



Proc network for Network Analytics/Graph Theory

Hans de Wit, Senior Data Scientist, Telenor Norway

Hans de Wit



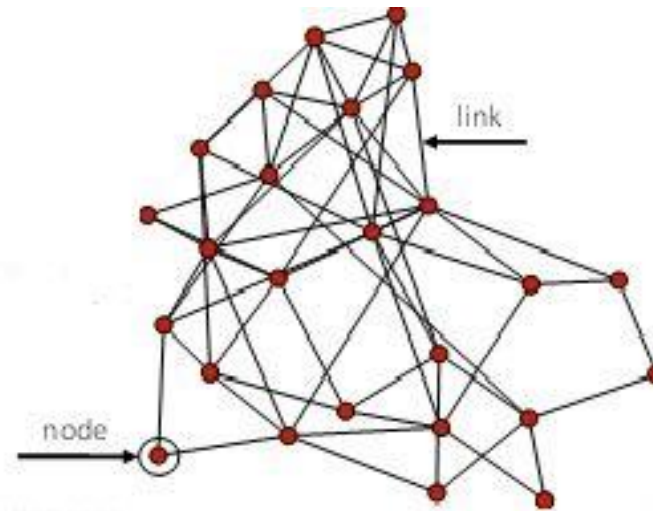
- **Telenor Mobile/IT Norway (since 2013)**
 - Advanced Analytics & Data Science Manager
- **ING Bank, The Netherlands**
 - Senior member 'model'/Innovation-team ING Retail Customer Intelligence
 - Member analytical campaign management ING Bank Customer Intelligence department, 1997-2005
- **ING Card, 2005-2008**
 - Direct Marketing, Credit Risk, Fraud
 - Master of Marketing (SRM) and bachelor of Commercial economics and Direct Marketing.

- **My passion:**
- **Making the unreal happen**



What is Network analytics (proc network) /Graph Theory

- Network analysis focuses primarily on analyzing relationships and patterns with a network, instead of on individual attributes
- To model these patterns and relationships, a single network consists of two fundamental components:
 - A set of Nodes(N)
 - A set of links(L) connecting the nodes
 - This network contains 25 nodes and 61 links



Network analytics/graph theory will become important in the (near) future. Invest your time in this technology!

NEWS

Gartner predicts exponential growth of graph technology

As both organizations and their databases struggle to manage growing amounts of data, experts see dramatic growth in the use of graph databases by 2025.



By Eric Avidon, News Writer

Published: 05 Oct 21

NEW YORK DAILY PRESS

BUSINESS CULTURE NEWS SPORTS TECH



OCTOBER 5 2021
Gartner predicts exponential progress of graph expertise
NEW YORK DAILY PRESS · TECH

As the dimensions and complexity of knowledge continues to escalate, organizations will more and more flip to graph expertise as a method of harnessing their information to drive decision-making.

Through the keynote handle on Oct. 5 of Graph + AI Summit Fall 2021, a hybrid in-person and digital convention hosted by graph database vendor TigerGraph, Gartner analyst Rita Sallam stated the analysis and advisory agency forecasts that 80% of knowledge and analytics improvements shall be made utilizing graph expertise by 2025.



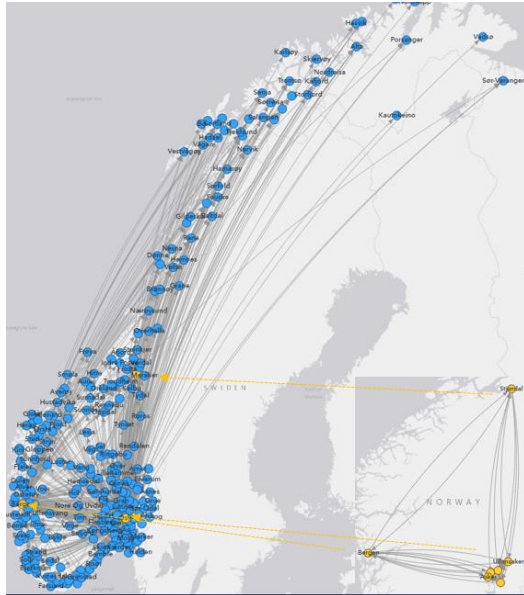
History of Graph theory. Seven Bridges of Königsberg

- By Leonhard Euler in 1736.
- The city of [Königsberg](#) in [Prussia](#) (now [Kaliningrad](#), [Russia](#)) was set on both sides of the [Pregel River](#), and included two large islands—[Kneiphof](#) and [Lomse](#)—which were connected to each other, or to the two mainland portions of the city, by seven bridges. The problem was to devise a walk through the city that would cross each of those bridges once and only once.

