Linked Open Social Data for Scientific Benchmarking (Supporting Information document)

Renato Fabbri[a,1,*], Osvaldo Novais de Oliveira Junior[a,1]

[a] *São Carlos Institute of Physics, São Paulo University, Brazil*

Abstract

This is a Supporting Information document which exposes ontological diagrams and auxiliary tables for the Linked Open Social Data (LOSD) database. The main document of the article is in [1].

Keywords: Big Data, Data Mining, Benchmark Data, Facebook, Twitter, IRC, Email, Complex Networks, Text Mining

Contents

1 General guidance **2**
2 Facebook data **2**
3 Twitter data **5**
4 IRC data **8**
5 Email data **11**
6 ParticipaBR data **14**
7 Cidade Democrática data **18**
8 AA data **22**
9 Snapshot references **25**
10 Trivial counts in each snapshot **29**

[*] Corresponding author

Email addresses: `fabbri@usp.br` (Renato Fabbri), `chu@ifsc.usp.br` (Osvaldo Novais de Oliveira Junior)

[1] *URL:* http://www.ifsc.usp.br/

1. General guidance

In this document we provide diagrams for the provenances in the LOSD: Facebook, Twitter, IRC, Email, ParticipaBR, Cidade Democrática and AA. Each provenance diagram was broken in two, one presents the relations among main classes (blue nodes) and data types (orange nodes), the other presents metadata on the snapshot. Every class instance is related to the snapshot instance by the triple `class_uri po:snapshot snapshot_uri`. Such triples are omitted for simplicity. Due to the large number of relations, the rendering of diagrams are automatized and displays some overlaps. Even so, the images are useful for grasping what is in current LOSD and for conducting explorations. Edges in the diagrams have:

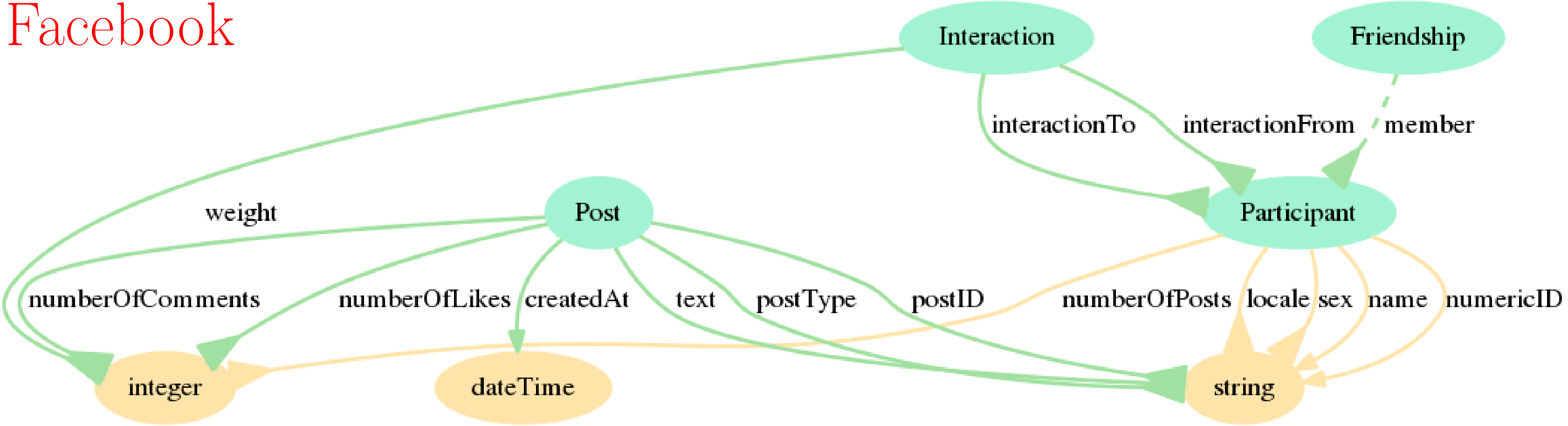
- green color if representing an OWL existential class restriction (all individuals from the class present at least one triple with the property as predicate);
- inverted nip if representing an OWL universal class restriction (all individuals presenting triples with the property as predicate are from the class);
- full edges (non-dashed) if representing a functional property axiom (there is at most one triple with the property as the predicate for each individual).

Furthermore, this document ends with two sets of tables, one with counts of triples, participants, edges/interactions/relations and characters, the other with references for snapshot groups, such as wikipedia or contact links.

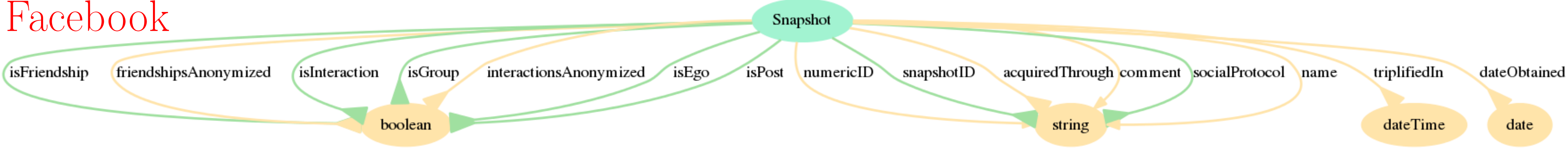
2. Facebook data

Each Facebook snapshot is yield by either an user, from which the friends constitute a friendship network, or a group, which participants can yield friendship and interaction networks and posts information with text and some metadata. Further information is found on the following diagrams, the tables on the end of this document or in the main document of this article [1].

Facebook


Interaction
Friendship
interactionTo
interactionFrom
member
weight
Post
Participant
numberOfComments
numberOfLikes
createdAt
text
postType
postID
numberOfPosts
locale
sex
name
numericID
integer
dateTime
string

