

Tesis de Grado de Ingeniería en Informática

Estudio de PeeringDB para identificación y extracción de siblings

Alumno: Augusto Arturi

Director: Dr. Ing. Esteban Carisimo

Facultad de Ingeniería, Universidad de Buenos Aires

27 de agosto de 2021



Objetivo



Contexto



Internet es una red de comunicaciones en la que operan múltiples **conglomerados**.

¿ Que es un Conglomerado ?



Un conglomerado es una compañía que tiene subsidiarias y asociados que opera en distintas unidades de negocio, generalmente de forma multinacional.

TELECOM



Personal



Fibertel

Cablevisión



Referencias

- Claro
- Claro (Ecuador)
- Claro (Colombia)
- Telcel (México)
- TracFone Wireless (USA)
- Ninguno (la compañía no opera en estos países)

Objetivo



Contexto



Internet es una red de comunicaciones en la que operan múltiples **conglomerados**.

Problemática



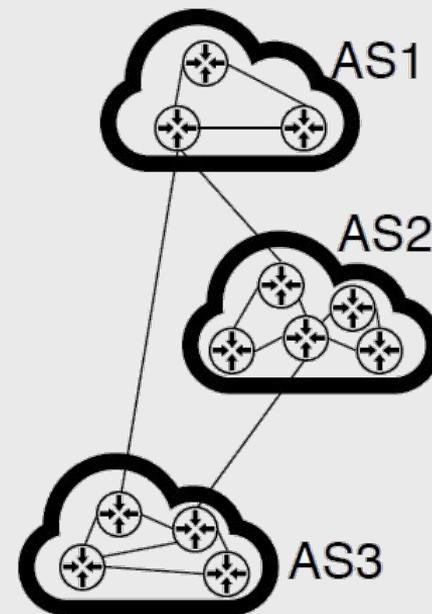
Las metodologías actuales son insuficientes para identificar el conjunto de **ASNs** operados por un mismo conglomerado.

Sistema Autónomo (AS)



Un AS es un conjunto de prefijos que son operados por una misma organización, con una única política de ruteo.

- Identificado por el Número de Sistema Autónomo (**ASN**)
- ASNs son dados por el IANA¹
- Los ASes son los **únicos habilitados** a solicitar direcciones IPs a los RIRs².



¹Internet Assigned Numbers Authority

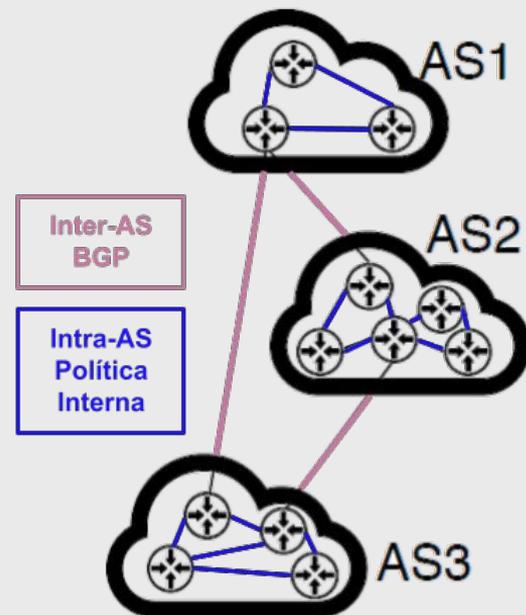
²Regional Internet Registry

Sistema Autónomo (AS)



Un AS es un conjunto de prefijos que son operados por una misma organización, con una única política de ruteo.

- Identificado por el Número de Sistema Autónomo (**ASN**)
- ASNs son dados por el IANA¹
- Los ASes son los **únicos habilitados** a solicitar direcciones IPs a los RIRs².



¹Internet Assigned Numbers Authority

²Regional Internet Registry

Objetivo



Contexto



Internet es una red de comunicaciones en la que operan múltiples **conglomerados**.

Problemática



Las metodologías actuales son insuficientes para identificar el conjunto de **ASNs** operados por un mismo conglomerado.

Propuesta



Mejorar la **representación** de los conglomerados en la red de Sistemas Autónomos.

Sistema Autónomo (AS)



Un AS es un conjunto de prefijos que son operados por una misma organización de red, con una única política de ruteo.

≈ 72.000 ASes activos



Sistema Autónomo (AS)



¿ Cómo se interconectan ?

Mediante un enlace físico.

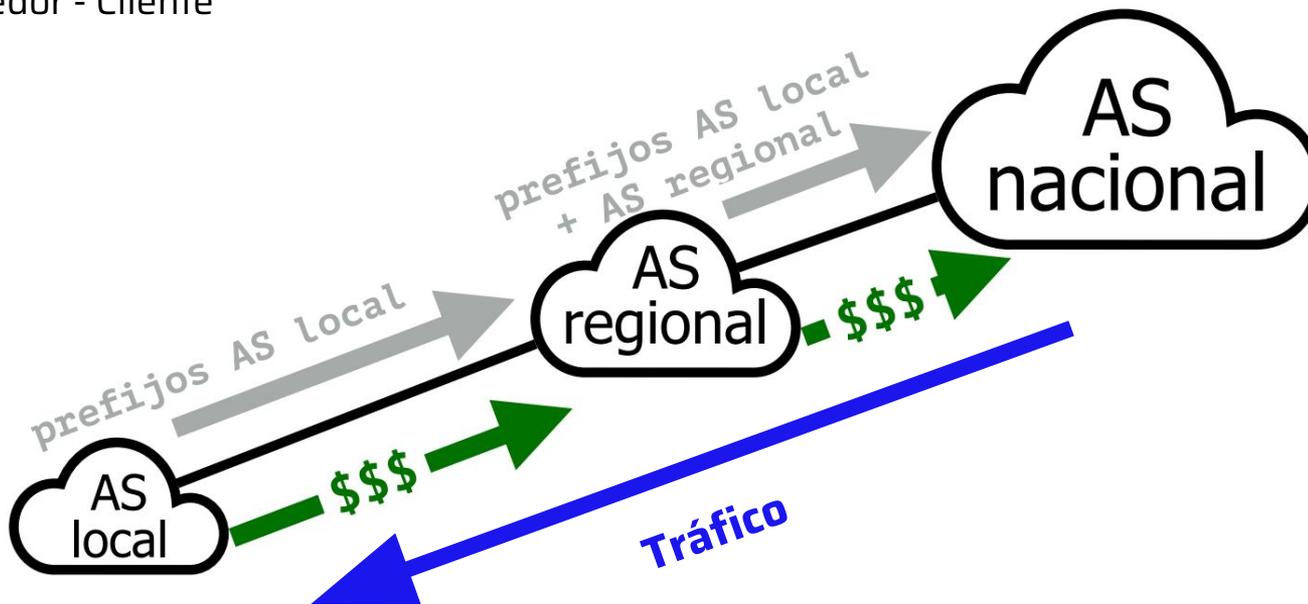
¿Para qué se interconectan?

Usuarios puedan interconectarse entre sí.