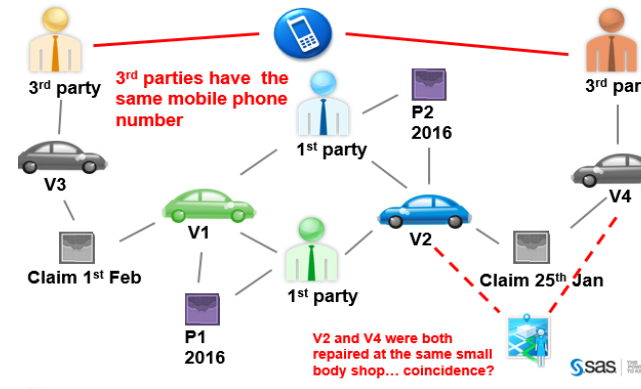


Network analytics use cases

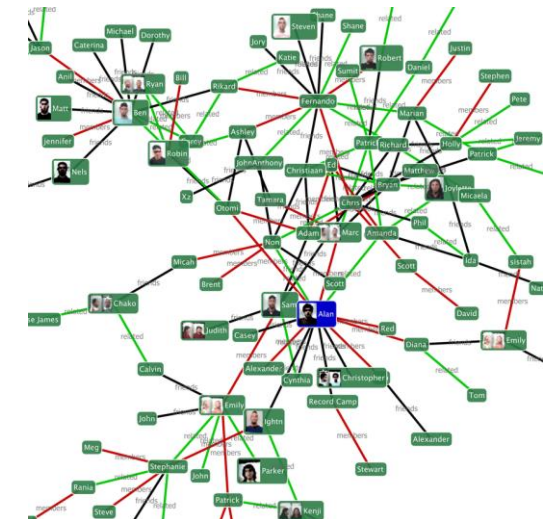
- Predict virus (covid-19) cases using mobile network data (movements)



- Fraud/anti-Money laundering
 - Creditcard
 - Insurance

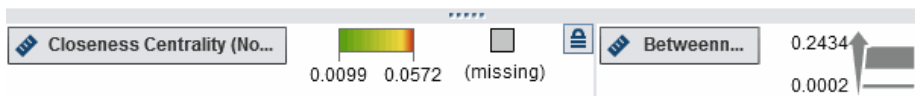
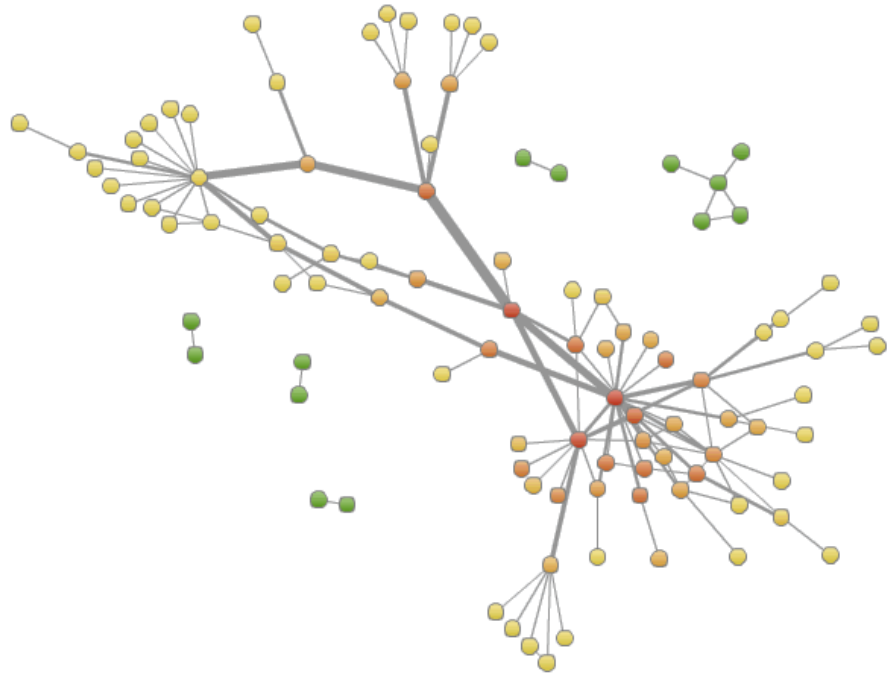