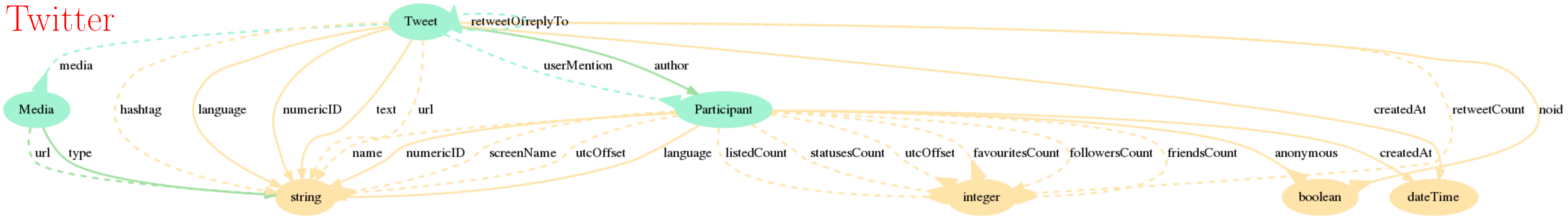
Facebook

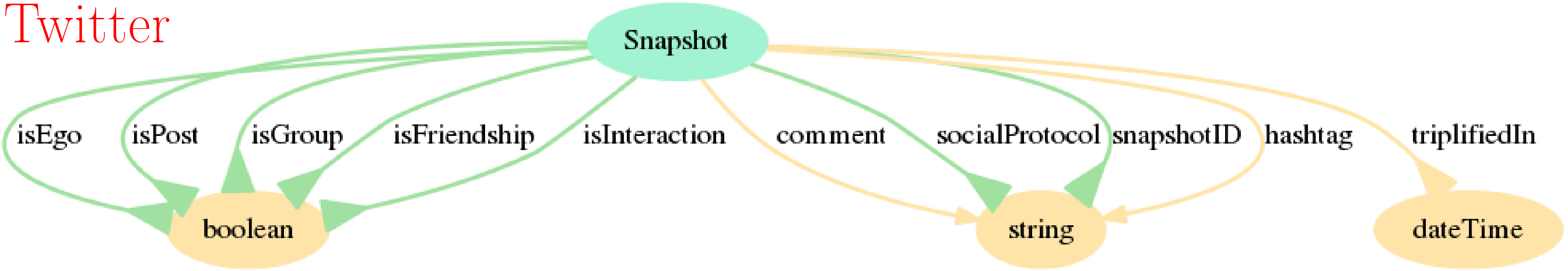

Snapshot
isFriendship
friendshipsAnonymized
isInteraction
isGroup
interactionsAnonymized
isEgo
isPost
numericID
snapshotID
acquiredThrough
comment
socialProtocol
name
triplifiedIn
dateObtained
boolean
string
dateTime
date

3. Twitter data

Each Twitter snapshot is yield by a hashtag. Retweets (`po:retweetOf` are usually considered to yield the interactions between users. Users are identified through authors `po:numericID` (global as given by Twitter API) or . The database present also `po:replyTo` and `po:userMention` which might also be useful in understanding the networking. Further information is found on the following diagrams, the tables on the end of this document or in the main document of this article [1].

Twitter

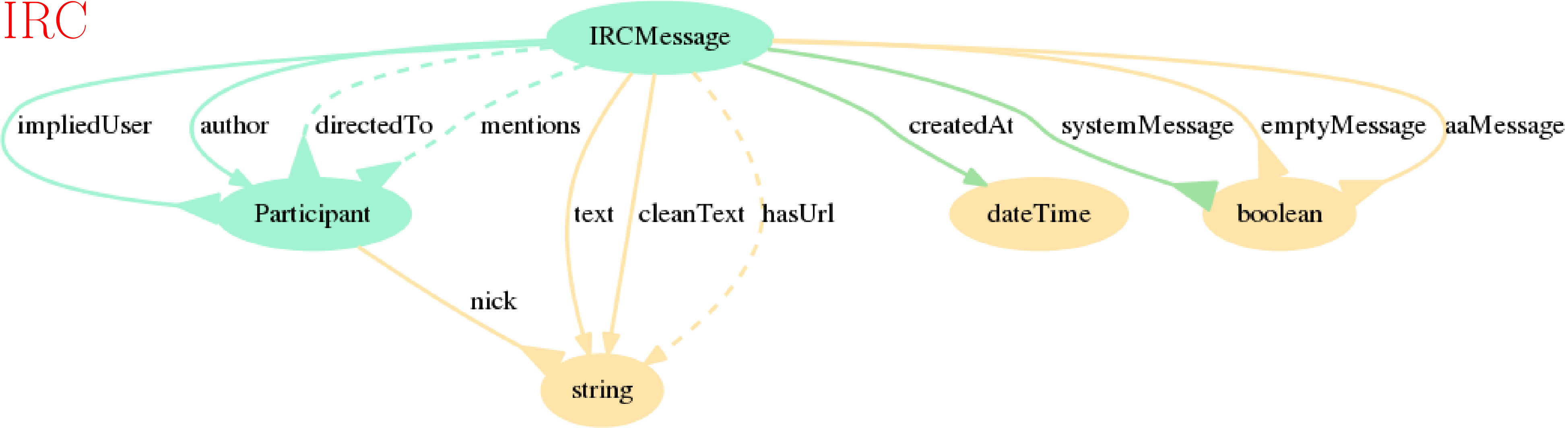

Tweet
retweetOfreplyTo
media
userMention
author
Media
hashtag
language
numericID
text
url
Participant
createdAt
retweetCount
noid
url
type
name
numericID
screenName
utcOffset
language
listedCount
statusesCount
utcOffset
favouritesCount
followersCount
friendsCount
anonymous
createdAt
string
integer
boolean
dateTime

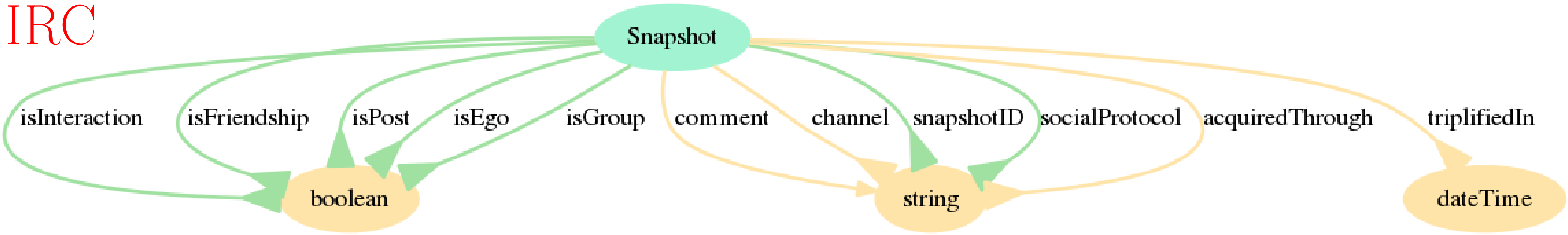

Twitter
Snapshot
isEgo
isPost
isGroup
isFriendship
isInteraction
comment
socialProtocol
snapshotID
hashtag
triplifiedIn
boolean
string
dateTime

4. IRC data

Each IRC snapshot is yield by an IRC channel. IRC messages are either server messages (e.g. join and exit channel) marked with `po:systemMessage true` and having an `po:impliedUser user_uri`, or user messages, which yield interactions through `po:directedTo` and `po:mentions` properties. Text messages without the user names are delivered through the `po:cleanText` property. Further information is found on the following diagrams, the tables on the end of this document or in the main document of this article [1].

