Network Analysis using SAS Visual Analytics

Falko Schulz

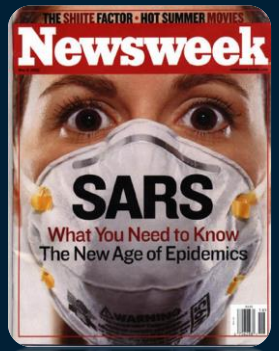
Distinguished Software Developer ▪ Business Intelligence Visualization R&D

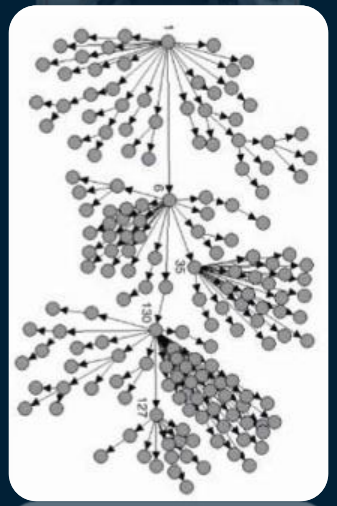
Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

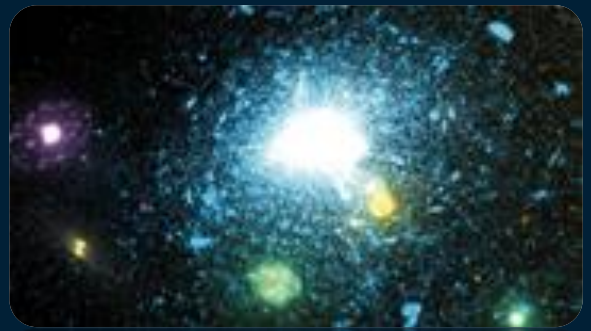
Agenda

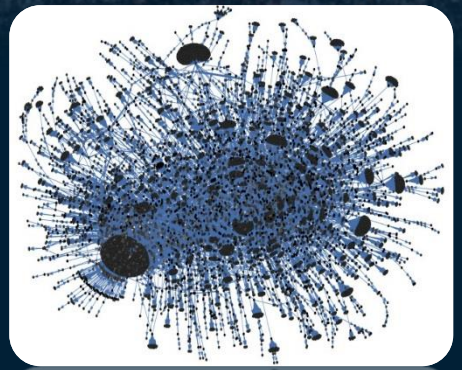
- Introduction
- Network analysis and visualization
- Data preparation
- Exploring networks in SAS Visual Analytics
- Examples
- Q&A

Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.









Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

Complex Systems

Synergy is the interaction of multiple elements in a system to produce an effect different from or greater than the sum of their individual effects.

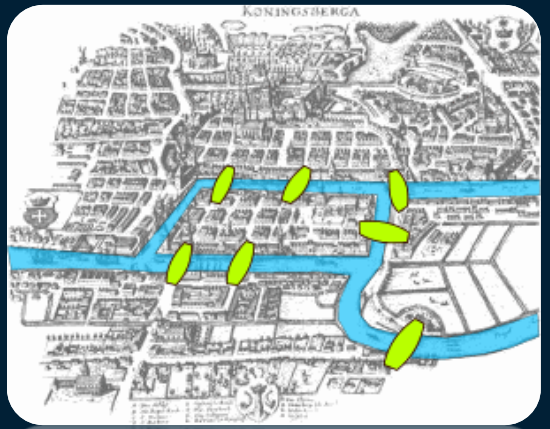
Emergence is the way complex systems and patterns arise out of a multiplicity of relatively simple interactions.

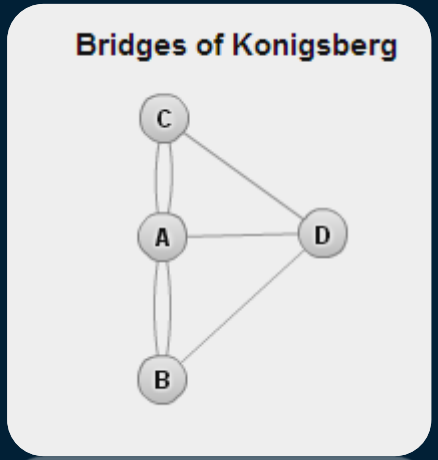
“The whole is greater than the sum of its parts.”