- Police
 - Cyber security
 - Crime investigation

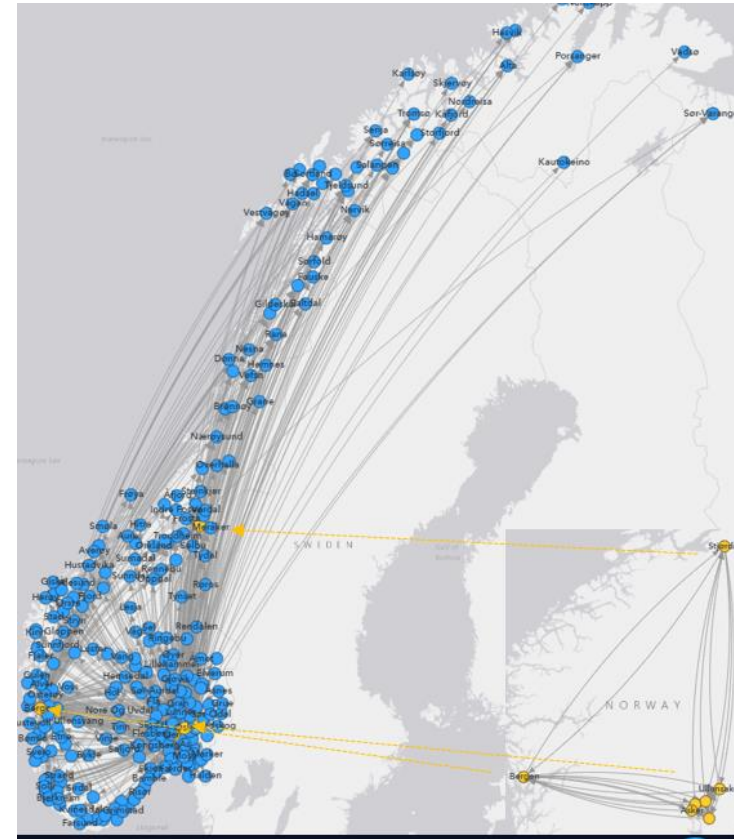


Visualize Networks in Sas Visual Analytics

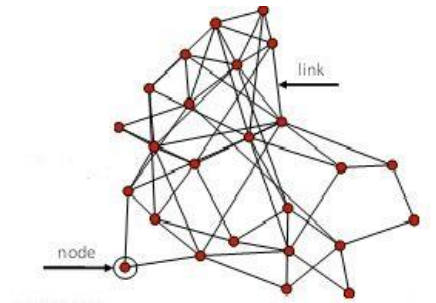
- Object 'Network analysis'



- Object 'Geo Network'



Network properties: Nodes



- A node is a general term use to denote an actor or object in the network.
- Depending on the network, nodes could represent individuals, products, countries, sportteams, blood types, business, school bus stops.
- In a network, nodes can be weighted or unweighted.

- Node classification
 - A weighed node represent a school bus stop, where the weight indicates the number of student at that location.
 - An unweighted node could represent an individual in a social network.

Unweighted	Weighted
	



Network properties: links

- A link represents an interaction between two nodes in a network.
- Depending on the network, links can be used to represent:
 - Whether a friendship exists between individuals
 - Whether a direct flight exists between airports
 - Whether two products were purchased in the same transaction
 - Whether two facilities are within a certain distance of each other
 - And so on.
- Network connectivity is influenced by the total number of links (L) in a network, and links can be directed or undirected, weighted or unweighted.