Relaciones entre ASes - p2c



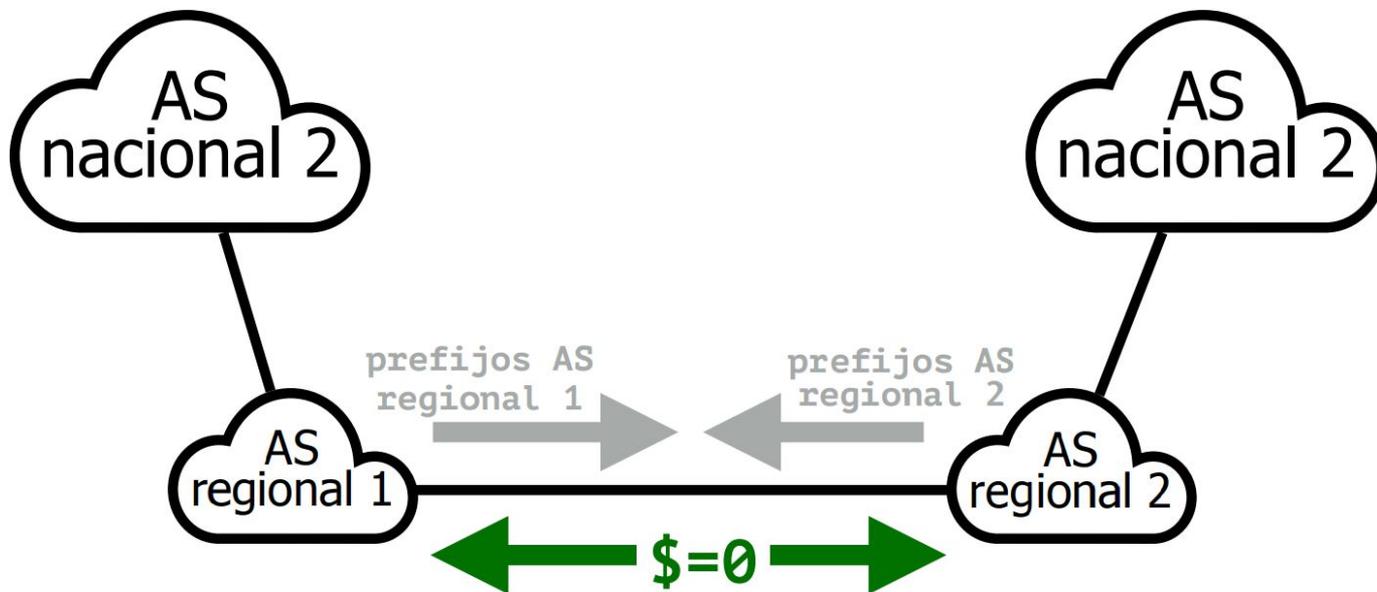
Proveedor - Cliente



Relaciones entre ASes - p2p



Entre pares



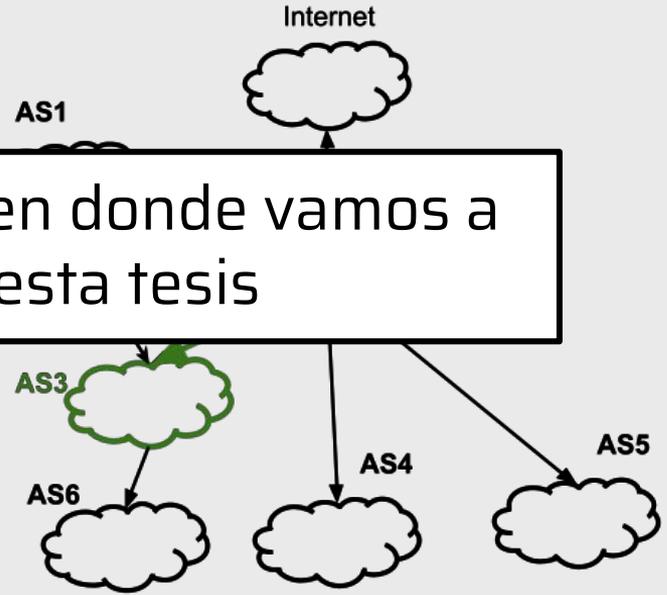
Siblings



ASNs que pertenecen al mismo conglomerado.

- No aplica para siblings
- Los siblings implementan políticas de ruteo flexibles.
- Es difícil inferir este tipo de relaciones a través de BGP.

Es la relación de **siblings** en donde vamos a hacer énfasis en esta tesis



Metodologia



Estado del arte



Cai et al. → Vínculo entre Sistemas Autónomos y Organizaciones

Los conglomerados son difíciles de vincular a infraestructuras

Estructuras de negocio complejas

Características propias de los protocolos de Internet

Segmentación acorde a su estructura comercial

BGP no requiere de información comercial para su funcionamiento



AS PATH

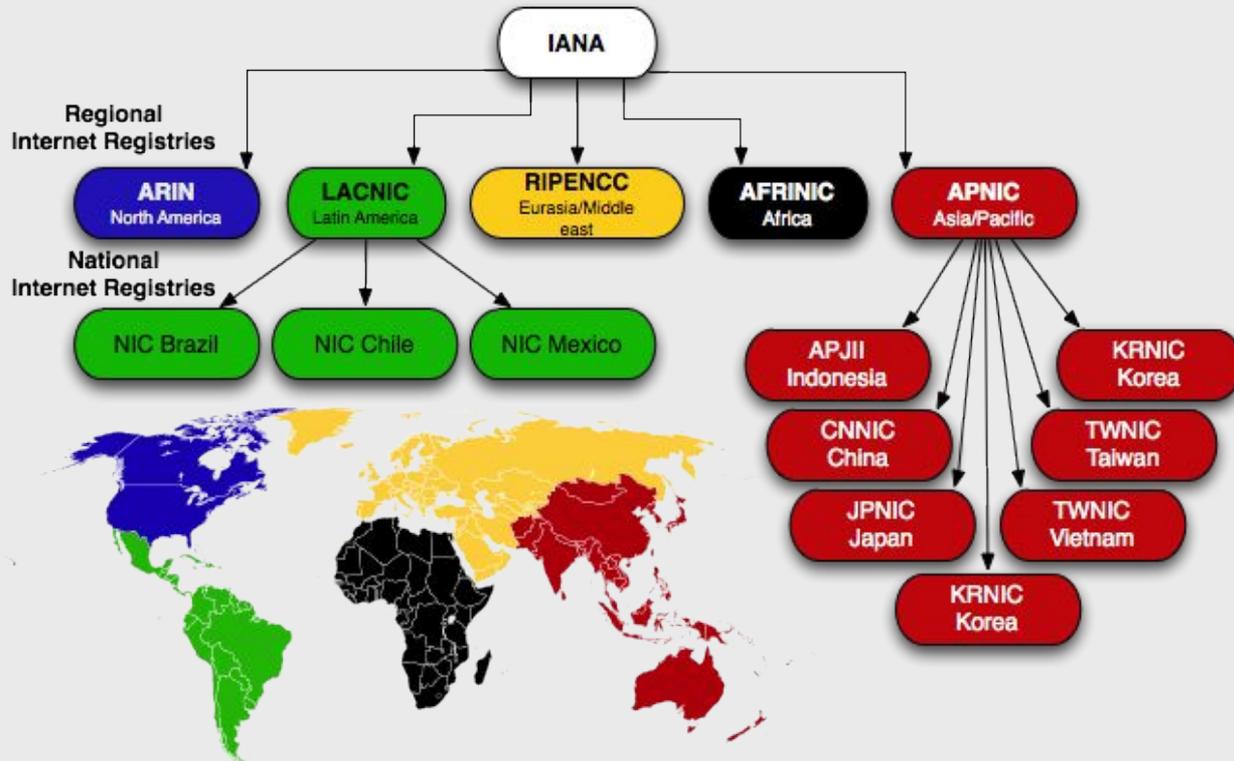
Estado del arte



Cai et al.  Vínculo entre Sistemas Autónomos y Organizaciones

Limitaciones

- ❖ Información semiestructurada, sin esquema consistente.
- ❖ Baja cohesión en los datos y separados por RIR.
- ❖ Información obsoleta e imprecisa.
- ❖ Límite de consultas.
- ❖ Problemas para vincular ASes delegados por distintos RIRs.



Estado del arte



Cai et al.  [Vínculo entre Sistemas Autónomos y Organizaciones](#)

Limitaciones

- ❖ Información semiestructurada, sin esquema consistente.
- ❖ Baja cohesión en los datos y separados por RIR.
- ❖ Información obsoleta e imprecisa.
- ❖ Límite de consultas.

PeeringDB



Es una organización, que facilita el intercambio de información relacionada con interconexión, principalmente para Operadores de Red.

- ❖ Es una herramienta pública para el crecimiento y entendimiento de Internet.
- ❖ La información disponible es representativa, correcta y actual.¹
- ❖ Es utilizado en investigaciones.²
- ❖ Base de datos orientada a las operaciones.

¹Aemen Lodhi y col. \Using PeeringDB to Understand the Peering Ecosystem".

²Timm Böttger, Felix Cuadrado y Steve Uhlig. \Looking for hypergiants in PeeringDB".

PeeringDB



¿ Por qué ?

- ❖ No se conoce ninguna otra alternativa que reúna tantos ASNs en un único lugar.
- ❖ La centralización simplifica.
- ❖ Existe un mapeo indexado entre ASNs y organizaciones.
- ❖ Auto-reportado.
- ❖ Es una base de datos aceptada y utilizada por la comunidad científica.

¿ Para qué ?

Existe evidencia de factibilidad para la inferencia de siblings a través de los campos:

notas, aka y org_id.

The Interconnection Database

Join. Search. Grow your network.

PeeringDB is a freely available, user-maintained, database of networks, and the go-to location for interconnection data. The database facilitates the global interconnection of networks at Internet Exchange Points (IXPs), data centers, and other interconnection facilities, and is the first step in making interconnection decisions.

The database is a non-profit, community-driven initiative run and promoted by volunteers. It is a public tool for the growth and good of the Internet. Join the community and support the continued development of the Internet.

Learn more [about](#) PeeringDB or [register](#).