-- Aristotle

Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

Graph Theory

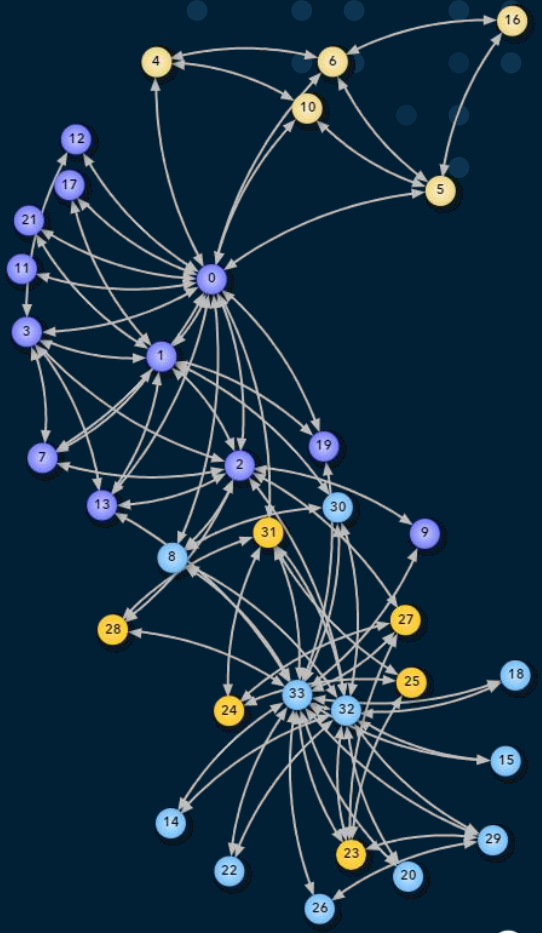




Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

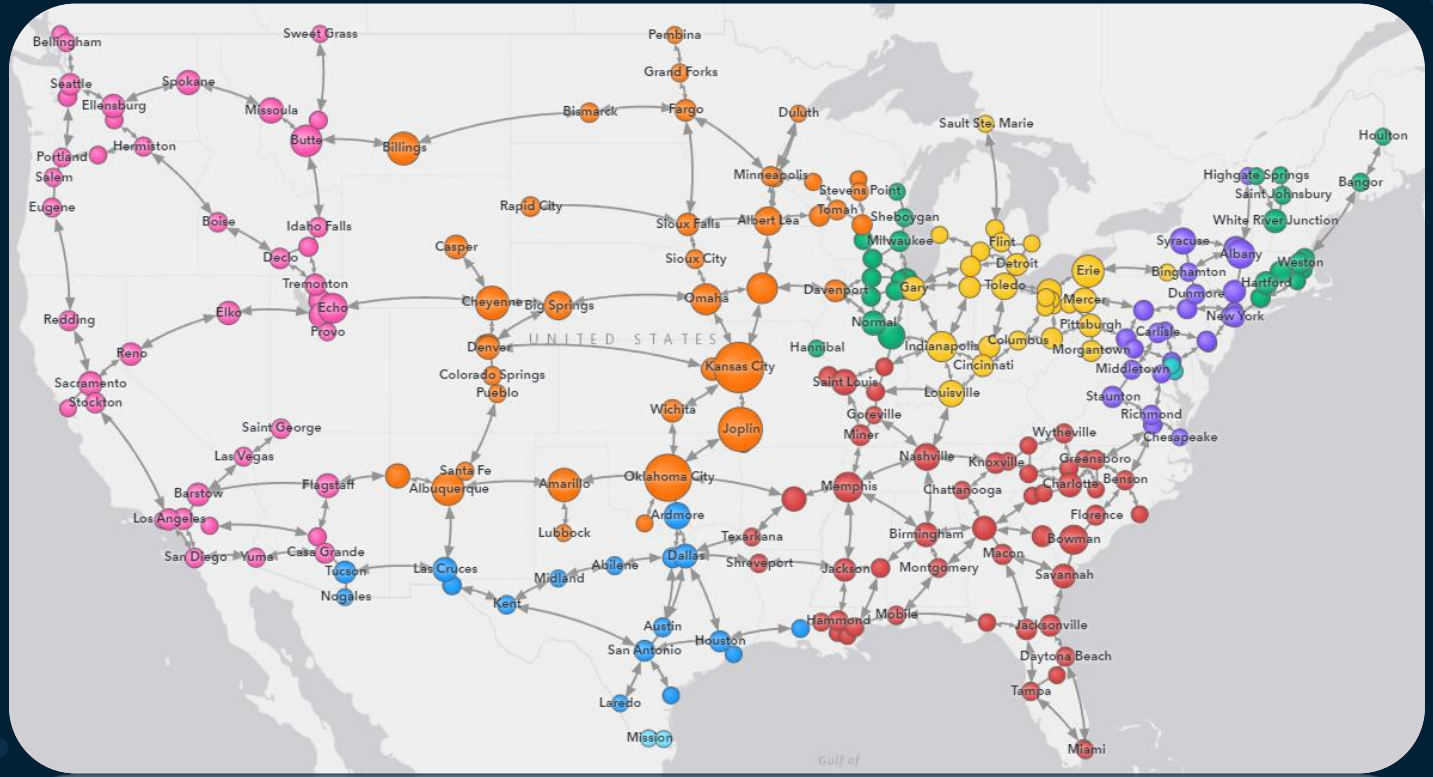
Network Visualization

- Nodes (*Vertices*)
  - Size, color
- Links (*Edges*)
  - Direction, width, color
- Layout
  - Dynamic (spring-force layout)
  - Mapped (geo overlay)



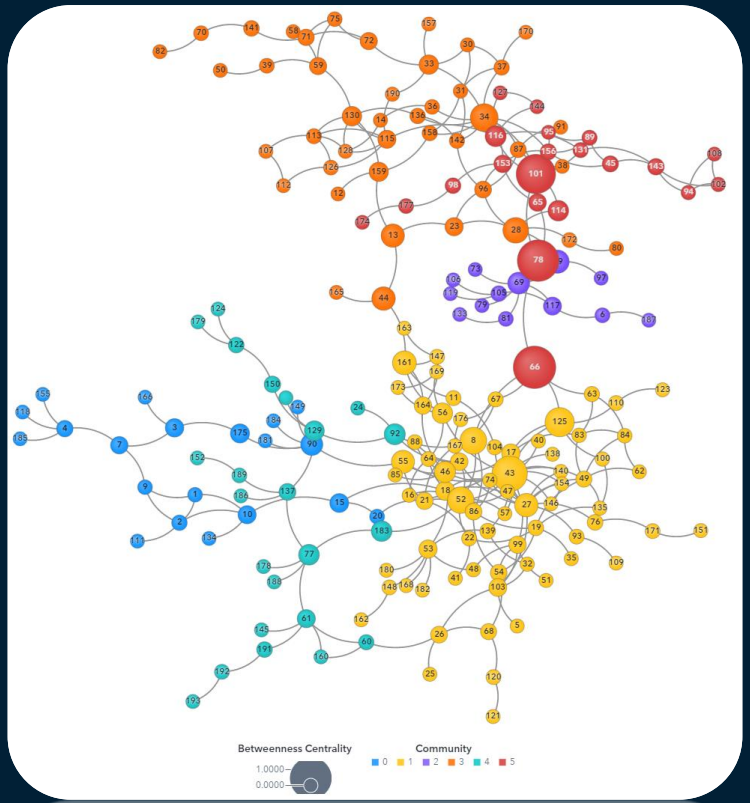
Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