IRC

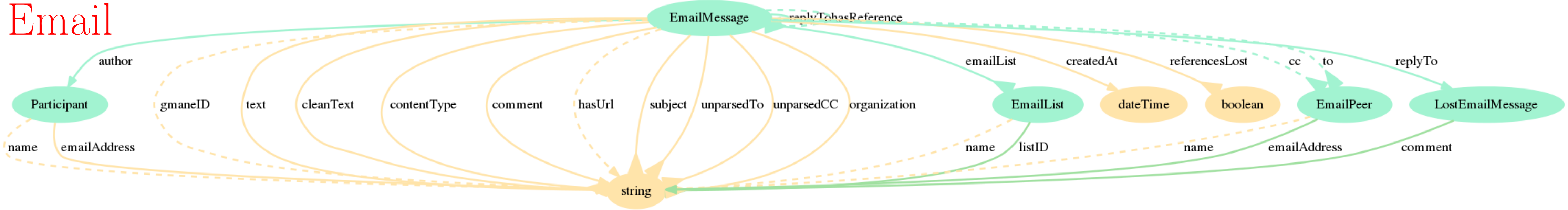

IRCMessage
impliedUser
author
directedTo
mentions
createdAt
systemMessage
emptyMessage
aaMessage
Participant
text
cleanText
hasUrl
dateTime
boolean
nick
string


IRC
Snapshot
isInteraction
isFriendship
isPost
isEgo
isGroup
comment
channel
snapshotID
socialProtocol
acquiredThrough
triplifiedIn
boolean
string
dateTime

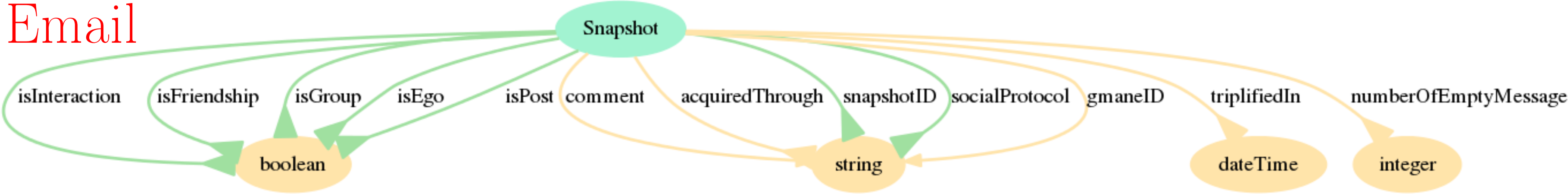
5. Email data

Each IRC snapshot is yield by an Email list. Interactions is yield through `po:replyTo` relations although `po:to` and `po:cc` can also be considered. the email body is given by `po:text` relations while `po:cleanText` has the text with lines removed where they are trivially from previous messages or computer code. Further information is found on the following diagrams, the tables on the end of this document or in the main document of this article [1].

Email


EmailMessage
replyTohasReference
author
emailList
createdAt
referencesLost
cc
to
replyTo
Participant
gmaneID
text
cleanText
contentType
comment
hasUrl
subject
unparsedTo
unparsedCC
organization
EmailList
dateTime
boolean
EmailPeer
LostEmailMessage
name
emailAddress
name
listID
name
emailAddress
comment
string

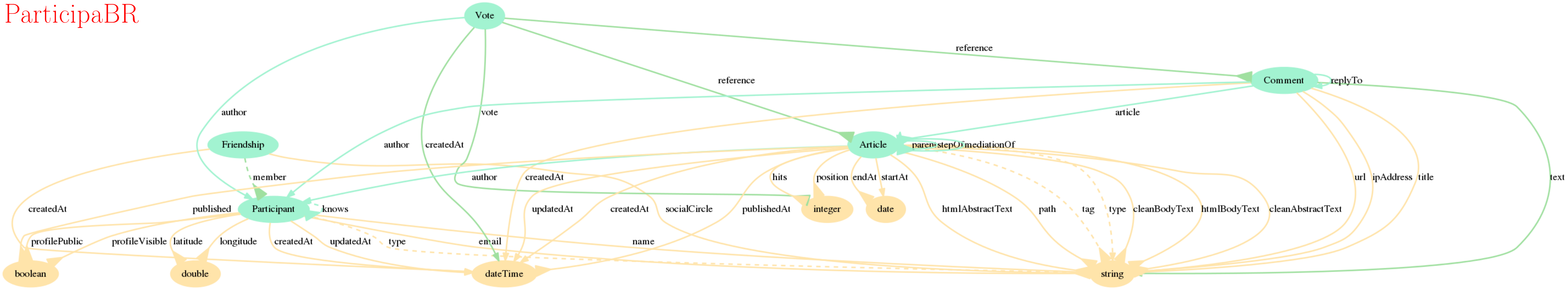
Email


Snapshot
isInteraction
isFriendship
isGroup
isEgo
isPost
comment
acquiredThrough
snapshotID
socialProtocol
gmaneID
triplifiedIn
numberOfEmptyMessages
boolean
string
dateTime
integer