ACTUALIZACIONES MÁS RECIENTES

Punto de intercambio de tráfico

[EPIX.Poznan](#)
4 hace horas

[STACIX](#)
7 hace horas

[RO-CIX](#)
15 hace horas

[MIX-BT](#)
16 hace horas

[GigaPIX - LAN 2](#)
1 hace días

Redes

[FLORIAN-MIFSUD](#)
(210921)

16 hace minutos

[COBASO-NET](#) (399866)
34 hace minutos

[Inexa Tecnologia LTDA.](#)
(264159)

54 hace minutos

[INEXA - Flavio Jose](#)
[Penso Junior - ME](#)

(262299)
57 hace minutos

[WELITON VIEIRA](#)
[BATISTA](#) (272514)
1 hace horas

Instalaciones

[Datacenter Services](#)
[Douala](#)

15 hace horas

[Xelent](#)
18 hace horas

[CAT Tower \(Bang-rak\)](#)
23 hace horas

[CNI DC Bali](#)
1 hace días

[Parkovyi DC](#)
2 hace días

© 2004-2021 PeeringDB

Todos los derechos reservados. Mediante el uso de este servicio, usted acepta adherirse a nuestra [AUP](#).

[2.28.0 - Privacy Policy](#)



[About](#)

[Registrarse](#)

[Buscar](#)

[Patrocinadores](#)

Recursos

[API](#)

[Documentación](#)

[FAQ](#)

[Governance](#)

[Estado](#)

Contáctenos

support@peeringdb.com

Estadísticas del Sistema Global

904 Punto de intercambio de tráfico

22688 Redes

4241 Instalaciones

40739 Connections to Exchanges

36381 Connections to Facilities

4391 Automated Networks

Punto de intercambio de tráfico (0)

Redes (11)

- [Google LLC \(15169\)](#)
- [Google LLC AS19527 \(19527\)](#)
- [Google Fiber, Inc. \(16591\)](#)
- [Google Private Cloud \(16550\)](#)
- [Google Cloud Korea \(139070\)](#)
- [Google Cloud Indonesia \(139190\)](#)
- [Google IT \(36385\)](#)
- [Google LLC AS36040 \(36040\)](#)
- [Google LLC AS43515 \(43515\)](#)
- [Google Corporate Network in APAC \(45566\)](#)
- [euNetworks Cloud Connect \(30834\)](#)

Instalaciones (0)

Organizations (2)

- [Google Fiber, Inc.](#)
- [Google LLC](#)



Google LLC

ORG_ID

Organización	<u>Google LLC</u>
También conocido como	Google, YouTube (for Google Fiber see AS16591 record)
Nombre Completo	
Sitio Web de la Empresa	https://about.google/intl/en/
ASN	15169
IRR as-set/route-set	AS-GOOGLE
URL del servidor de ruta	
URL de Looking Glass	
Tipo de red	Content
Prefijos IPv4	15000
Prefijos IPv6	10000
Tasa de Tráfico	No divulgado
Tasa de tráfico	Mostly Outbound
Alcance Geográfico	Global
Protocolos compatibles	<input checked="" type="radio"/> IPv4 Unicast <input type="radio"/> Multicast <input checked="" type="radio"/> IPv6 <input type="radio"/> Never via route servers
Última actualización	2020-02-10T21:20:22Z
Public Peering Info Updated	2021-08-12T00:12:31
Peering Facility Info Updated	2021-05-27T16:49:33
Contact Info Updated	2020-01-22T04:24:08Z
Notas	<p>Peering Requests: https://isp.google.com/iwantpeering</p> <p>Peering Operational Issues: Contact noc@google.com 24x7</p> <p>We have a generally open peering policy: https://peering.google.com/#/options/peering</p> <p>This link also has information about our traffic delivery and management practices.</p> <p>Please note: not all Google content and services may be available at each PoP or Exchange.</p> <p>Google manages the following ASNs: AS36040, AS43515, AS36561, AS19527, AS139070, AS139190, AS16550</p>

Google LLC

También conocido como	
Nombre Completo	
Sitio web	http://www.google.com
Dirección 1	1600 Amphitheatre Parkway
Dirección 2	
Floor	
Suite	
Ubicación	Mountain View, California, 94043
Código de país	US
Geocodificación	No se pudieron obtener los datos de Geocodigo para esta entidad a este tiempo. Esto se hace automáticamente al hacer cambios a los campos de direcciones
Última actualización	2019-03-20 15:06:00
Notas	

Redes

Filtro

Nombre	ASN
Google Cloud Indonesia	139190
Google Cloud Korea	139070
Google Corporate Network in APAC	45566
Google IT	36385
Google LLC	15169
Google LLC AS19527	19527
Google LLC AS36040	36040
Google LLC AS43515	43515
Google Private Cloud	16550

SIBLINGS

Redes

Filtro

Nombre	ASN
Google Cloud Indonesia	139190
Google Cloud Korea	139070
Google Corporate Network in APAC	45566
Google IT	36385
Google LLC	15169
Google LLC AS19527	19527
Google LLC AS36040	36040
Google LLC AS43515	43515
Google Private Cloud	16550

Organización	Google LLC
También conocido como	Google, YouTube (for Google Fiber see AS16591 record)
Nombre Completo	
Sitio Web de la Empresa	https://about.google/intl/en/
ASN	15169
IRR as-set/route-set	AS-GOOGLE
URL del servidor de ruta	
URL de Looking Glass	
Tipo de red	Content
Prefijos IPv4	15000
Prefijos IPv6	10000
Tasa de Tráfico	No divulga
Tasa de tráfico	Mostly Outgoing
Alcance Geográfico	Global
Protocolos compatibles	IPv4, IPv6, BGP servers
Última actualización	2020-02-10T21:20:22Z
Public Peering Info Updated	2021-08-12T00:12:31
Peering Facility Info Updated	2021-05-27T16:49:33
Contact Info Updated	2020-01-22T04:24:08Z
Notas	<p>Peering Requests: https://isp.google.com/iwantpeering Peering Operational Issues: Contact noc@google.com 24x7</p> <p>We have a generally open peering policy: https://peering.google.com/#/options/peering This link also has information about our traffic delivery and management practices.</p> <p>Please note: not all Google content and services may be available at each PoP or Exchange.</p> <p>Google manages the following ASNs: AS36040, AS43515, AS36561, AS19527, AS139070, AS139190, AS16550</p>

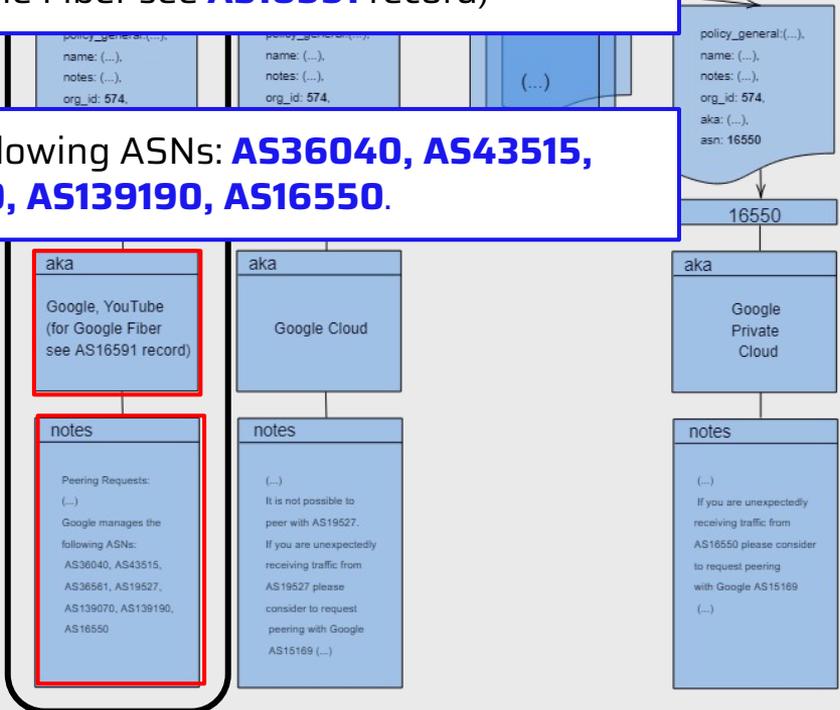
SIBLINGS

website: (...),
 city: (...),
 updated: (...),
 created: (...),
 id: **574**
 country": (...),
 (...)