Network Layout



Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

Network Layout



Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

Decorate network data

- Centrality measures
  - *Betweenness:* the number of shortest paths a node is on.
    - likely path for information flows
  - *Eigenvector centrality:* the centrality of a node's neighbors.
    - popular among popular actors.
  - *Degree:* number of connections.
    - number of friends, followers, etc.
- Community detection
  - Clustering

Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

Data Preparation

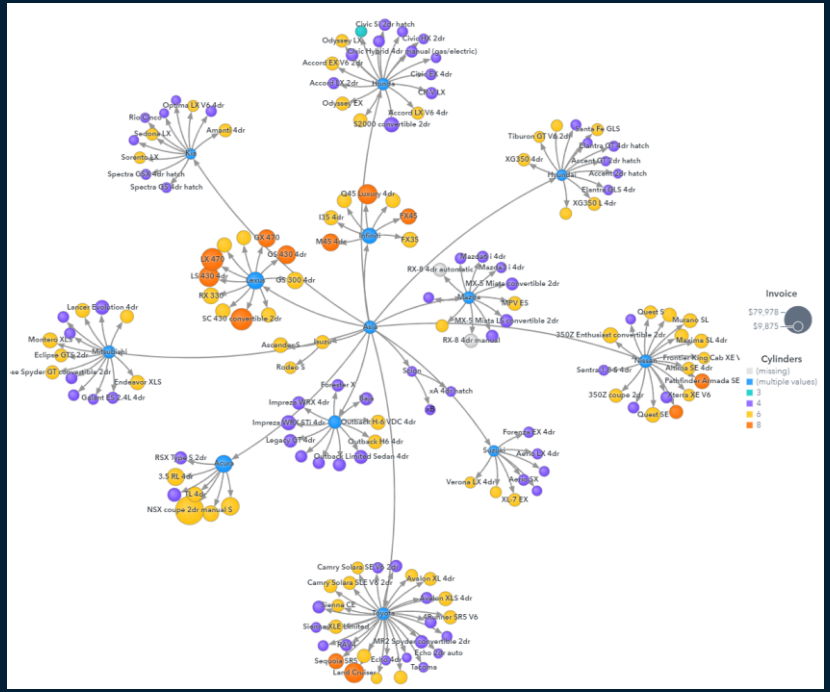
- SAS Visual Analytics supports two types of input data sources
  - Hierarchical network structure
  - Ungrouped network structure

Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

Hierarchical data structure

- Common data structure
- Creates a multi-level graph by using a hierarchy or a set of categories

ORIGIN	MAKE	MODEL	CYLINDERS	INVOICE	MSRP	HORSEPOWER
Asia	Honda	Civic LX 4dr	4	$14,531	$15,850	115
Asia	Honda	Civic Si 2dr hatch	4	$17,849	$19,490	160
Asia	Toyota	Land Cruiser	8	$47,986	$54,765	325
USA	Ford	Escape XLS	6	$20,907	$22,515	201

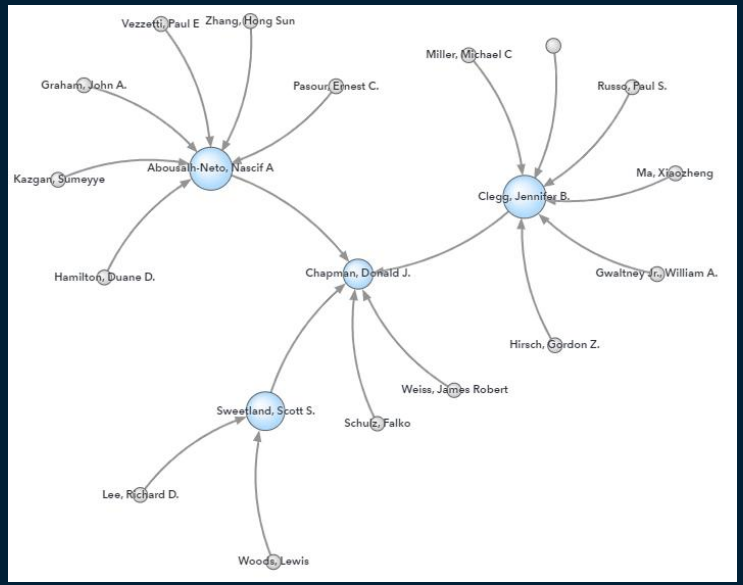


Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

Ungrouped data structure

- Used to represent traditional node-link data sets
- Requires one source (from) and one target (to) column

EMPLOYEE	MANAGER	SALARY	STARTDT
EMP1	MGR1	40000	17-Sep-02
EMP2	MGR2	67000	12-Mar-01
EMP3	MGR2	55000	04-Apr-04
MGR1	MGR2	85000	23-Jun-96
MGR2		98000	5-May-95



Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

Demo

sas.com

Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

BRISBANE

Optimal City Tour using SAS Optimization

All In Brewing CO
Aether Brewing
The Village Social
Stafford Tavern
Moreton Island
The Popular Beer
Wine & Lager Bistro
Cannon Hill Tavern
Shucks Bar
Gallery of Modern Art
Regatta Hotel
Three Brews
Slipstrem Brewing Co
Lone Pine Koala Sanctuary
Ballistic Beer Co
Bacchus Brewing Co
White Lies Brewing
Jacobs Creek Wines
Sirromet Wines
Tap'd Bar Ipswich

Distance...
224

Duration
16:47

Stop (1) or Place (2)
2
1

Distance (km)
8.86
0.02

Transportation Type
Bus
Train/Tram
Walking

Places
31

Stops
359

POWERED BY esri

Network Analysis and Network Optimization in SAS® Viya®

Copyright © SAS Institute Inc. All rights reserved.