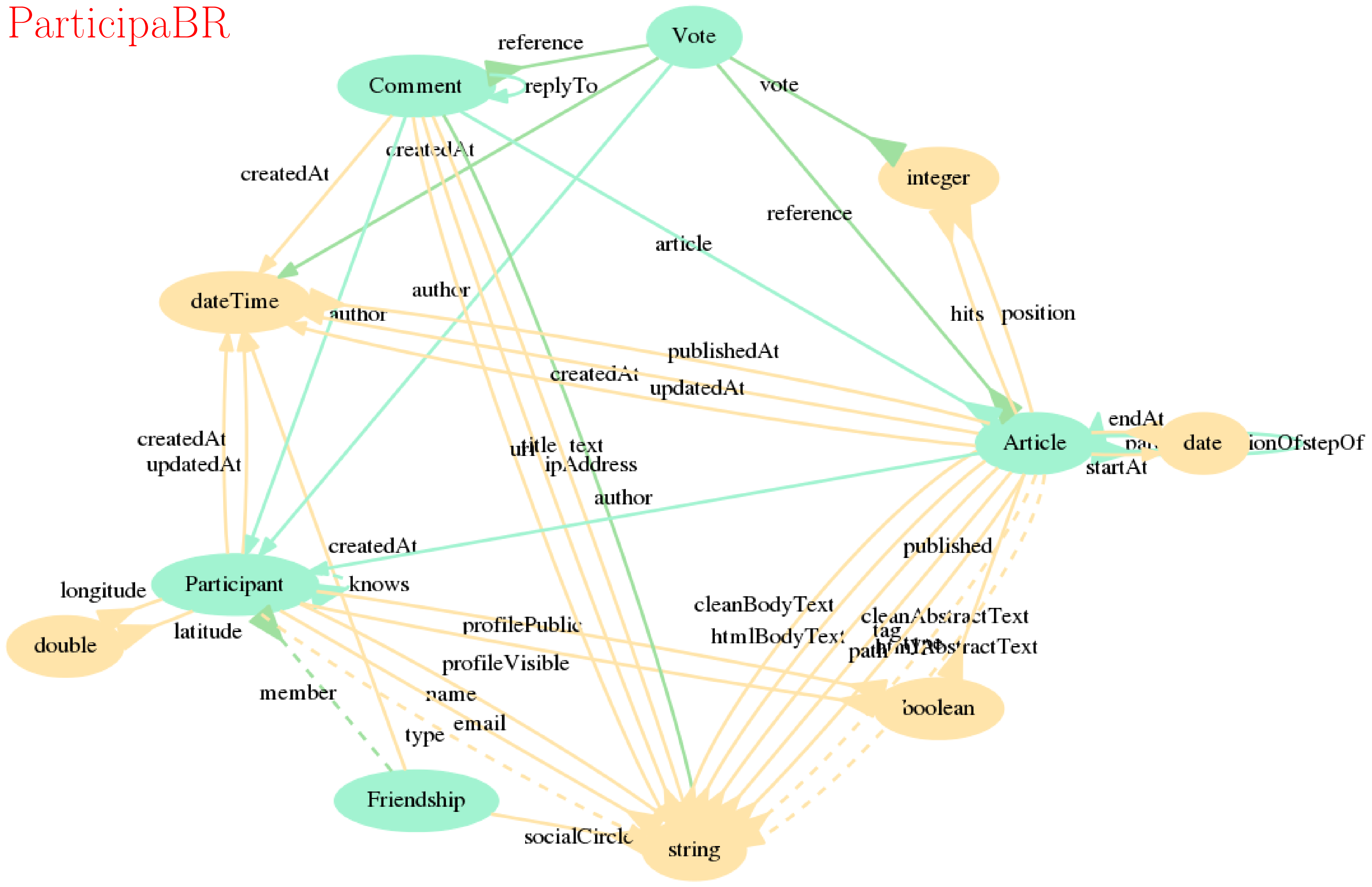
6. ParticipaBR data

The ParticipaBR snapshot is yield by a data dump donated by the system administrators of the federal portal of social participation ParticipaBR. Articles can have parent articles (`po:parent`), be step of a collection of articles (`po:stepOf`) and be a mediation of other articles (`po:mediationOf`). Interactions are yield by comments which are `po:replyTo` other comments or which are made directly to an article. This snapshot holds also friendship structures. The language used is mainly Brazilian Portuguese, but English and Spanish are also incident. Due to the higher complexity of the diagram, an additional figure is given rendered with another layout algorithm Further information is found on the following diagrams, the tables on the end of this document or in the main document of this article [1].