Aka: Google, Youtube (for Google Fiber see **AS16591** record)

Notes: Google manages the following ASNs: **AS36040, AS43515, AS36561, AS19527, AS139070, AS139190, AS16550.**

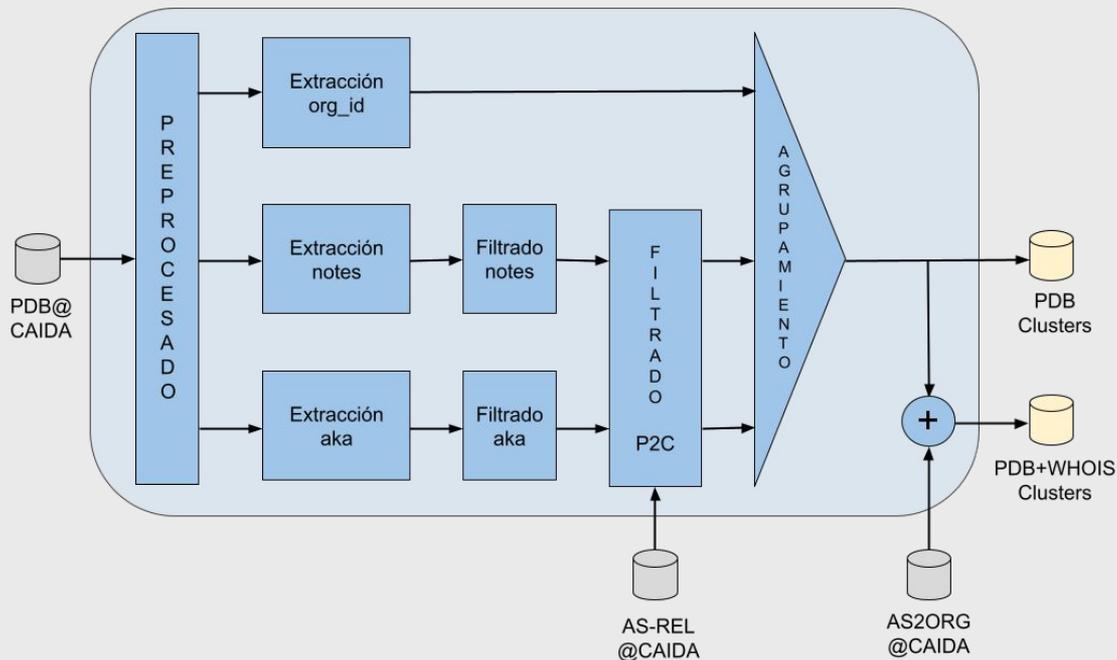


Metodología

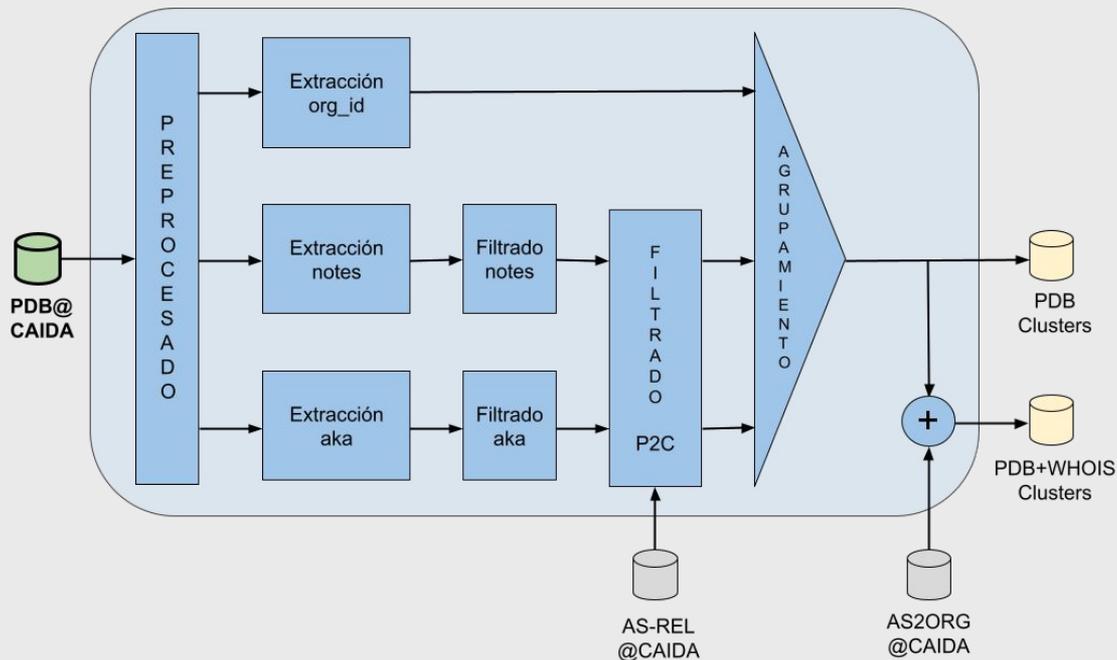


Convertir la información presente en [PeeringDB](#) en una fuente de datos estructurada que reporte los [siblings](#) de un [conglomerado](#).