Network direction and link classification

- Network direction

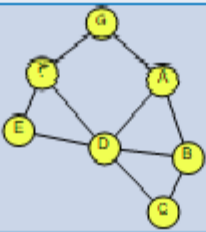
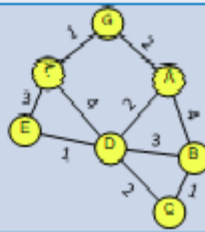
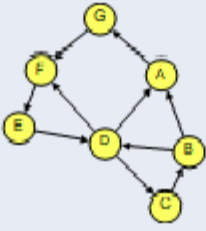
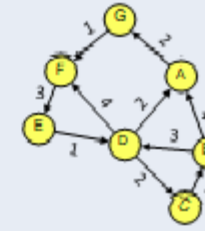
- Undirected 

- Node A is linked to node B
- Node B is linked to node A

- Directed 

- Node A points to node B
- Node B points to node A

Link Classification

	Unweighted	Weighted
Undirected		
Directed		

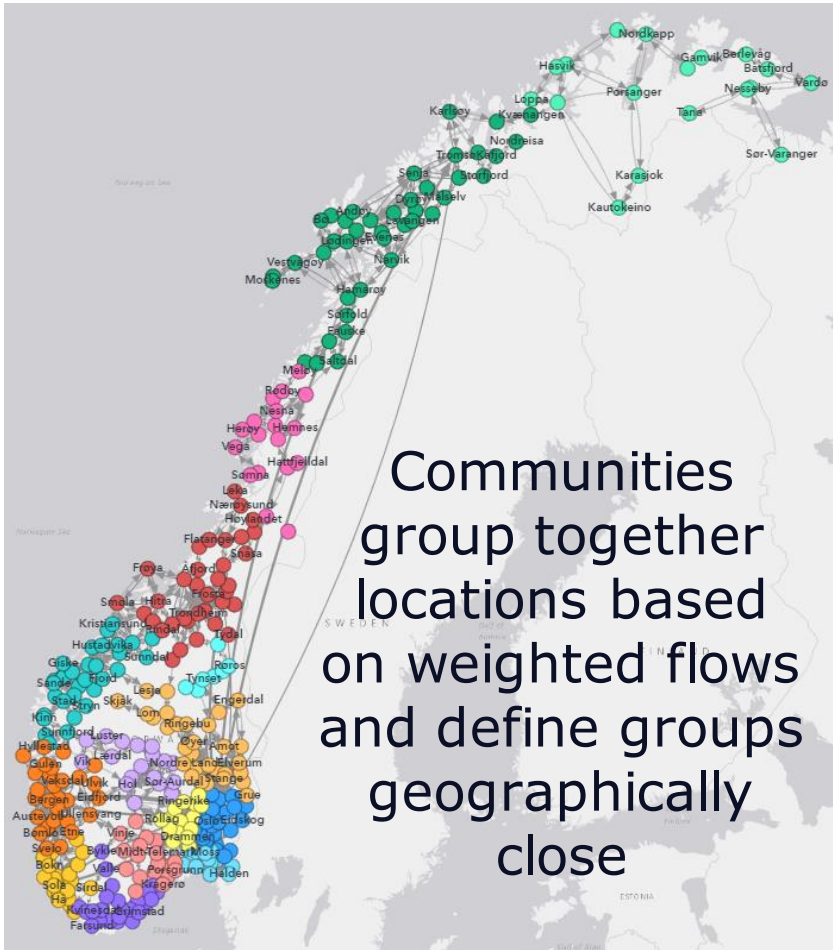


Different kind of network analysis

- Centrality measures
 - Network density
 - Degree centrality
 - Betweenness and closeness centrality
 - Influence centrality
 - Hub and authority
 - Rankpage centrality (like google)
- Subnetworks
 - Connected and biconnected components
 - Maximal cliques
 - Community detection
 - Path, shortest path and cycles
 - Pattern matching
 - Core deco
- Bipartite networks
- Network optimization (optimization license required)



Finding Communities



```
proc network
```

```
loglevel = &loglevel;
```

```
direction = undirected
```

```
links = &_worklib..output_dp_od_adj_und
```

```
outnodes = &_worklib..output_n1_nodes_comm;
```

```
linksVar
```

```
from = loc_from
```

```
to = loc_to
```

```
weight = &weight_und_var.;
```

```
community
```

```
algorithm = louvain
```

```
resolution_list = 1;
```