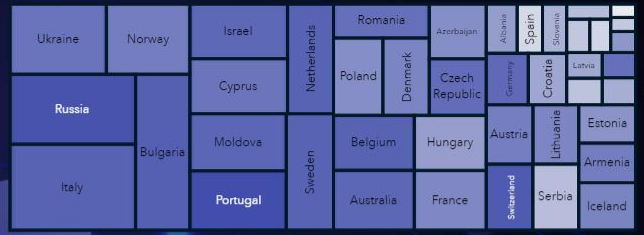
EUROVISION SONG CONTEST

POLITICAL OR GEOGRAPHICAL VOTING

Data: Janne Spijkervet, Eurovision Dataset, https://github.com/Spijkervet/eurovision-dataset
Design: Falko.Schulz@sas.com, 2021
Product: SAS Visual Analytics on SAS Viya
Last updated: Feb 01, 2021

The nearly 60-year-old song contest is watched by an estimated 125 million viewers each year. Started in 1956, it has been held very year (with the exception of 2020) - making it the longest-running annual international televised music competition.

All time favourites



Points
744
23

Rank (Avg/Finals)
25.00
2.16

Vote Category
Jury
Televoting

Edition
Finals
Semi-Finals
Semi-Finals 1
Semi-Finals 2

Contest Year
2009 to 2019
1975 2019

Points given
10 to 12
1 12

Number of years 12 points given to a country

4

Greece to Cyprus

Number of years a country has been in the top 5 ranking

3

Sweden

Most points given to a country

48

Greece to Cyprus

Best average ranked country

5.1

Italy

Highest scored country

744

Italy

Lowest scored country

23

Finland

Frequency
8
1

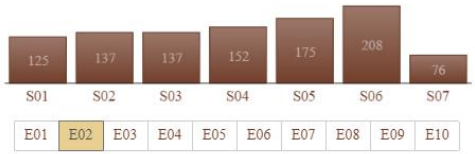
Points
96
10

SAS Viya

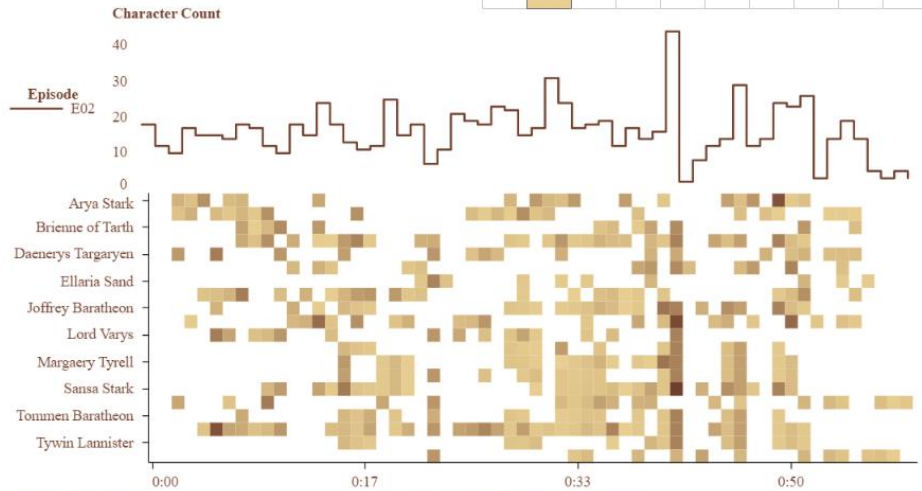
GAME OF THRONES

Data: https://github.com/jeffreylancaster/game-of-thrones
Last updated: March, 2019
Design: Falko.Schulz@sas.com