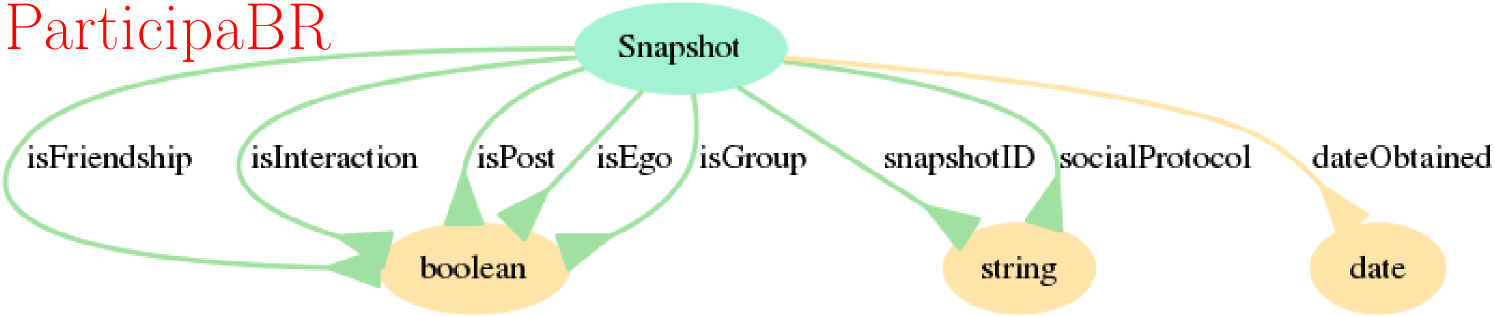
ParticipaBR



ParticipaBR



ParticipaBR



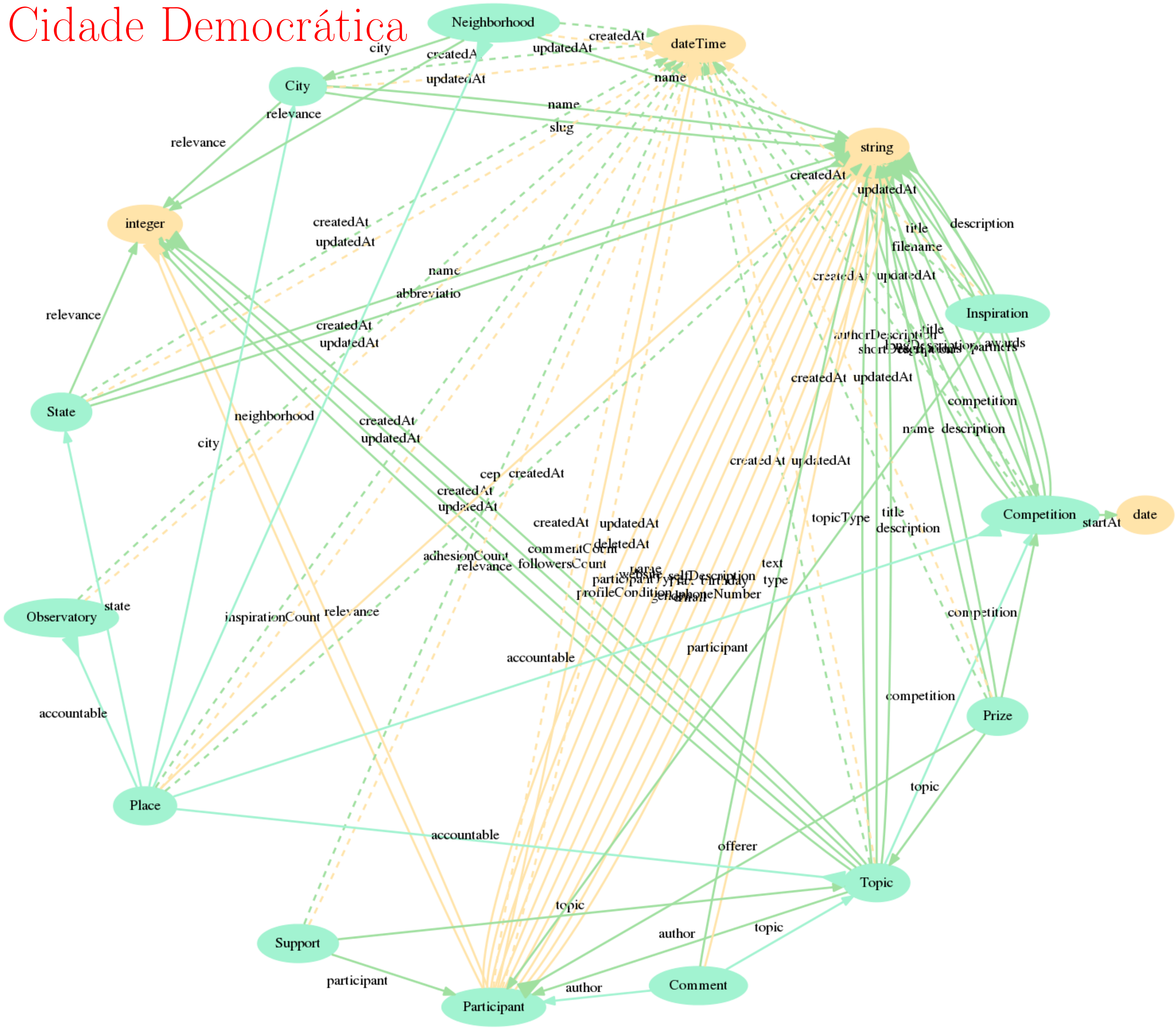
7. Cidade Democrática data

The Cidade Democrática snapshot is yield by a data dump donated by the system administrators of the civil society social participation portal Cidade Democrática. This snapshot holds a complex structure of both Topics/Inspirations/ Observatories/Supports/Competitions/Prizes and of State/City/Neighborhood/Place. The language used is mainly Brazilian Portuguese. Due to the higher complexity of the diagram, an additional figure is given rendered with another layout algorithm Further information is found on the following diagrams, the tables on the end of this document or in the main document of this article [1].

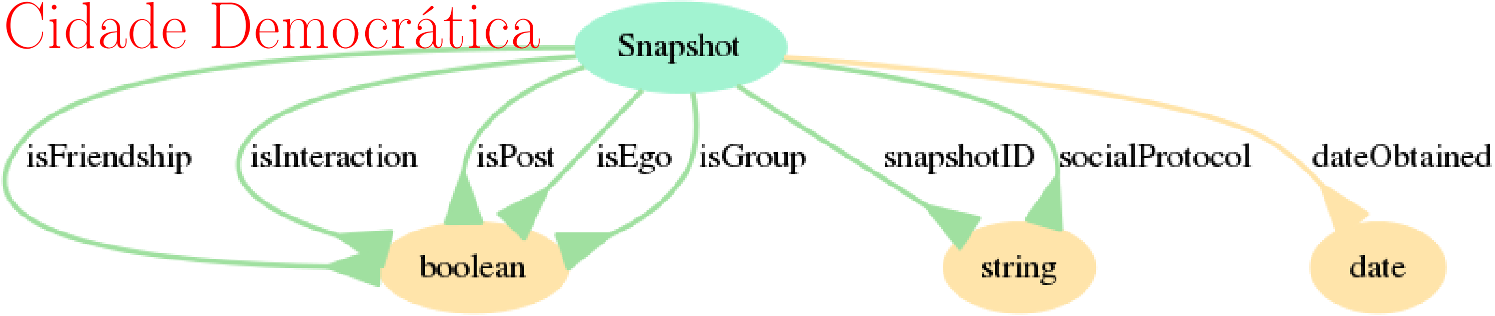
Cidade Democrática

Prize
Comment
Place
Support
topic
topic
accountable
neighborhood
topic
competition
Inspiration
accountable
offerer
author
participant
Topic
state
city
Neighborhood
accountable
description
name
text
competition
participant
type
createdAt
cep
competition
author
createdAt
city
updatedAt
createdAt
updatedAt
createdAt
updatedAt
description
filename
title
Competition
Participant
topicType
title
description
adhesionCount
name
followersCount
relevance
commentCount
State
updatedAt
createdAt
createdAt
City
relevance
updatedAt
createdAt
updatedAt
Observatory
startAt
authorDescription
awards
longDescription
partners
regulations
shortDescription
title
selfDescription
website
name
email
birthday
fax
gender
participantType
phoneNumber
profileCondition
inspirationCount
relevance
abbreviatio
name
name
createdAt
updatedAt
createdAt
updatedAt
deletedAt
relevance
slug
relevance
createdAt
updatedAt
updatedAt
createdAt
createdAt
updatedAt
date
string
integer
dateTime