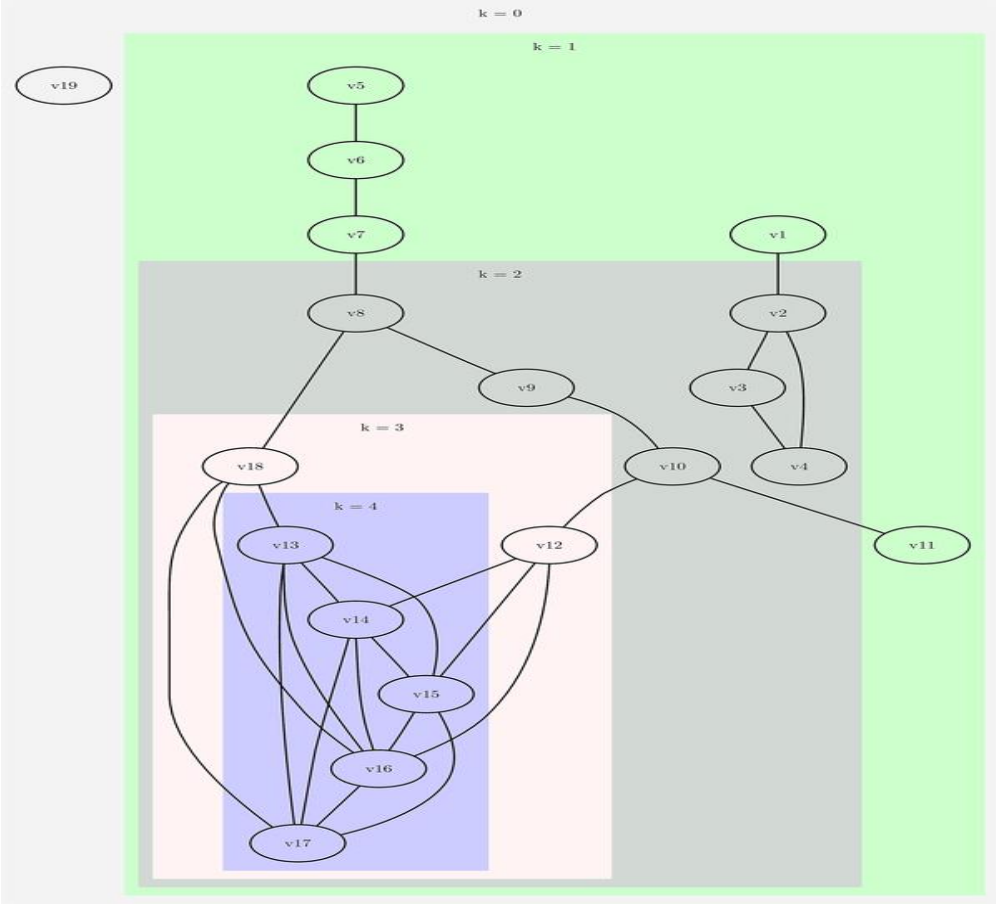
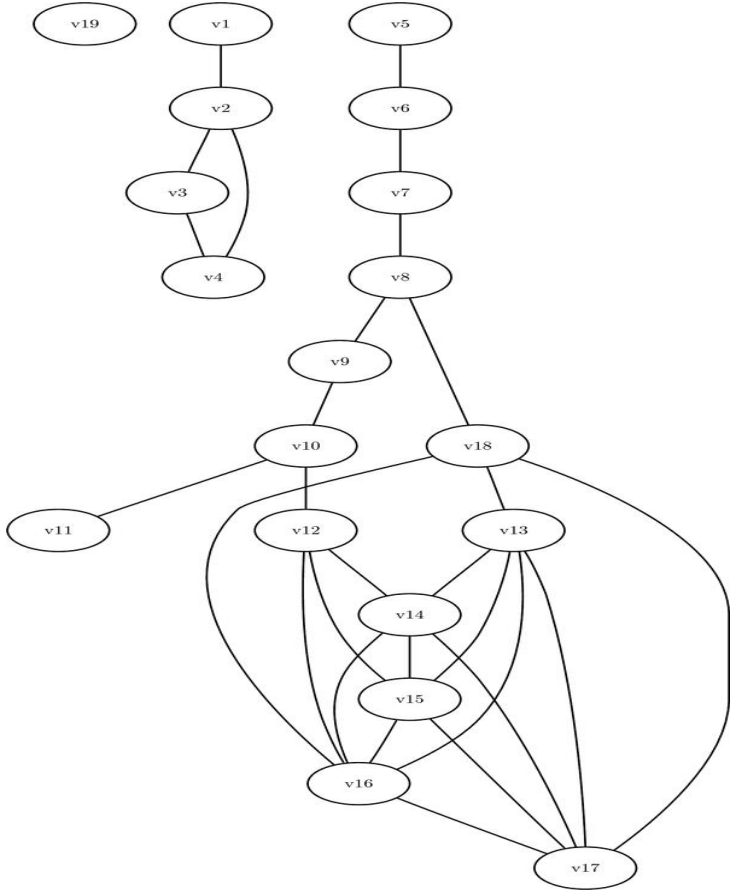
```
by date;
```

```
display / excludeall;
```