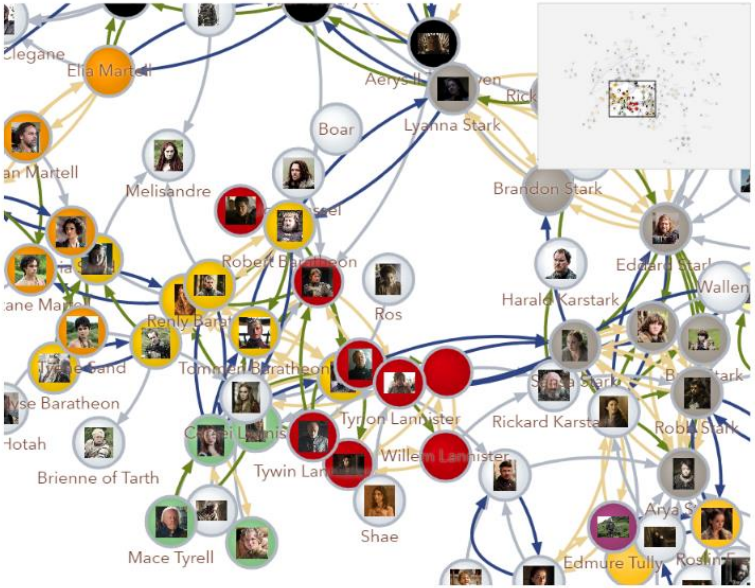
Character Count by Season



Character Count by Scene Start (min) grouped by Episode

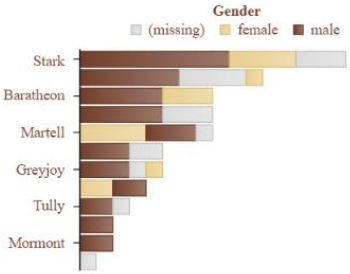


Social Network

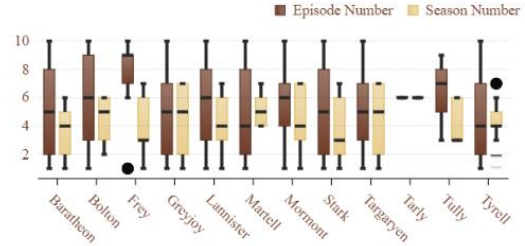


SAS Viya

House Members



Screentime by Episode and Season

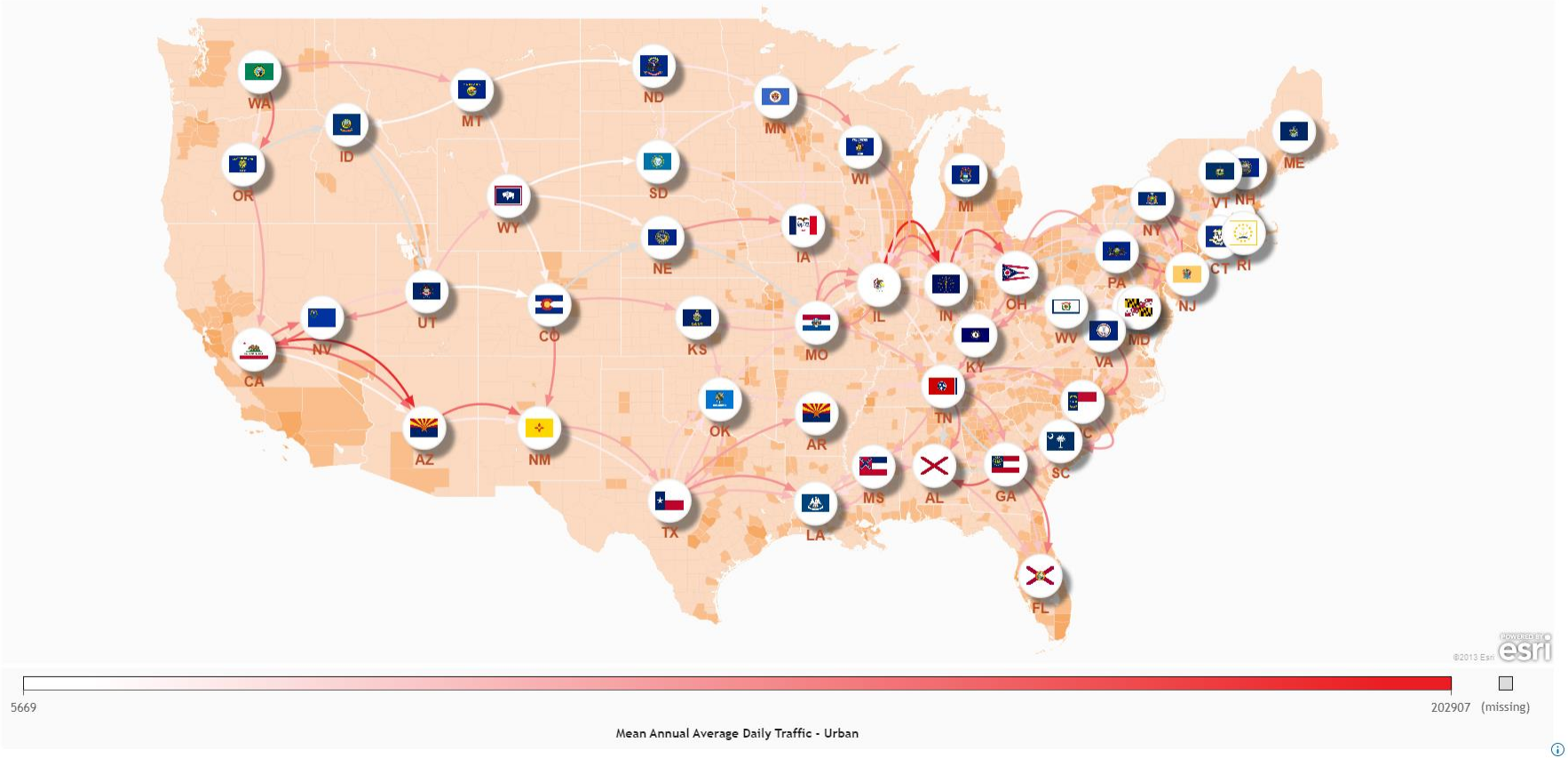


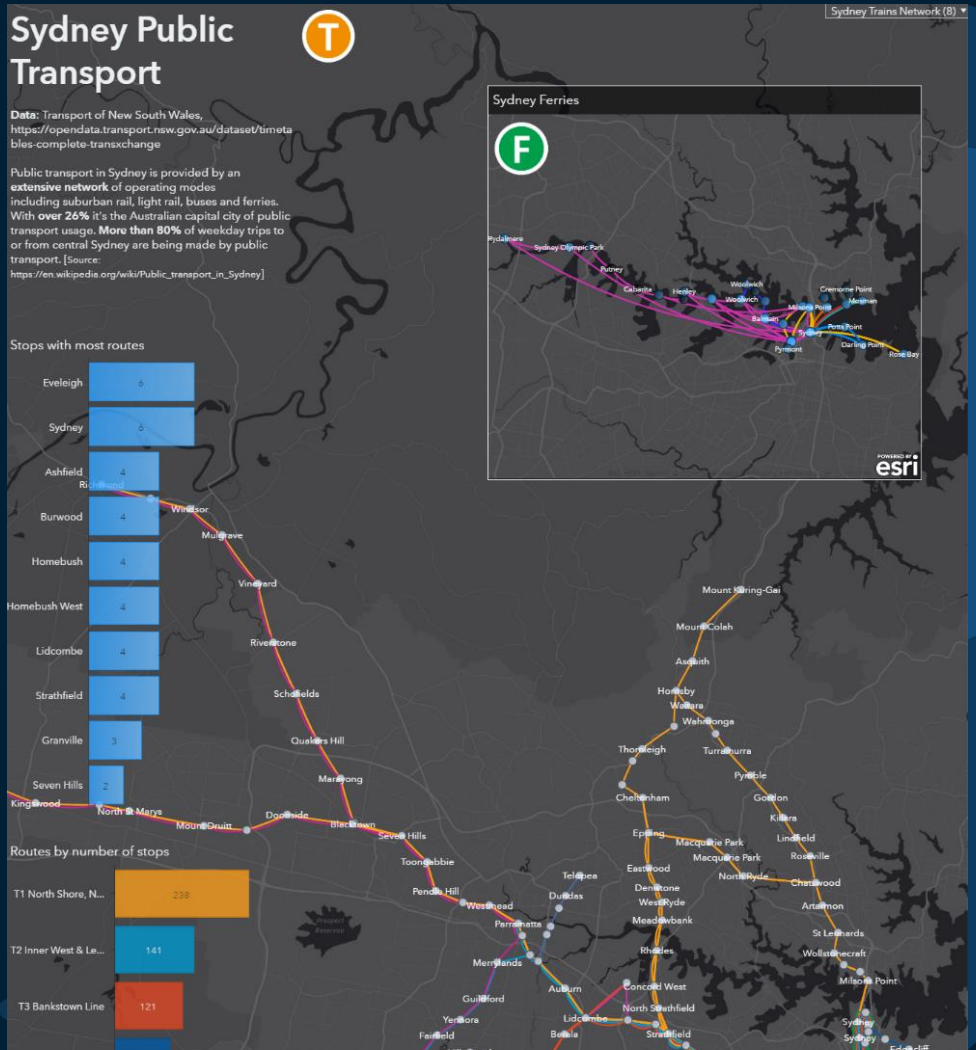
Screentime by Character

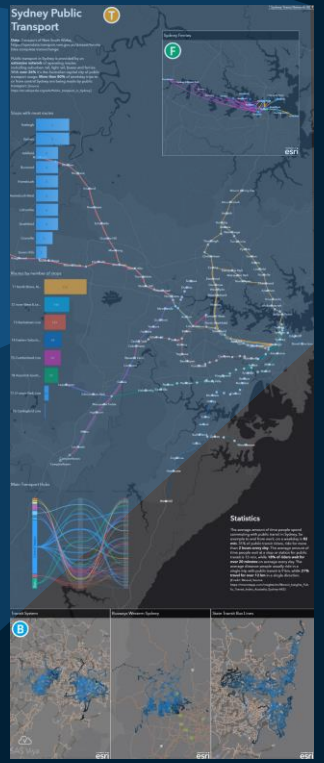
Copyright © SAS Institute Inc. All rights reserved.

Interstates Network

United States Interstates Network




Sydney Public Transport
Sydney Trains Network (8)
Data: Transport of New South Wales, https://opendata.transport.nsw.gov.au/dataset/timetables-complete-transxchange
Public transport in Sydney is provided by an extensive network of operating modes including suburban rail, light rail, buses and ferries. With over 26% it's the Australian capital city of public transport usage. More than 80% of weekday trips to or from central Sydney are being made by public transport. [Source: https://en.wikipedia.org/wiki/Public_transport_in_Sydney]