Cidade Democrática

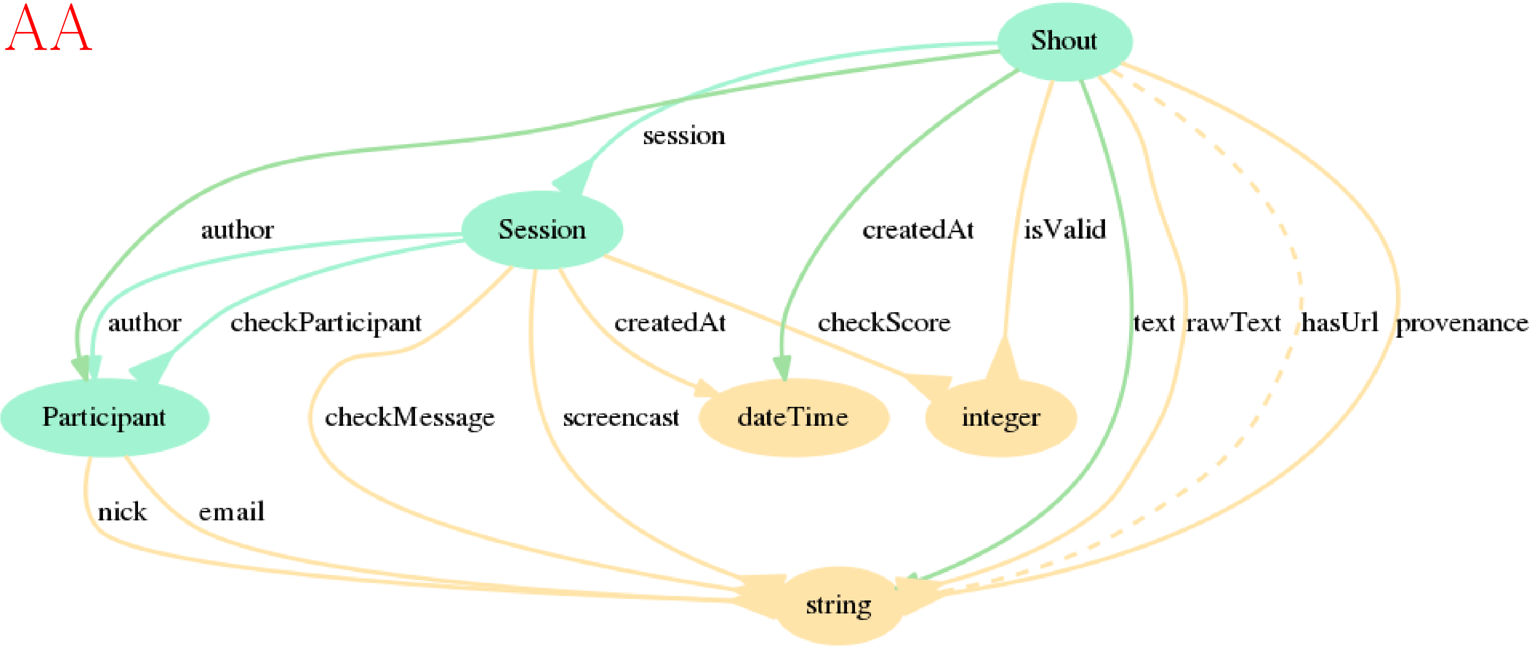


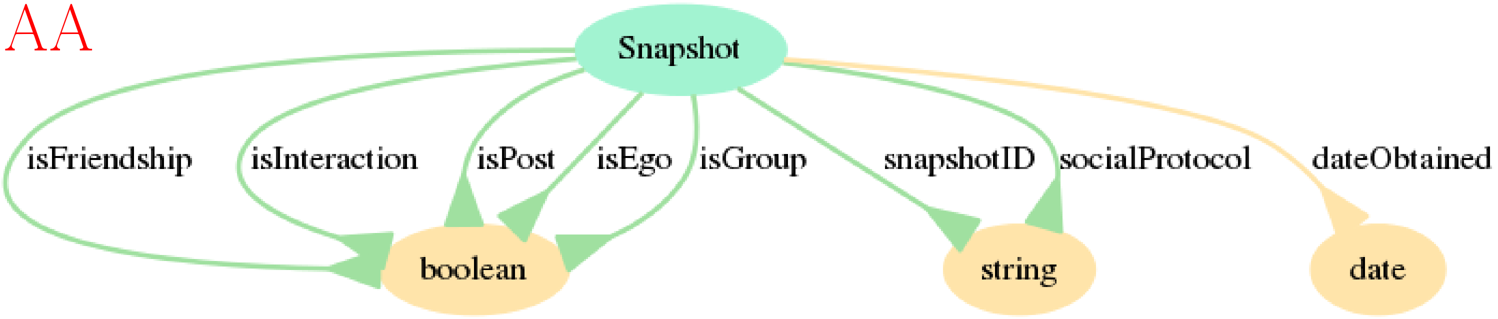
Cidade Democrática



8. AA data

The AA (Algorithmic Autoregulation) snapshots are yield by a data dump donated by the system administrators and by a mined IRC log. The system pursue simplicity and most of data consists of detached shouts with `po:text` and `po:author`. Further information is found on the following diagrams, the tables on the end of this document or in the main document of this article [1].


AA
Shout
session
Session
author
createdAt
isValid
author
checkParticipant
createdAt
checkScore
text
rawText
hasUrl
provenance
Participant
checkMessage
screencast
dateTime
integer
nick
email
string


AA
Snapshot
isFriendship
isInteraction
isPost
isEgo
isGroup
snapshotID
socialProtocol
dateObtained
boolean
string
date

9. Snapshot references

Table 1: All the Facebook snapshots are either the result of individuals who downloaded their data (and donated to the first author) or data downloaded from groups. In the first case, it is senseless to present references. In the second case, we present the group name and a link to a post in the group where data and figures were delivered to the group.

group name	**url(s)**
Adorno Nao Eh Enfeite	https://www.facebook.com/groups/265217103529531/permalink/525654127485826/
Ativistas Da Inclusao Digital	https://www.facebook.com/groups/423602557691243/permalink/525201037531394/
Ciencias Com Fronteiras	https://www.facebook.com/groups/contraaexclusao/permalink/269103356558439/
Computer Art	https://www.facebook.com/groups/computerart/permalink/259389137529870/
Coolmeia	https://www.facebook.com/groups/coolmeia/permalink/380091142098291/ , https://www.facebook.com/groups/coolmeia/permalink/489757754464962/
Democracia Direta Ja	https://www.facebook.com/groups/ddjbrasil/permalink/347023325397298/
Democracia Pura	https://www.facebook.com/groups/democraciapura/permalink/310907215704321/
Economia	https://www.facebook.com/groups/economa1/permalink/586007714743535/
Economia Criativa Digital	https://www.facebook.com/groups/economiacriativadigital/permalink/438313682916103/
Educacoes E Aprendizagens XXI	https://www.facebook.com/groups/geaxxi/permalink/433229973421182/
Latesfip	https://www.facebook.com/groups/183557128478424/permalink/266610616839741/
Living Bridges Planet	https://www.facebook.com/groups/livingbridgesplanet/permalink/352950408144951/
Mobilizacoes Culturais Interior SP	https://www.facebook.com/groups/131639147005593/permalink/144204529082388/
Partido Pirata	https://www.facebook.com/groups/partidopiratabrasil/permalink/10151409024509317/
Politicas Culturas Brasileiras	https://www.facebook.com/groups/pcult/permalink/519626544747423/
Praca Popular	https://www.facebook.com/groups/215924991863921/permalink/319279541528465/
Rede Tranzmidias	https://www.facebook.com/groups/318333384951196/permalink/346658712118663/
Silicon Valley Global Network	https://www.facebook.com/groups/109971182359978/permalink/589326757757749/
Solidarity Economy	https://www.facebook.com/groups/9149038282/permalink/10151461945623283/
Study Group SNA	https://www.facebook.com/groups/140630009439814/permalink/151470598355755/
Tecnoxamanismo	https://www.facebook.com/groups/505090906188661/permalink/733144993383250/ , https://www.facebook.com/groups/505090906188661/permalink/733157380048678/

Table 2: Different Twitter snapshots are yield by different hashtags. In this table we present each snapshot with the respective hashtag and a reference to the subject.

snapshot hashtag	**observation**	**reference**
#arenaNETmundial	a Brazilian discussion hub about free culture, democracy and the internet	http://www.participa.br/netmundial
#art	tweets with the generic hashtag #art	https://en.wikipedia.org/wiki/Art
#ChennaiFloods	heavy rainfall generated by the annual northeast monsoon in November–December 2015	https://en.wikipedia.org/wiki/2015_South_Indian_floods
#dilma	the 36th President of Brazil	https://en.wikipedia.org/wiki/Dilma_Rousseff
#ForaDilma	2015-16 anti-government protests in Brazil	https://en.wikipedia.org/wiki/2015-16_protests_in_Brazil
#ForaCunha	2015-16 anti-corruption protests in Brazil	https://en.wikipedia.org/wiki/2015-16_protests_in_Brazil
#fuck	tweets with the generic hashtag #fuck	https://en.wikipedia.org/wiki/Fuck
#game	tweets with the generic hashtag #game	https://en.wikipedia.org/wiki/Game
#god	tweets with the generic hashtag #god	https://en.wikipedia.org/wiki/God
#MAMA2015	the grand 2015 Mnet Asian Music Awards	https://en.wikipedia.org/wiki/2015_Mnet_Asian_Music_Awards
#music	tweets with the generic hashtag #music	https://en.wikipedia.org/wiki/Music
#obama	the 44th President of the United States	https://en.wikipedia.org/wiki/Barack_Obama
#python	the Python programming language	https://en.wikipedia.org/wiki/Python_(programming_language)
#QuartaSemRacismoClubeSDV	an anti-racism netweaving	https://twitter.com/hashtag/quartasemracismoclubesdv
#science	tweets with the generic hashtag #science	https://en.wikipedia.org/wiki/Science
#SnapDetremura	reference for Snapchat about a celebrated person	https://twitter.com/detremura

