint graphs from different classes

SNAP networks are also available from SuiteSparse Matrix Collection by Tim Davis.

Social networks

Name	Type	Nodes	Edges	Description
ego-Facebook	Undirected	4,039	88,234	Social circles from Facebook (anonymized)
ego-Gplus	Directed	107,614	13,673,453	Social circles from Google+
ego-Twitter	Directed	81,306	1,768,149	Social circles from Twitter
soc-Epinions1	Directed	75,879	508,837	Who-trusts-whom network of Epinions.com
soc-LiveJournal1	Directed	4,847,571	68,993,773	LiveJournal online social network
soc-Pokec	Directed	1,632,803	30,622,564	Pokec online social network
soc-Slashdot0811	Directed	77,360	905,468	Slashdot social network from November 2008
soc-Slashdot0922	Directed	82,168	948,464	Slashdot social network from February 2009
wiki-Vote	Directed	7,115	103,689	Wikipedia who-votes-on-whom network
wiki-RfA	Directed, Signed	10,835	159,388	Wikipedia Requests for Adminship (with text)
gemsec-Deezer	Undirected	143,884	846,915	Gemsec Deezer dataset
gemsec-Facebook	Undirected	134,833	1,380,293	Gemsec Facebook dataset
soc-RedditHyperlinks	Directed, Signed, Temporal, Attributed	55,863	858,490	Hyperlinks between subreddits on Reddit
soc-sign-bitcoin-otc	Weighted, Signed, Directed, Temporal	5,881	35,592	Bitcoin OTC web of trust network
soc-sign-bitcoin-alpha	Weighted, Signed, Directed, Temporal	3,783	24,186	Bitcoin Alpha web of trust network
comm-f2f-Resistance	Weighted, Directed, Temporal	451	3,126,993	Dynamic face-to-face interaction network between group of people
musae-twitch	Undirected	34,118	429,113	Social networks of Twitch users.
musae-facebook	Undirected	22,470	171,002	Facebook page-page network with page names.
act-mooc	Bipartite, Directed, Attributed, Temporal	7,143	411,749	Student actions on a MOOC platform, with student drop-out binary labels.
musae-github	Undirected	37,700	289,003	Social network of Github developers.
feather-deezer-social	Undirected	28,281	92,752	Social network of Deezer users from Europe.
feather-lastfm-social	Undirected	7,624	27,806	Social network of LastFM users from Asia.
twitch-gamers	Undirected	168,114	6,797,557	Social network of Twitch users.

Networks with ground-truth communities

Name	Type	Nodes	Edges	Communities	Description
com-LiveJournal	Undirected, Communities	3,997,962	34,681,189	287,512	LiveJournal online social network
com-Friendster	Undirected, Communities	65,608,366	1,806,067,135	957,154	Friendster online social network
com-Orkut	Undirected, Communities	3,072,441	117,185,083	6,288,363	Orkut online social network
com-Youtube	Undirected, Communities	1,134,890	2,987,624	8,385	Youtube online social network
com-DBLP	Undirected, Communities	317,080	1,049,866	13,477	DBLP collaboration network
com-Amazon	Undirected, Communities	334,863	925,872	75,149	Amazon product network
email-Eu-core	Directed, Communities	1,005	25,571	42	E-mail network
wiki-topcats	Directed, Communities	1,791,489	28,511,807	17,364	Wikipedia hyperlinks

Communication networks

Name	Type	Nodes	Edges	Description
email-EuAll	Directed	265,214	420,045	Email network from a EU research institution
email-Enron	Undirected	36,692	183,831	Email communication network from Enron
wiki-Talk	Directed	2,394,385	5,021,410	Wikipedia talk (communication) network
comm-f2f-Resistance	Weighted, Directed, Temporal	451	3,126,993	Dynamic face-to-face interaction network between group of people

Citation networks

Name	Type	Nodes	Edges	Description
cit-HepPh	Directed, Temporal, Labeled	34,546	421,578	Arxiv High Energy Physics paper citation network
cit-HepTh	Directed, Temporal, Labeled	27,770	352,807	Arxiv High Energy Physics paper citation network
cit-Patents	Directed, Temporal, Labeled	3,774,768	16,518,948	Citation network among US Patents