Sydney Ferries
Stops with most routes
Eveleigh 6
Sydney 6
Ashfield 4
Burwood 4
Homebush 4
Homebush West 4
Lidcombe 4
Strathfield 4
Granville 3
Seven Hills 2
Routes by number of stops
T1 North Shore, N... 238
T2 Inner West & Le... 141
T3 Bankstown Line 121
Richmond
Windsor
Mulgrave
Vineyard
Riverstone
Schofields
Quakers Hill
Marayong
Kingswood
North St Marys
Mount Druitt
Doonside
Blacktown
Seven Hills
Toongabbie
Pendle Hill
Westmead
Parramatta
Merrylands
Guildford
Yenora
Fairfield
Telopea
Dundas
Auburn
Lidcombe
Berala
Mount Kuring-Gai
Mount Colah
Asquith
Hornsby
Wallara
Wahroonga
Thornleigh
Turramurra
Pymble
Cheltenham
Gordon
Killara
Lindfield
Epping
Macquarie Park
Macquarie Park
Rosevile
Eastwood
North Ryde
Chatswood
Denistone
West Ryde
Meadowbank
Artarmon
St Leonards
Rhodes
Wollstonecraft
Concord West
Milsons Point
North Strathfield
Strathfield
Sydney
Sydney
Edgecliff
Pyrmont
Darling Point
Rose Bay
Sydney Olympic Park
Putney
Cabarita
Henley
Woolwich
Woolwich
Balmain
Milsons Point
Cremorne Point
Neutral Bay
Potts Point
Sydney
esri