Table 3: Different IRC snapshots are yield by different channels. In this table we present each snapshot with the respective channel and a reference to the subject.

snapshot channel	observation	reference
#foradoeixo	a Brazilian network of culture related collectives	https://pt.wikipedia.org/wiki/Fora_do_Eixo
#hackerspace-cps	a hackerspace in Campinas, Brazil	https://lhc.net.br/wiki/P%C3%A1gina_principal
#hackerspaces-br	Brazilian hackerspaces channel	https://garoa.net.br/wiki/Hackerspaces_Brasileiros
#labmacambira	Brazilian channel for the lab-Macambira collective	http://labmacambira.sourceforge.net/

Table 4: Different Email snapshots are yield by different email lists. In this table we present each snapshot with the respective list and a reference to the subject.

Gmane ID	observation	reference
gmane.linux.audio.users	the Linux Audio Users	http://linuxaudio.org
gmane.politics.organizations.metareciclagem	a network about technology and social transformation	https://metareciclagem.github.io
gmane.linux.audio.devel	the Linux Audio Developers	http://lists.linuxaudio.org/listinfo/linux-audio-dev
gmane.comp.gcc.libstdc++.devel	the C++ standard library	https://gcc.gnu.org/libstdc++/

Table 5: References for the snapshots of the detached instances ParticipaBR, Cidade Democrática and AA.

social protocol	observations	reference
ParticipaBR	a Brazilian federal portal of social participation	http://www.participa.br/
Cidade Demorática	a Brazilian civil society portal of social participation	http://www.cidadedemocratica.org.br/
AA	the Algorithmic Autoregulation software development methodology	[2]

10. Trivial counts in each snapshot

Table 6: Number of triples (ntriples), number of relations/interactions/edges (nedges), number of participants (nparticipants) and number of characters (nchars) in each LOSD snapshot.

snapshot id	ntriples	nedges	nparticipants	nchars
aa-irc-legacy-labmacambira_lalenia.txt	53,140	0	117	558,466
aa-mongo-legacy	22,773	0	37	240,172
aa-mysql-legacy	790,796	0	157	2,753,354
cidadedemocratica-legacy	915,852	0	23,079	6,871,848
facebook-legacy-AdornoNaoEhEnfeite29032013	8,459	1,292	293	26,113
facebook-legacy-AntonioAnzoategui18022013	1,676	328	52	0
facebook-legacy-AtivistasDaInclusaoDigital09032013	25,642	5,592	306	0
facebook-legacy-Auricultura10042013	19,515	3,898	412	14,015
facebook-legacy-BrunoMialich31012013	40,794	9,320	502	0
facebook-legacy-CalebLuporini13042013	105,962	24,653	1,050	0
facebook-legacy-CalebLuporini19022013	104,746	24,391	1,026	0
facebook-legacy-CienciasComFronteiras29032013	110,734	23,302	2,921	0
facebook-legacy-ComputerArt10032013	260,050	62,819	1,342	0
facebook-legacy-Coolmeia06032013	76,063	16,534	1,202	0
facebook-legacy-DanielPenalva18022013	3,519	682	113	0
facebook-legacy-DemocraciaDiretaJa14032013	258,679	59,323	3,053	54,443
facebook-legacy-DemocraciaDiretaJa14072013	257,490	59,781	3,607	58,035
facebook-legacy-DemocraciaPura06042013	32,227	6,730	627	65,062
facebook-legacy-Economia14042013	238,790	54,001	3,587	52,664
facebook-legacy-EconomiaCriativaDigital03032013	185,905	43,128	1,684	0
facebook-legacy-EducacoesEAprendizagensXXI02032013	106,918	24,802	1,285	0
facebook-legacy-GabrielaThume19022013	18,581	4,108	307	0
facebook-legacy-GrahamForrest28012013	1,370	185	90	0
facebook-legacy-LailaManuelle17012013	201,071	48,572	969	0
facebook-legacy-LarissaAnzoategui20022013	24,824	5,191	580	0
facebook-legacy-Latesfip08032014	11,554	2,009	306	0
facebook-legacy-LivingBridgesPlanet29032013	149,708	32,494	3,077	52,808
facebook-legacy-LuisCirne07032013	16,619	3,390	437	0
facebook-legacy-MariliaMelloPisani10042013	84,770	19,040	1,230	0
facebook-legacy-Mirtes16052013	39,415	9,075	445	0
facebook-legacy-MobilizacoesCulturaisInteriorSP13032013	26,518	6,096	298	0
facebook-legacy-PartidoPirata23032013	45,495	8,537	1,943	36,313