Collaboration networks

Name	Type	Nodes	Edges	Description
ca-AstroPh	Undirected	18,772	198,110	Collaboration network of Arxiv Astro Physics
ca-CondMat	Undirected	23,133	93,497	Collaboration network of Arxiv Condensed Matter
ca-GrQc	Undirected	5,242	14,496	Collaboration network of Arxiv General Relativity
ca-HepPh	Undirected	12,008	118,521	Collaboration network of Arxiv High Energy Physics
ca-HepTh	Undirected	9,877	25,998	Collaboration network of Arxiv High Energy Physics Theory

Web graphs

Name	Type	Nodes	Edges	Description
web-BerkStan	Directed	685,230	7,600,595	Web graph of Berkeley and Stanford
web-Google	Directed	875,713	5,105,039	Web graph from Google
web-NotreDame	Directed	325,729	1,497,134	Web graph of Notre Dame
web-Stanford	Directed	281,903	2,312,497	Web graph of Stanford.edu

Product co-purchasing networks

Name	Type	Nodes	Edges	Description
amazon0302	Directed	262,111	1,234,877	Amazon product co-purchasing network from March 2 2003
amazon0312	Directed	400,727	3,200,440	Amazon product co-purchasing network from March 12 2003
amazon0505	Directed	410,236	3,356,824	Amazon product co-purchasing network from May 5 2003
amazon0601	Directed	403,394	3,387,388	Amazon product co-purchasing network from June 1 2003
amazon-meta	Metadata	548,552	1,788,725	Amazon product metadata: product info and all reviews on around 548,552 products.

Internet peer-to-peer networks

Name	Type	Nodes	Edges	Description
p2p-Gnutella04	Directed	10,876	39,994	Gnutella peer to peer network from August 4 2002
p2p-Gnutella05	Directed	8,846	31,839	Gnutella peer to peer network from August 5 2002

p2p-Gnutella06	Directed	8,717	31,525	Gnutella peer to peer network from August 6 2002
p2p-Gnutella08	Directed	6,301	20,777	Gnutella peer to peer network from August 8 2002
p2p-Gnutella09	Directed	8,114	26,013	Gnutella peer to peer network from August 9 2002
p2p-Gnutella24	Directed	26,518	65,369	Gnutella peer to peer network from August 24 2002
p2p-Gnutella25	Directed	22,687	54,705	Gnutella peer to peer network from August 25 2002
p2p-Gnutella30	Directed	36,682	88,328	Gnutella peer to peer network from August 30 2002
p2p-Gnutella31	Directed	62,586	147,892	Gnutella peer to peer network from August 31 2002

Road networks

Name	Type	Nodes	Edges	Description
roadNet-CA	Undirected	1,965,206	2,766,607	Road network of California
roadNet-PA	Undirected	1,088,092	1,541,898	Road network of Pennsylvania
roadNet-TX	Undirected	1,379,917	1,921,660	Road network of Texas

Autonomous systems graphs

Name	Type	Nodes	Edges	Description
as-733 (733 graphs)	Undirected	103-6,474	243-13,233	733 daily instances(graphs) from November 8 1997 to January 2 2000
as-Skitter	Undirected	1,696,415	11,095,298	Internet topology graph, from traceroutes run daily in 2005
as-Caida (122 graphs)	Directed	8,020-26,475	36,406-106,762	The CAIDA AS Relationships Datasets, from January 2004 to November 2007
Oregon-1 (9 graphs)	Undirected	10,670-11,174	22,002-23,409	AS peering information inferred from Oregon route-views between March 31 and May 26 2001
Oregon-2 (9 graphs)	Undirected	10,900-11,461	31,180-32,730	AS peering information inferred from Oregon route-views between March 31 and May 26 2001