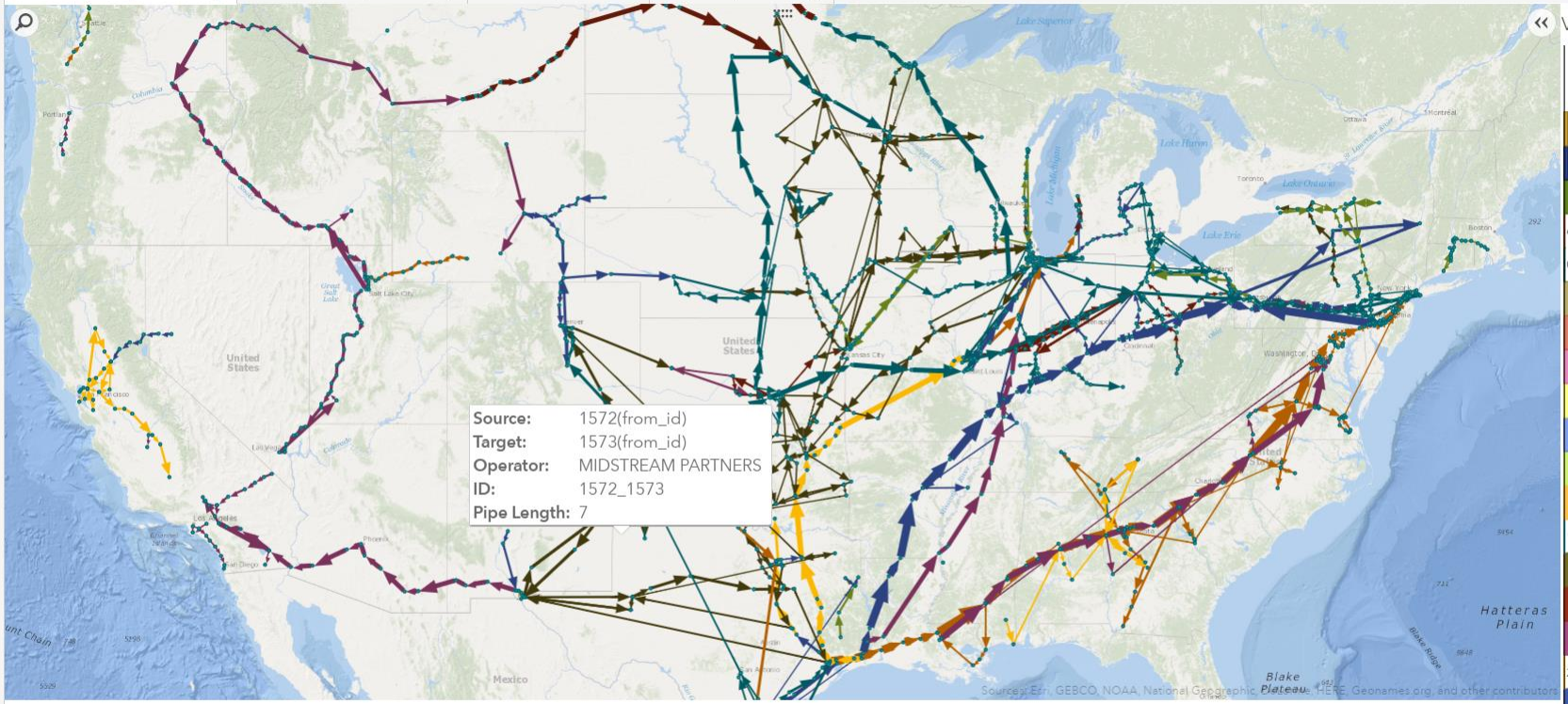

Sydney Public Transport
Statistics
esri

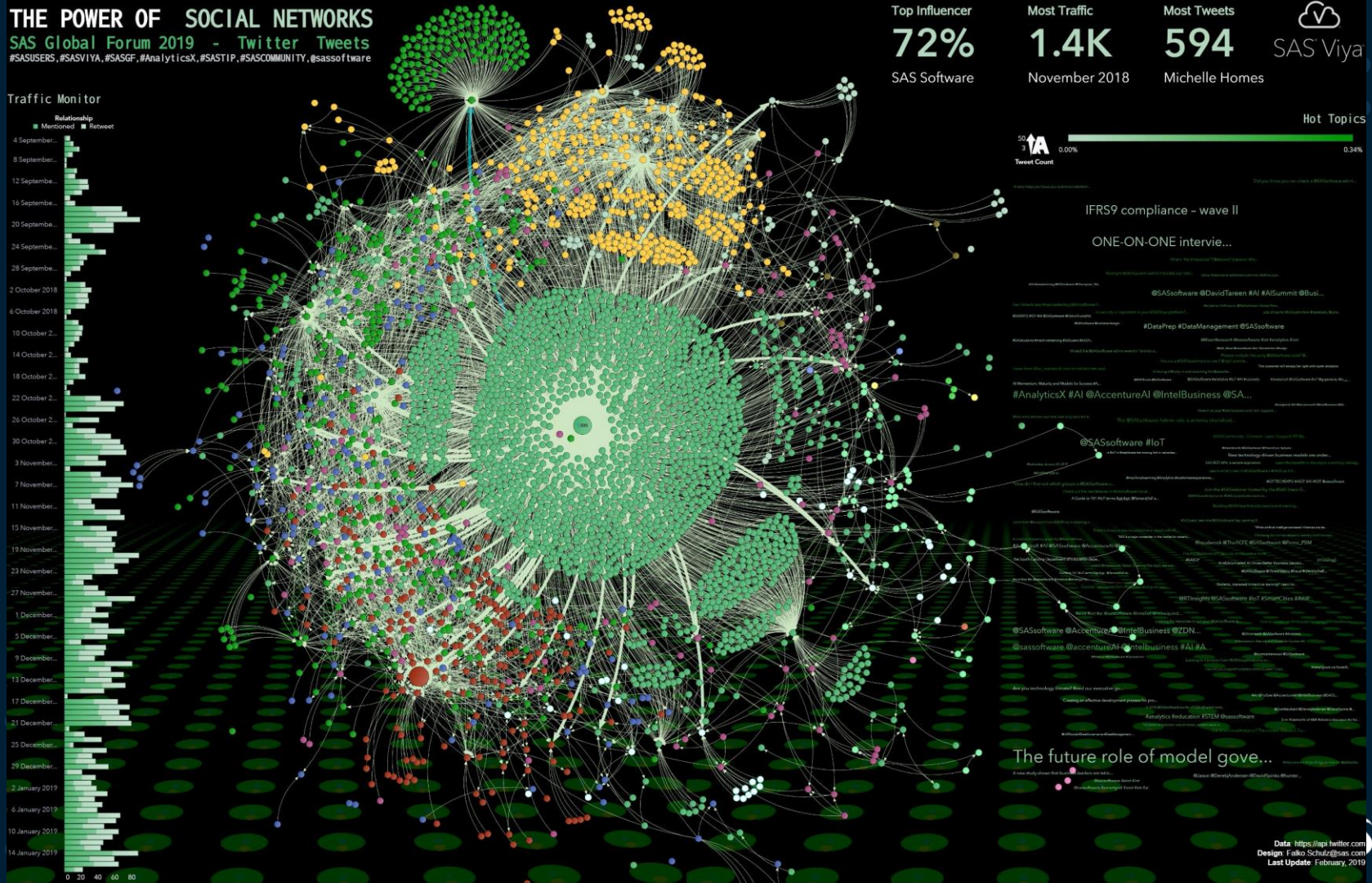
US Energy Resources

US Petroleum Pipelines | US Crude Oil Pipelines | US Powerplants | NC Powerplants | +