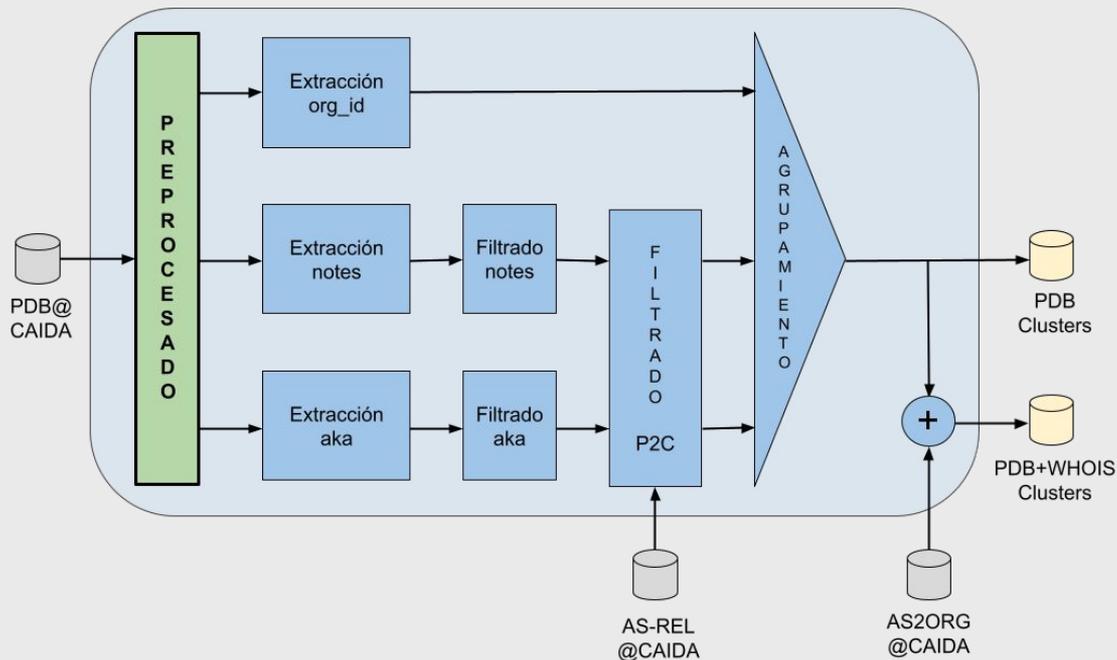
Metodología de inferencia de siblings



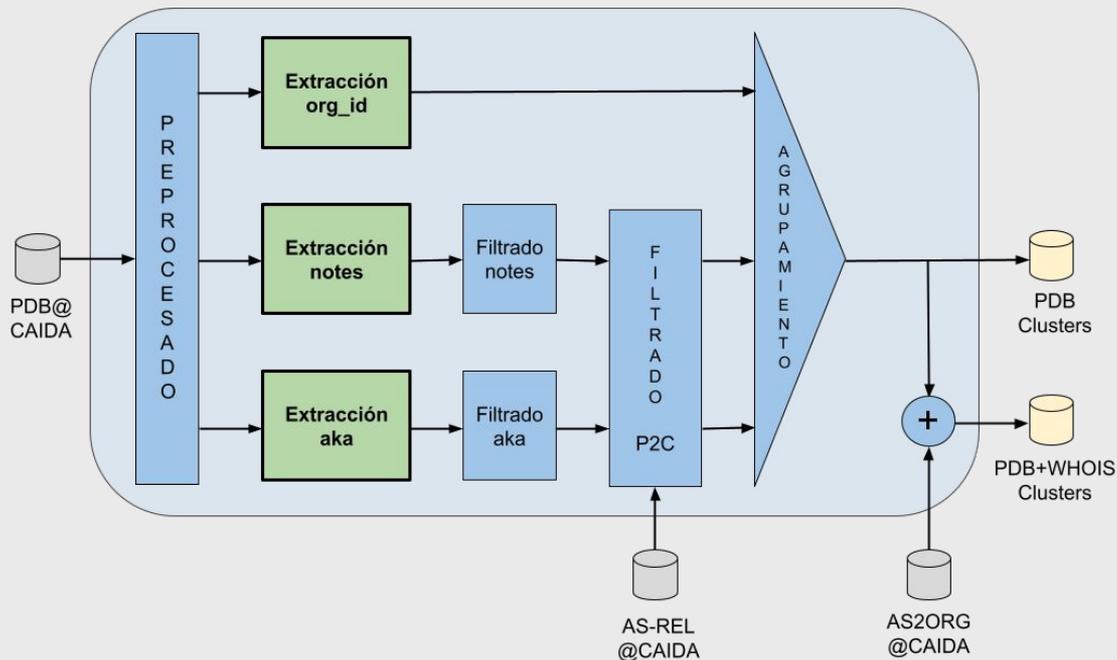
Metodología de inferencia de siblings



Metodología de inferencia de siblings



Metodología de inferencia de siblings



Extracción de Sistemas Autónomos



Números que potencialmente
sean ASNs en campos
de **texto libre**



notes y **aka**

¿ **Cómo lo hacemos ?**

ASNs reportados en
macroentidad



org_id

Regexes Notas



Directas

Indirectas

La **salida** de la extracción es un listado de ASNs

Abreviatura de Sistema Autónomo + Número

Expresión que antecede a uno o más ASes

AS, ASN, ASNS, ASS y ASES

Also manages
ASN behind
The following ASNS
Merging
We control
ASN under

Regexes Aka



Campo conciso y directo



\d{4,8}

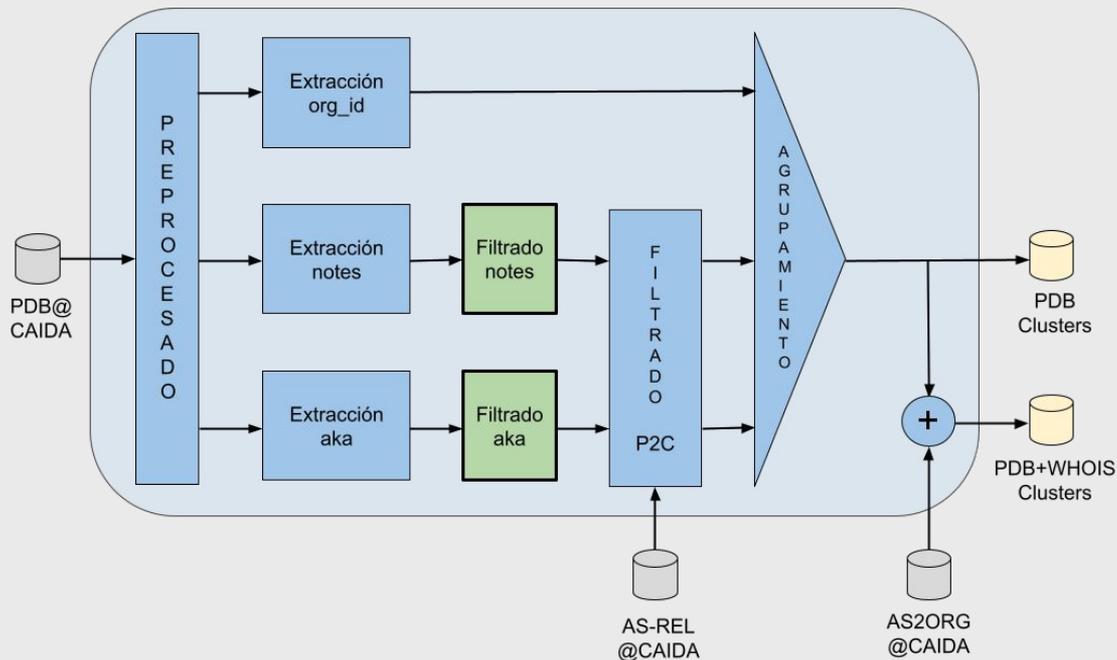
Entrada

```
Easynet (4589), MDNX (8190),  
NETCOM(5571), Solution1,  
CI-NET(8844), GRIFFIN(20500),  
Telecomplete/Fused(6320)
```

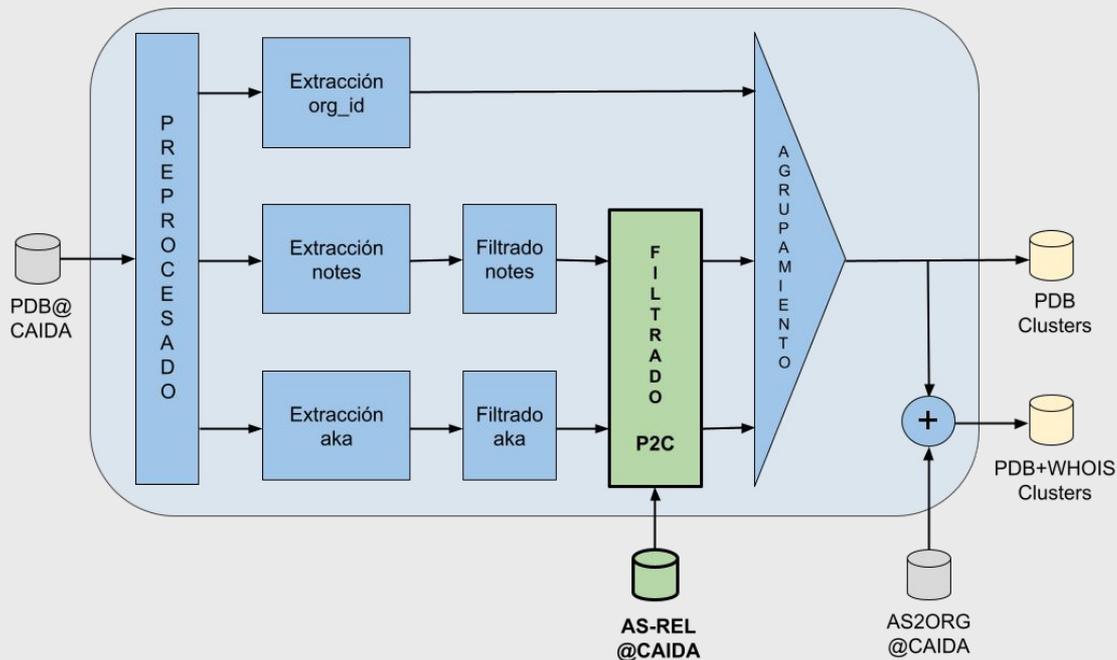
Salida

```
['4589', '8190', '5571',  
'8844', '20500', '6320']
```

Metodología de inferencia de siblings



Metodología de inferencia de siblings



Filtro p2c



La extracción genera un listado de ASNs

¿Son siblings?

¿Cómo sabemos qué relación tiene cada par de ASNs?

AS-RELATIONSHIPS [CAIDA]



ASN primario es aquel que fue registrado por el operador y del que se infiere el listado

Filtro p2c



Maxihost (AS262287)

asn: 262287
notes:(...)
aka:(...)

(...)
MAXIHOST OWNS A TIER 3 COMPLIANT
DATACENTER IN SÃO PAULO, WHERE ITS
HEADQUARTER IS LOCATED.
SEE MORE AT [HTTPS://WWW.MAXIHOST.COM/](https://www.maxihost.com/)
WE CONNECT DIRECTLY WITH THE FOLLOWING ISPS.
- ALGAR (AS16735)
- SPARKLE (AS6762)
- VOXILITY (AS3223)
- GTT (AS3257)
- COGENT (AS174)
- (...)

ASNs inferidos

174, 3223, 3257, 6762, 16735, 262287

ASNx

262287



PROV. CLI. PROV. PROV. PROV.

Telecom AR (AS7303)

asn: 7303
notes:(...)
aka:(...)

Telecom Argentina is the major broadband
and mobile provider in Argentina, with
more than 4.1 MM broadband subscribers,
4 MM fixed lines and 20 MM mobile lines.
Other ASN under 7303 are 10481
and 10318.