Signed networks

Name	Type	Nodes	Edges	Description
soc-RedditHyperlinks	Directed, Signed, Temporal, Attributed	55,863	858,490	Hyperlinks between subreddits on Reddit
soc-sign-bitcoin-otc	Weighted, Signed, Directed, Temporal	5,881	35,592	Bitcoin OTC web of trust network
soc-sign-bitcoin-alpha	Weighted, Signed, Directed, Temporal	3,783	24,186	Bitcoin Alpha web of trust network
soc-sign-epinions	Directed	131,828	841,372	Epinions signed social network
wiki-Elec	Directed, Bipartite	~7,000	~100,000	Wikipedia adminship election data
wiki-RfA	Directed, Signed	10,835	159,388	Wikipedia Requests for Adminship (with text)
soc-sign-Slashdot081106	Directed	77,357	516,575	Slashdot Zoo signed social network from November 6 2008
soc-sign-Slashdot090216	Directed	81,871	545,671	Slashdot Zoo signed social network from February 16 2009
soc-sign-Slashdot090221	Directed	82,144	549,202	Slashdot Zoo signed social network from February 21 2009

Location-based online social networks

Name	Type	Nodes	Edges	Description
loc-Gowalla	Undirected, Geo-Location	196,591	950,327	Gowalla location based online social network
loc-Brightkite	Unirected, Geo-Location	58,228	214,078	Brightkite location based online social network

Wikipedia networks, articles, and metadata

Name	Type	Nodes	Edges	Description
wikispeedia	Navigation paths	4,604	119,882	Navigation paths on the Wikipedia hyperlink network, collected by the human-computation game *Wikispeedia*
wiki-Vote	Directed	7,115	103,689	Wikipedia who-votes-on-whom network
wiki-Talk	Directed	2,394,385	5,021,410	Wikipedia talk (communication) network
wiki-Elec	Bipartite	~7,000	~100,000	Wikipedia adminship election data
wiki-RfA	Directed, Signed	10,835	159,388	Wikipedia Requests for Adminship (with text)
wiki-meta	Edits	2.3M users, 3.5M pages	250M edits	Complete Wikipedia edit history (who edited what page)

| wiki-hoaxes | Wiki markup | 64 | | Public Wikipedia hoaxes |
| musae-wiki | undirected | 19,109 | 400,832 | Wikipedia page page network with traffic information. |

Temporal networks

Name	Type	Nodes	Temporal Edges	Static Edges	Description
soc-RedditHyperlinks	Directed, Signed, Temporal, Attributed	55,863	858,490	858,490	Hyperlinks between subreddits on Reddit
sx-stackoverflow	Directed, Temporal	2,601,977	63,497,050	36,233,450	Comments, questions, and answers on Stack Overflow
sx-mathoverflow	Directed, Temporal	24,818	506,550	239,978	Comments, questions, and answers on Math Overflow
sx-superuser	Directed, Temporal	194,085	1,443,339	924,886	Comments, questions, and answers on Super User
sx-askubuntu	Directed, Temporal	159,316	964,437	596,933	Comments, questions, and answers on Ask Ubuntu
wiki-talk-temporal	Directed, Temporal	1,140,149	7,833,140	3,309,592	Users editing talk pages on Wikipedia
email-Eu-core-temporal	Directed, Temporal	986	332,334	24,929	E-mails between users at a research institution
CollegeMsg	Directed, Temporal	1,899	20,296	59,835	Messages on a Facebook-like platform at UC-Irvine
soc-sign-bitcoin-otc	Weighted, Signed, Directed, Temporal	5,881	35,592		Bitcoin OTC web of trust network
soc-sign-bitcoin-alpha	Weighted, Signed, Directed, Temporal	3,783	24,186		Bitcoin Alpha web of trust network
act-mooc	Bipartite, Directed, Attributed, Temporal	7,143	411,749		Student actions on a MOOC platform, with student drop-out binary labels.
comm-f2f-Resistance	Weighted, Directed, Temporal	451	3,126,993		Dynamic face-to-face interaction network between group of people

User Actions

Name	Type	Nodes	Edges/Actions	Description
act-mooc	Bipartite, Directed, Attributed, Temporal	7,143	411,749	Student actions on a MOOC platform, with student drop-out binary labels.