Continued on next page

Table 6 – *Continued from previous page*

snapshot id	**ntriples**	**nedges**	**nparticipants**	**nchars**
facebook-legacy-PedroPauloRocha10032013	215,888	50,591	1,932	0
facebook-legacy-PeterForrest28012013	8,156	1,829	120	0
facebook-legacy-PoliticasCulturasBrasileiras08032013	178,289	41,690	1,278	69,756
facebook-legacy-PracaPopular16032013	4,539	932	65	4,249
facebook-legacy-RafaelReinehr09042013	174,423	39,586	2,297	0
facebook-legacy-RamiroGiroldo20022013	9,928	2,020	264	0
facebook-legacy-RedeTranzmidias02032013	25,111	4,940	391	54,907
facebook-legacy-RenatoFabbri02032013	93,134	21,579	974	0
facebook-legacy-RenatoFabbri03032013	93,690	21,711	978	0
facebook-legacy-RenatoFabbri11072013	114,552	26,440	1,256	0
facebook-legacy-RenatoFabbri18042013	104,156	24,072	1,124	0
facebook-legacy-RenatoFabbri20012013	86,416	20,085	868	0
facebook-legacy-RenatoFabbri29112012	82,093	19,083	823	0
facebook-legacy-RicardoFabbri18022013	11,372	2,327	344	0
facebook-legacy-RitaWu08042013	83,895	18,935	1,165	0
facebook-legacy-RonaldCosta12062013	31,338	6,557	730	0
facebook-legacy-SiliconValleyGlobalNetwork27042013	77,158	15,740	2,130	50,251
facebook-legacy-SolidarityEconomy12042013	14,230	2,404	525	67,774
facebook-legacy-StudyGroupSNA05042013	5,604	480	448	25,474
facebook-legacy-THackDay26032013	1,844	420	41	0
facebook-legacy-Tecnoxamanismo08032014	11,106	2,069	318	0
facebook-legacy-Tecnoxamanismo15032014	14,589	2,702	450	0
facebook-legacy-ThaisTeixeira19022013	26,424	6,088	296	0
facebook-legacy-VilsonVieira18022013	19,688	4,334	336	0
facebook-legacy-ViniciusSampaio18022013	90,515	21,360	725	0
facebook-legacy-avlab_BarthorLaZule22022014	16,005	3,513	279	0
facebook-legacy-avlab_CalebLuporini25022014	125,577	29,268	1,215	0
facebook-legacy-avlab_CamilaBatista23022014	21,138	4,476	462	0
facebook-legacy-avlab_CarlosDiego25022014	171,744	39,401	2,020	0
facebook-legacy-avlab_CristinaMekitarian23022014	24,647	5,572	337	0
facebook-legacy-avlab_DanielGonzales23022014	196,406	45,318	2,162	0
facebook-legacy-avlab_FelipeBrait23022014	1,228,605	299,082	4,611	0
facebook-legacy-avlab_FelipeVillela22022014	2,475	477	81	0
facebook-legacy-avlab_JoaoMeirelles25022014	52,371	11,649	825	0
facebook-legacy-avlab_JoaoMekitarian23022014	88,765	20,821	783	0
facebook-legacy-avlab_JulianaSouza23022014	129,757	29,942	1,427	0
facebook-legacy-avlab_KarinaGomes22022014	9,073	1,906	207	0

Continued on next page

Table 6 – *Continued from previous page*

snapshot id	**ntriples**	**nedges**	**nparticipants**	**nchars**
facebook-legacy-avlab_LucasOliveira26022014	62,871	14,764	545	0
facebook-legacy-avlab_MarcelaLucatelli25022014	138,733	31,647	1,735	0
facebook-legacy-avlab_MariliaPisani25022014	114,765	25,830	1,635	0
facebook-legacy-avlab_NatachaRena22022014	642,769	154,758	3,391	0
facebook-legacy-avlab_OrlandoCoelho22022014	5,149	848	251	0
facebook-legacy-avlab_PalomaKliss25022014	493,774	119,520	2,242	0
facebook-legacy-avlab_PedroRocha25022014	346,883	81,910	2,749	0
facebook-legacy-avlab_RenatoFabbri22022014	124,703	28,780	1,369	0
facebook-legacy-avlab_SarahLuporini25022014	505,853	121,502	2,835	0
facebook-legacy-avlab_SatoBrasil25022014	1,519,394	371,249	4,914	0
facebook-legacy-ego_MarceloSaldanha19112014	130,556	29,440	1,828	0
facebook-legacy-ego_MariliaPisani06052014	122,231	27,581	1,701	0
facebook-legacy-ego_MassimoCanevacci19062013	273,328	59,995	4,764	0
facebook-legacy-ego_RenatoFabbri06022014	123,993	28,606	1,367	0
facebook-legacy-ego_RenatoFabbri19112014	153,410	35,514	1,622	0
facebook-legacy-ego_VJPixel23052014	231,608	54,752	1,800	0
facebook-legacy-posavlab_AnaCelia18032014	53,939	12,167	753	0
facebook-legacy-posavlab_ElenaGarnelo04032014	93,472	21,723	940	0
facebook-legacy-posavlab_FabiBorges08032014	159,584	36,592	1,888	0
facebook-legacy-posavlab_GeorgeSanders08032014	108,071	24,706	1,321	0
facebook-legacy-posavlab_GrazielleMachado18032014	24,264	5,254	464	0
facebook-legacy-posavlab_RenatoFabbri19032014	129,360	29,890	1,400	0
facebook-legacy-posavlab_RicardoPoppi18032014	76,104	17,234	1,024	0
gmane-legacy-comp.gcc.libstdcpp.devel1-20000	324,051	14,786	1,036	30,126,252
gmane-legacy-linux.audio.devel1-20000	377,307	17,076	1,232	26,969,596
gmane-legacy-linux.audio.users1-20000	349,304	16,362	1,147	25,065,928
gmane-legacy-politics.organizations.metareciclagem1-20000	338,679	15,230	477	54,260,954
irc-legacy-foradoeixo	685,623	4,308	3,318	3,777,424
irc-legacy-hackerspace-cps	253,614	1,860	607	1,059,675
irc-legacy-hackerspaces-br	980,556	210	347	8,420,840
irc-legacy-labmacambira	1,535,463	58,525	1,561	8,358,970
participabr-legacy	159,602	2,207	3,825	2,045,617
twitter-legacy-ChennaiFloods	6,793,705	101,824	46,493	23,237,802
twitter-legacy-ForaCunha	88,963	1,656	2,747	372,131
twitter-legacy-ForaDilma	26,818	668	659	113,810
twitter-legacy-MAMA2015	20,356,960	426,558	33,080	75,358,785
twitter-legacy-QuartaSemRacismoClubeSDV	367,460	5,000	5,785	1,635,867

Continued on next page

Table 6 – *Continued from previous page*

snapshot id	**ntriples**	**nedges**	**nparticipants**	**nchars**
twitter-legacy-SnapDetremura	26,834	461	621	124,448
twitter-legacy-arenaNETmundial	388,134	15,797	5,898	2,825,121
twitter-legacy-art	2,814,803	32,655	30,486	9,539,413
twitter-legacy-dilma	78,424	1,692	2,274	332,005
twitter-legacy-fuck	3,631	40	93	14,727
twitter-legacy-game	229,992	1,548	4,682	1,229,910
twitter-legacy-god	1,514,365	20,132	22,117	5,560,140
twitter-legacy-music	8,150,863	39,456	54,006	21,116,573
twitter-legacy-obama	1,143,873	20,623	20,330	4,481,840
twitter-legacy-python	5,786	4	17	1,758
twitter-legacy-science	369,013	3,673	7,156	1,910,216

References

[1] O. N. d. O. J. Renato Fabbri, Linked open social data for scientific benchmarking, https://github.com/ttm/linkedOpenSocialData/raw/master/paper.pdf (2016).

[2] R. Fabbri, R. Fabbri, V. Vieira, D. Penalva, D. Shiga, M. Mendonça, A. Negrão, L. Zambianchi, G. S. Thumé, The algorithmic autoregulation software development methodology/a metodologia de desenvolvimento de software autorregulação algorítmica, Revista Electronica de Sistemas de Informaçao 13 (2) (2014) 1.