ASNs inferidos

7303, 10481, 10318

ASNx

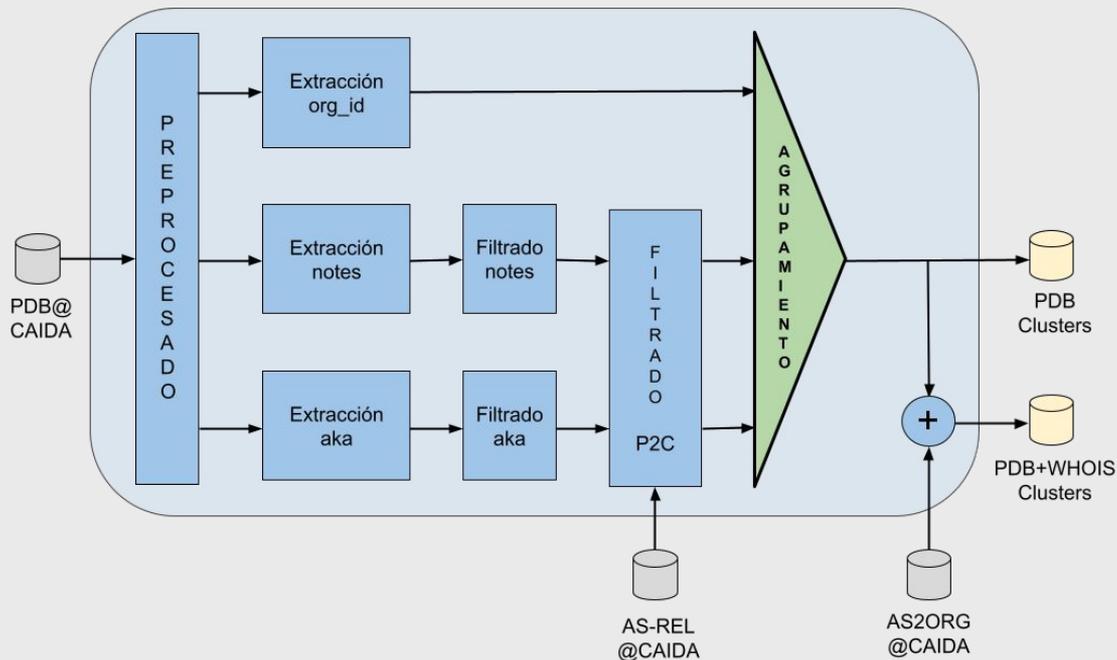
7303



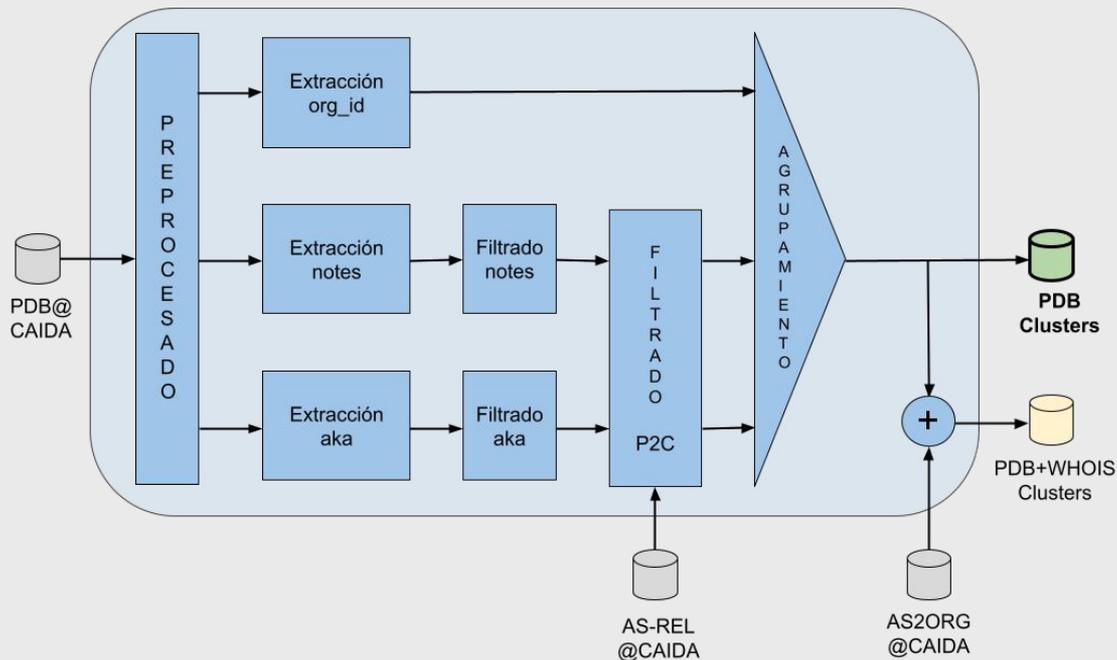
10318

Máximo un proveedor del
ASN primario

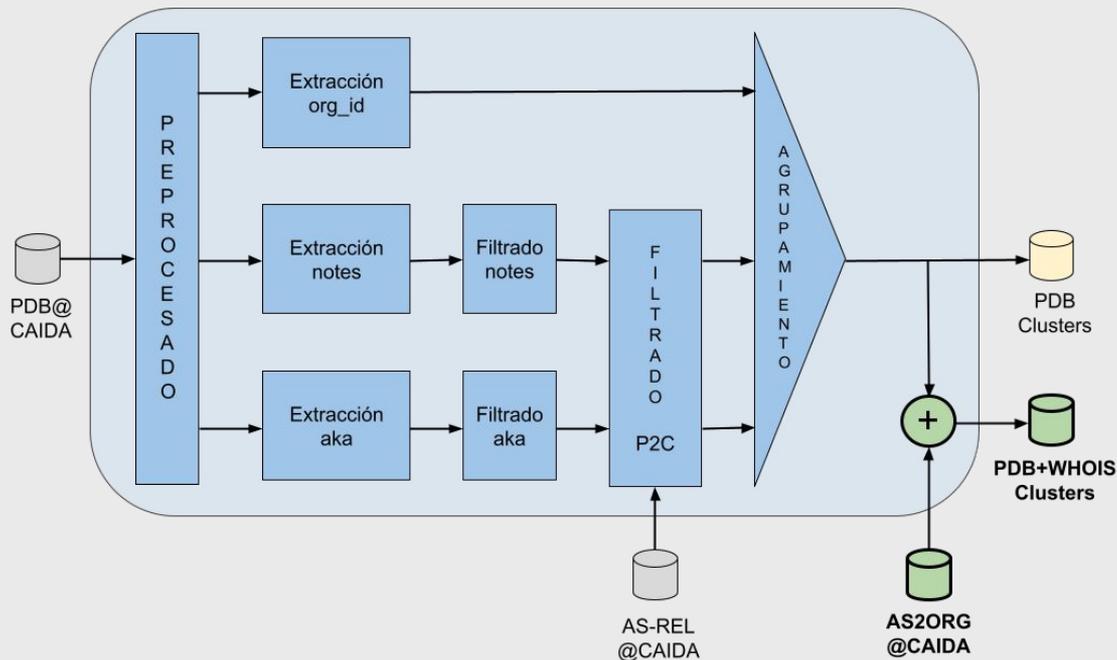
Metodología de inferencia de siblings



Metodología de inferencia de siblings



Metodología de inferencia de siblings



AS2ORG



WhoIs

AS2ORG+PDB

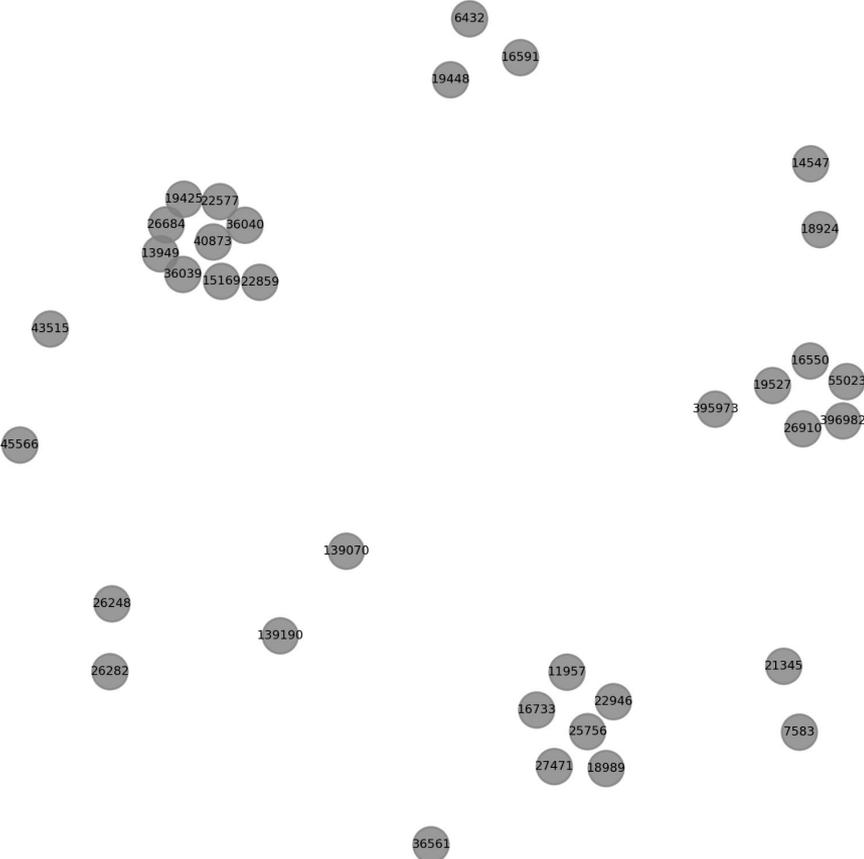


PeeringDB

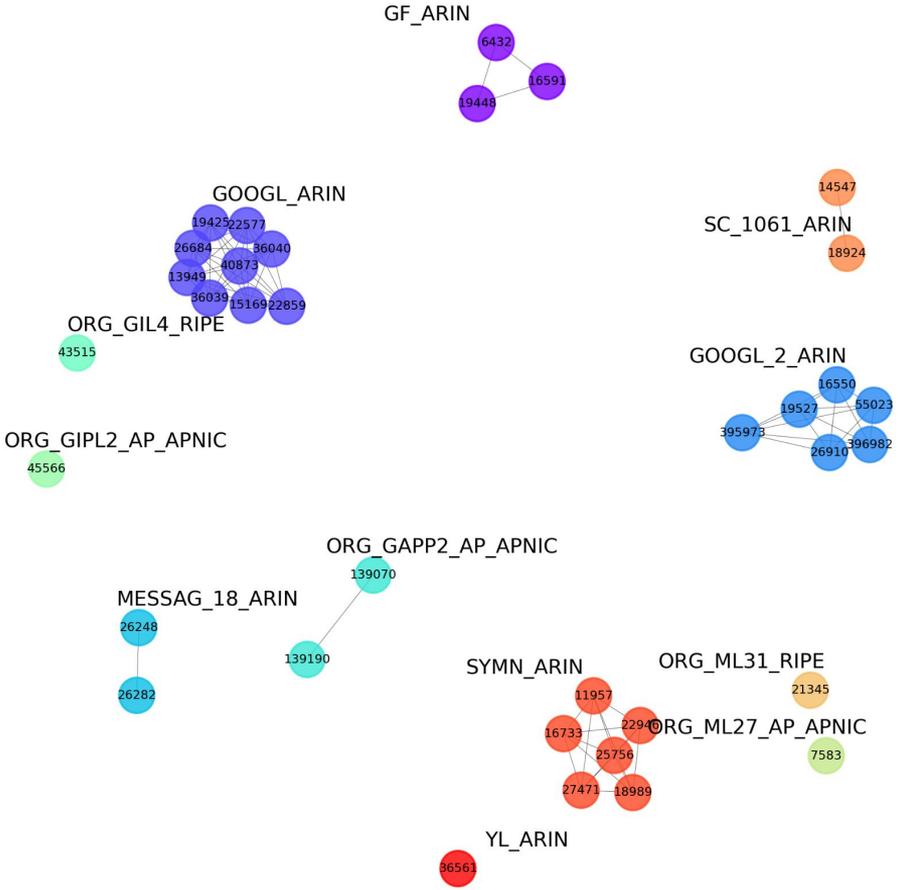