Memetracker and Twitter

Name	Type	Nodes	Edges	Description
twitter7	Tweets	17,069,982 users	476,553,560 tweets	A collection of 476 million tweets collected between June-Dec 2009
memetracker9	Memes	96 million	418 million links	Memetracker phrases and hyperlinks between 96 million blog posts from Aug 2008 to Apr 2009
ksc-time-series	Time Series	2,000		Time series of volume of 1,000 most popular Memetracker phrases and 1,000 most popular Twitter hashtags
higgs-twitter	Tweets	456,631	14,855,875	Spreading processes of the announcement of the discovery of a new particle with the features of the Higgs boson on 4th July 2012.

Online Communities

Name	Type	Number of items	Description
soc-RedditHyperlinks	Subreddit hyperlinks	858,490 links between 55,863 subreddits	Hyperlinks between subreddits on Reddit
web-RedditEmbeddings	Reddit Embeddings	118,381 users and 51,278 subreddits	Embeddings of users and subreddits
web-RedditPizzaRequests	Reddit requests	5,671 submissions	Textual requests for pizza with outcome labels
web-Reddit	Reddit submissions	132,308 submissions	Resubmitted content on reddit.com
web-flickr	Images	2,316,948 related images	Images sharing common metadata on Flickr

Online Reviews

Name	Type	Number of items	Description
web-BeerAdvocate	Beer reviews	1,586,259 beer reviews	Beer reviews from BeerAdvocate
web-RateBeer	Beer reviews	2,924,127 beer reviews	Beer reviews from RateBeer

web-CellarTracker	Wine reviews	2,025,995 wine reviews	Wine reviews from CellarTracker
web-Amazon	Amazon reviews (all categories)	34,686,770 product reviews	Reviews from Amazon
web-FineFoods	Food reviews	568,454 food reviews	Food reviews from Amazon
web-Movies	Movie reviews	7,911,684 movie reviews	Movie reviews from Amazon

Face-to-Face Communication Networks

Name	Type	Nodes	Edges	Description
comm-f2f-Resistance	Weighted, Directed, Temporal	451	3,126,993	Dynamic face-to-face interaction network between group of people

Graph classification datasets

Name	Type	Number of graphs	Description
Deezer Ego-nets	Undirected, Unattributed	9,629	Ego-nets of European Deezer users.
GitHub Stargazers	Undirected, Unattributed	12,725	Communities of developers who starred repositories.
Reddit Threads	Undirected, Unattributed	203,088	Reddit discussion and non-discussion based threads.
Ego-Nets	Undirected, Unattributed	127,094	Ego-Nets of Twitch users in the partnership program.

Network types

- **Directed** : directed network
- **Undirected** : undirected network
- **Bipartite** : bipartite network
- **Multigraph** : network has multiple edges between a pair of nodes
- **Temporal** : for each node/edge we know the time when it appeared in the network
- **Labeled** : network contains labels (weights, attributes) on nodes and/or edges

Network statistics

Dataset statistics

Nodes	Number of nodes in the network
Edges	Number of edges in the network
Nodes in largest WCC	Number of nodes in the largest weakly connected component
Edges in largest WCC	Number of edges in the largest weakly connected component
Nodes in largest SCC	Number of nodes in the largest strongly connected component
Edges in largest SCC	Number of edges in the largest strongly connected component
Average clustering coefficient	Average clustering coefficient
Number of triangles	Number of triples of connected nodes (considering the network as undirected)
Fraction of closed triangles	Number of connected triples of nodes / number of (undirected) length 2 paths
Diameter (longest shortest path)	Maximum undirected shortest path length (sampled over 1,000 random nodes)
90-percentile effective diameter	90-th percentile of undirected shortest path length distribution (sampled over 1,000 random nodes)

Citing SNAP

We encourage you to cite our datasets if you have used them in your work. You can use the following BibTeX citation:

```
@misc{snapnets,
  author       = {Jure Leskovec and Andrej Krevl},
  title        = {{SNAP Datasets}: {Stanford} Large Network Dataset Collection},
  howpublished = {\url{http://snap.stanford.edu/data}},
  month        = jun,
  year         = 2014
}
```