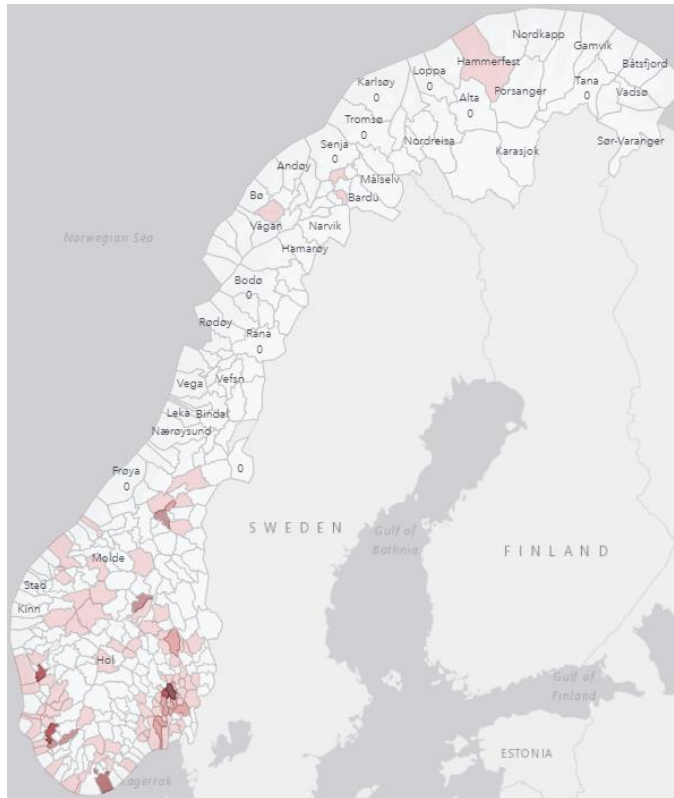
```
run;
```



K-core decomposition. How many connection



Outcome of K-core decomposition and next weeks Covid-19 case where highly correlated.



```
proc network
```

```
loglevel = &loglevel
```

```
direction = directed
```

```
links = &_worklib..output_dp_od_adj
```

```
outnodes = &_worklib..output_n1_nodes_score;
```

```
linksVar
```

```
from = loc_from
```

```
to = loc_to
```

```
weight = &weight_var.;
```

```
core;
```

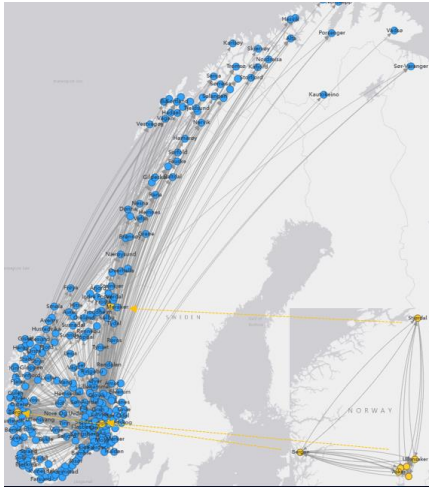
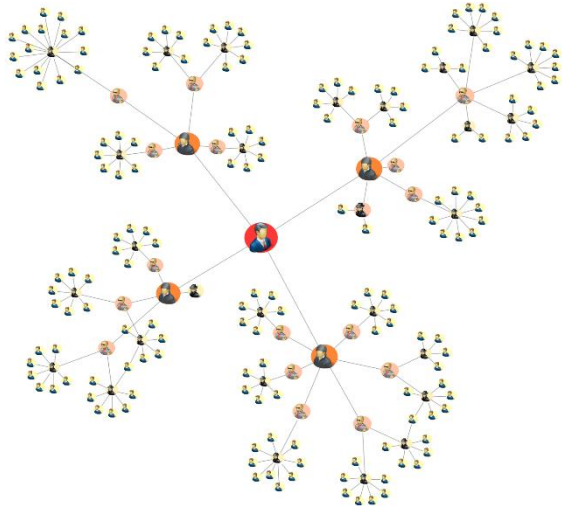
```
by date;
```

```
display / excludeall;
```

```
run;
```