Data | Objects | Outline

Options | Roles | Actions | Rules | Filters | Ranks




THE POWER OF SOCIAL NETWORKS
SAS Global Forum 2019 - Twitter Tweets
#SASUSERS,#SASVIYA,#SASGF,#AnalyticsX,#SASTIP,#SASCOMMUNITY,@sassoftware
Top Influencer
72%
SAS Software
Most Traffic
1.4K
November 2018
Most Tweets
594
Michelle Homes
SAS Viya
Traffic Monitor
Relationship
Mentioned
Retweet
4 September...
8 September...
12 Septembe...
16 Septembe...
20 Septembe...
24 Septembe...
28 Septembe...
2 October 2018
6 October 2018
10 October 2...
14 October 2...
18 October 2...
22 October 2...
26 October 2...
30 October 2...
3 November...
7 November...
11 November...
15 November...
19 November...
23 November...
27 November...
1 December...
5 December...
9 December...
13 December...
17 December...
21 December...
25 December...
29 December...
2 January 2019
6 January 2019
10 January 2019
14 January 2019
0 20 40 60 80
Hot Topics
Tweet Count
0.00%
0.34%
IFRS9 compliance – wave II
ONE-ON-ONE intervie...
@SASsoftware @DavidTareen #AI #AISummit @Busi...
#DataPrep #DataManagement @SASsoftware
#AnalyticsX #AI @AccentureAI @IntelBusiness @SA...
@SASsoftware #IoT
@SASsoftware @Accenture... @IntelBusiness @ZDN...
@sassoftware @accentureAI @intelbusiness #AI #A...
Analytics #education #STEM @sassoftware
The future role of model gove...
Data: https://api.twitter.com
Design: Falko.Schulz@sas.com
Last Update: February, 2019
sas

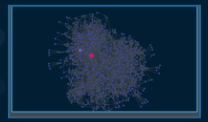
COVID-19 Scientific Literature Search and Text Analysis

Explore the interactive co-citation network

SAS Viya

Select an authority group for the top 30 most populated communities, identified by the top three topics in that community, to see an associated network of co-citations. Zoom in on the top percentage of authoritative papers below. Each node in the network represents a paper or reference. The color of a node represents a paper's authority score. A higher value here means higher importance in the network based on a number of references from direct linked papers as well as their references.

Co-citation network for topic **"Respiratory Exacerbation, Respiratory Tract Infection or Patient Care and Mortality"**



Centrality: Authority

100.00%

0.00%

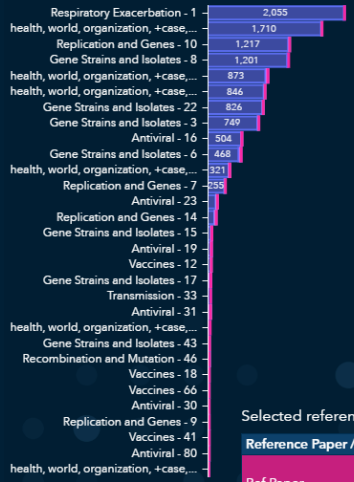
Centrality: Authority (Category)

<30% 30%-49% 100%

Frequen...

1

Top 30 most populated communities



Authority Group

<30%
30%-49%
50%-79%
80%-89%
90%-99%
100%

Authoritativeness

0% to 100%

0% 100%

Selected reference paper and first degree citations

Reference Paper / Cited By	Centrality: Author... ▼	Title	Authors	Publish Tin
Ref Paper	100.0...	Identification of a new human coronavirus	Lia van der Hoek,Krzysztof Pyrc,Maarten Jebbink,Wilma Vermeulen-Oost,Ron Berkhout,Katja Wolthers,Pauline Wertheim-van Dillen,Jos Kaandorp,Joke Spaargaren,Ben Berkhout	2004-03-2
Cited by	35.45%	Croup Is Associated with the Novel Coronavirus NL63	Lia Van Der Hoek,Klaus Sure,Gabriele Ihorst,Alexander Stang,Krzysztof Pyrc,Maarten Jebbink,Gudula Petersen,Johannes Forster,Ben Berkhout,Klaus Berla	2005-08-2
Cited by	27.15%	A novel pancoronavirus RT-PCR assay: frequent detection of human coronavirus NL63 in children hospitalized with respiratory tract infections in Belgium	Elien Moës,Leen Vijgen,Els Keyaerts,Kalina Zlateva,Sandra Li,Piet Maes,Krzysztof Pyrc,Ben Berkhout,Lia van der Hoek,Marc Van Ranst	2005-02-0
Cited by	25.16%	Identification of a Novel Polyomavirus from Patients with Acute Respiratory Tract Infections	Anne Gaynor,Michael Nissen,David Whiley,Ian Mackay,Stephen Lambert,Guang Wu,Daniel Brennan,Gregory Storch,Theo Sloots,David Wang	2007-05-0

Double click on a paper topic shows more detailed information about the selected community. Double click on a paper searches the data base for related papers based on the papers title.

Last updated: Thursday, May 28, 2020 02:40:23 AM