+

WhoIs

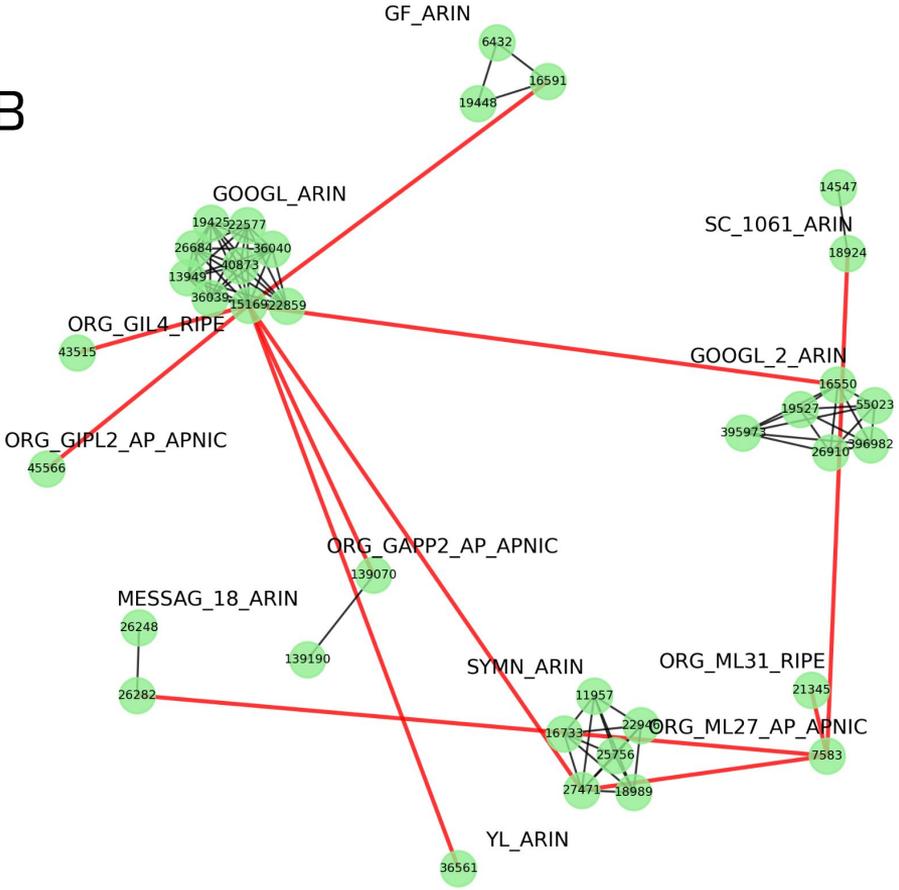
Nivel ASes



Nivel ASes con AS2ORG



Nivel ASes con AS2ORG+PDB





Resultados





Resultados de Metodologia con PeeringDB

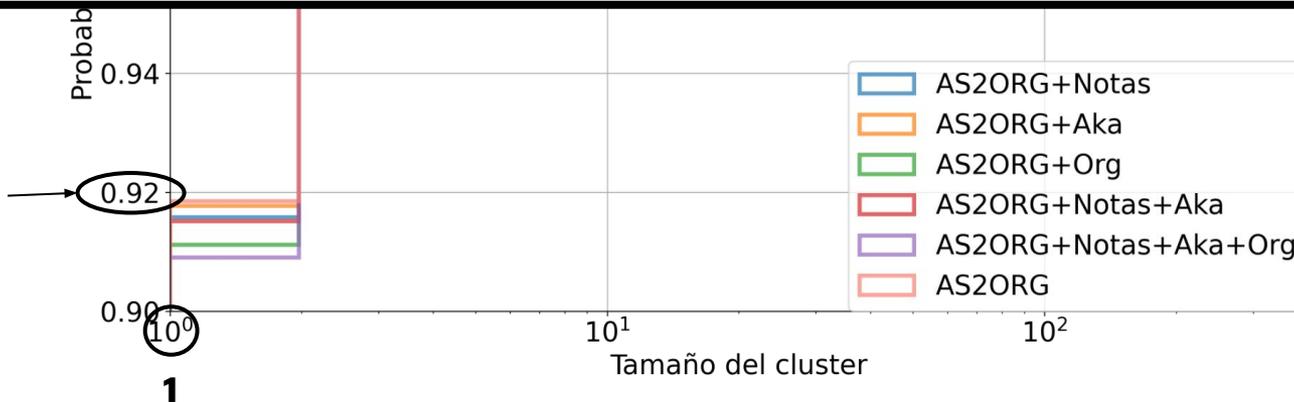
<i>Campo</i>	<i>Subconjuntos Extraidos</i>	<i>Agrupamientos</i>	<i>Cantidad de ASNs</i>
Notes	660	452	1048
Aka	226	212	361
Org_id	18.369	18.369	20.245
Notes+Aka	886	638	1364
Notes+Aka+Org_id	19.255	18.179	20.438

Metodologia mas completa

Resultados



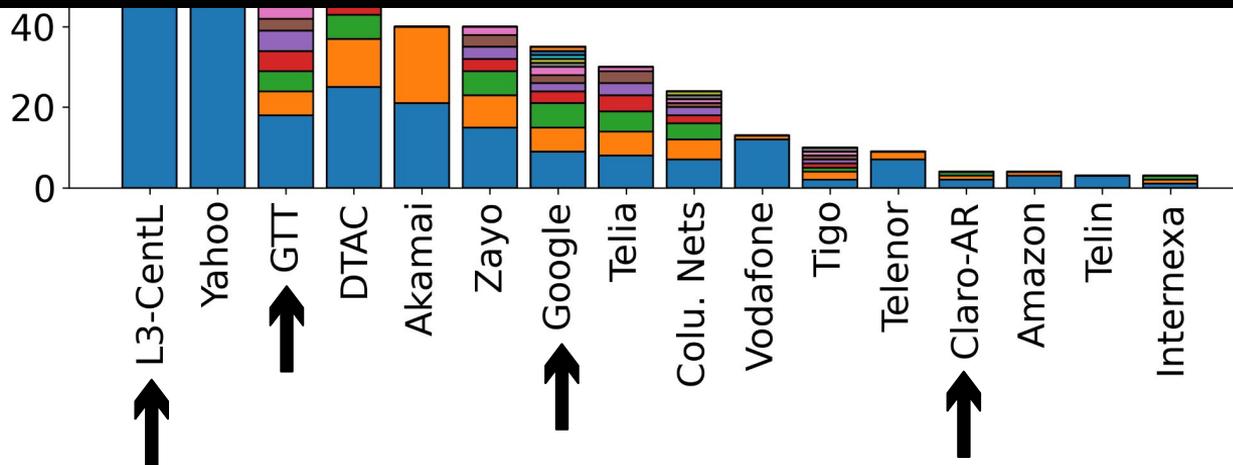
Impacto por sibling inferido > Cantidad



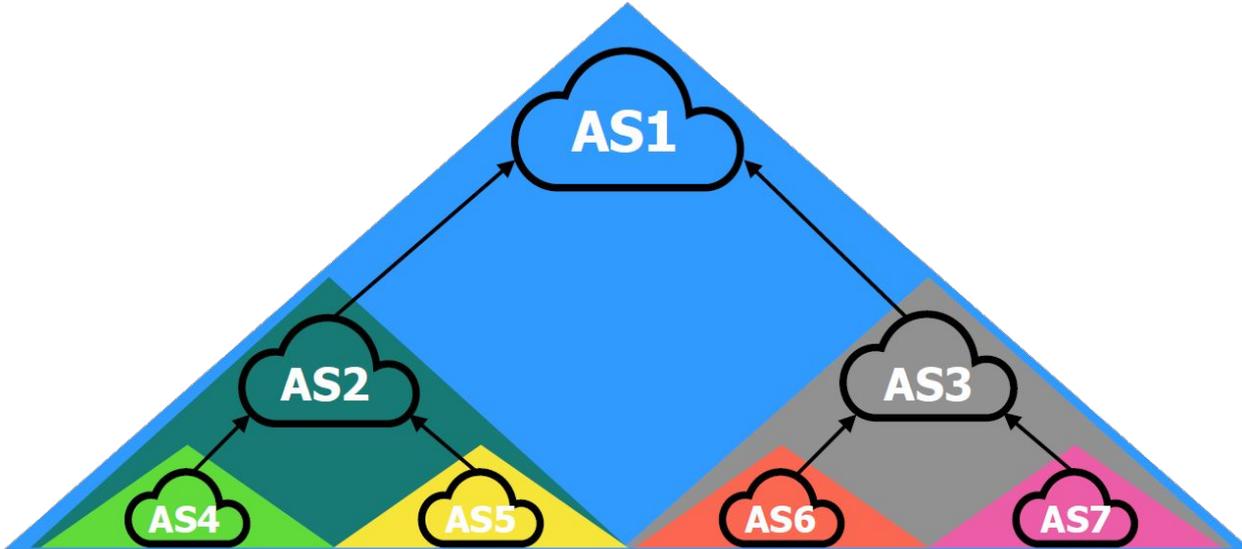
Resultados



La metodología es capaz de detectar **fusiones**



Resultados



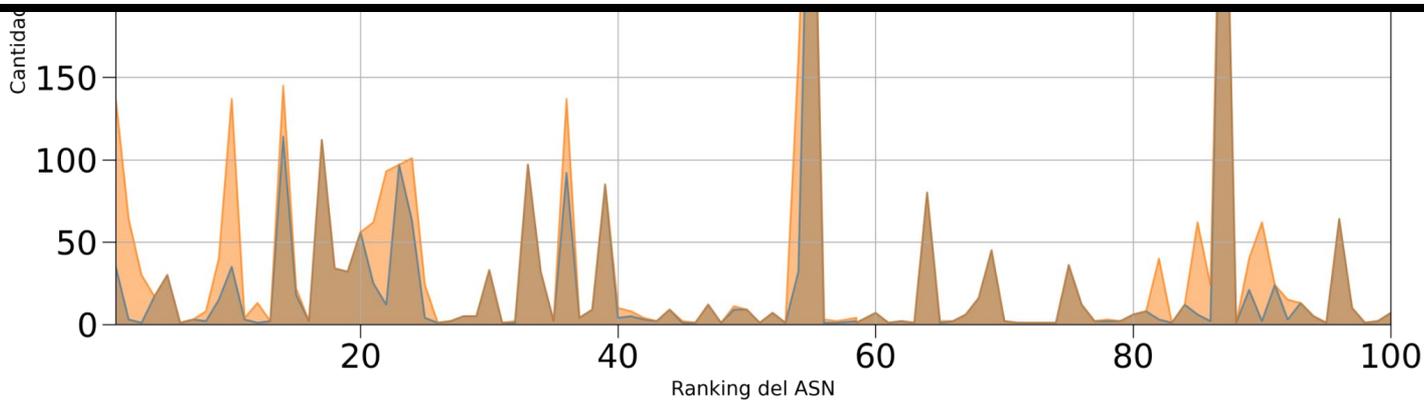
Resultados



Analisis por ranking segun ASRANK



Existe una **contribución** en el 33 % del top 100



Contribuciones



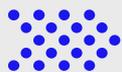
Las metodologías actuales son realmente incapaces de plasmar el tamaño de los conglomerados, existen falsos negativos.



Si bien PeeringDB cuenta con el [22.6%](#) de los ASNs delegados por los RIRs, su información **es representativa de la red**.



Combinar PeeringDB con las metodologías existentes es una mejor alternativa para representar a los conglomerados.



PDB es representativa para las **grandes redes** y **grandes tránsitos** como vimos en [ASRANK](#).

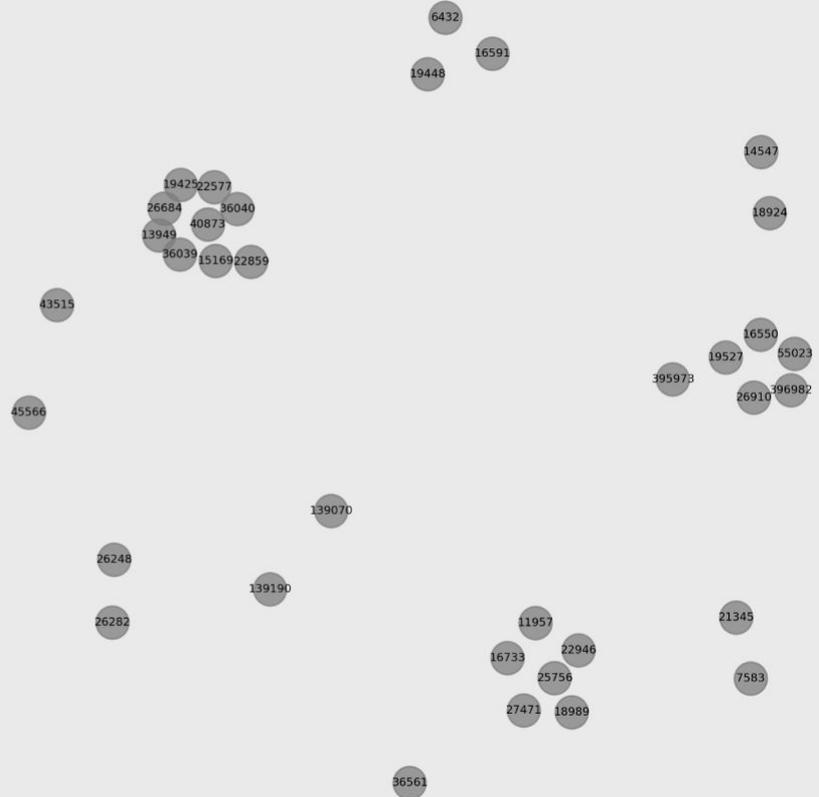
Tesis de Grado de Ingeniería en Informática

Estudio de PeeringDB para identificación y extracción de siblings

Alumno: Augusto Arturi
Director: Dr. Ing. Esteban Carisimo



Nivel de ASes





Gracias