To identify leader , followers in a network. (between closeness) centrality



```
proc network  loglevel = &loglevel.  direction = directed

links = &_worklib..output_dp_od_adj

outnodes = &_worklib..output_n1_nodes_centrality;

linksvar from = loc_from to = loc_to weight =
&weight_var.;

centrality degree = both influence = weight
clusteringcoef close = weight

closenopath = diameter

between = weight

betweennorm = false

hub = weight

auth = weight

pagerank = weight;

by date;

display / excludeall;

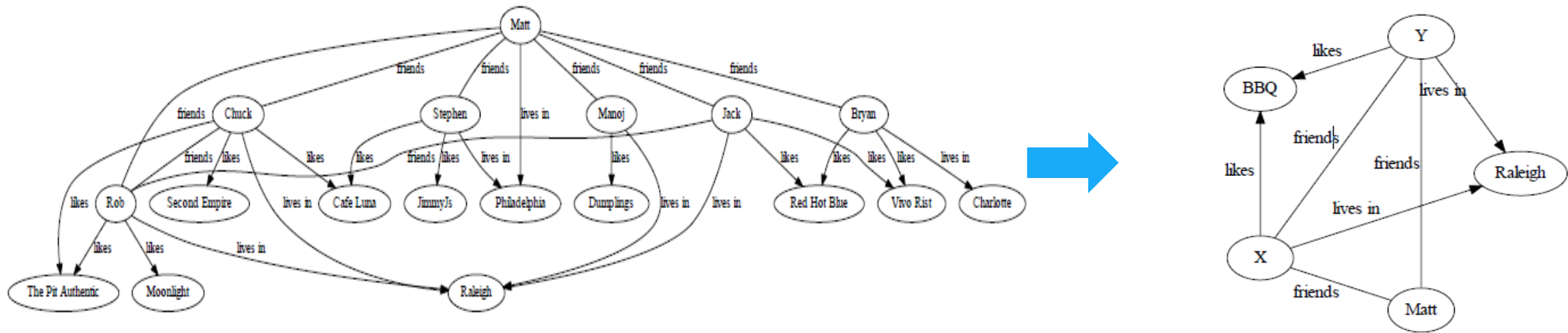
run;
```



Pattern Matching in a Social Network

Looking for friends from Matt that like bbq and lives Raleigh.

Figure 4 Social Network G

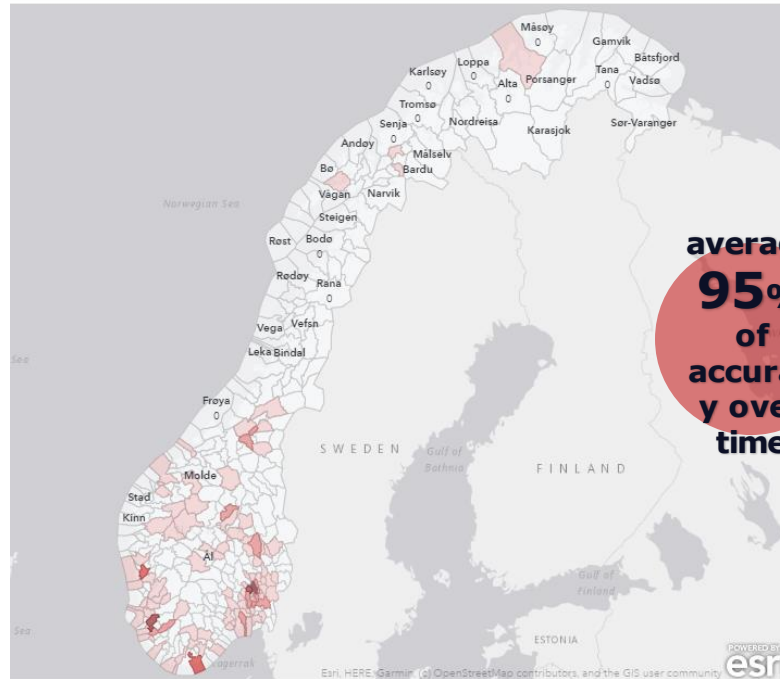


The outcome of network features are accurate

Predicting locations with **new cases** using Machine Learning

Machine learning models) using NETWORK features accurately classify where new infections occur.

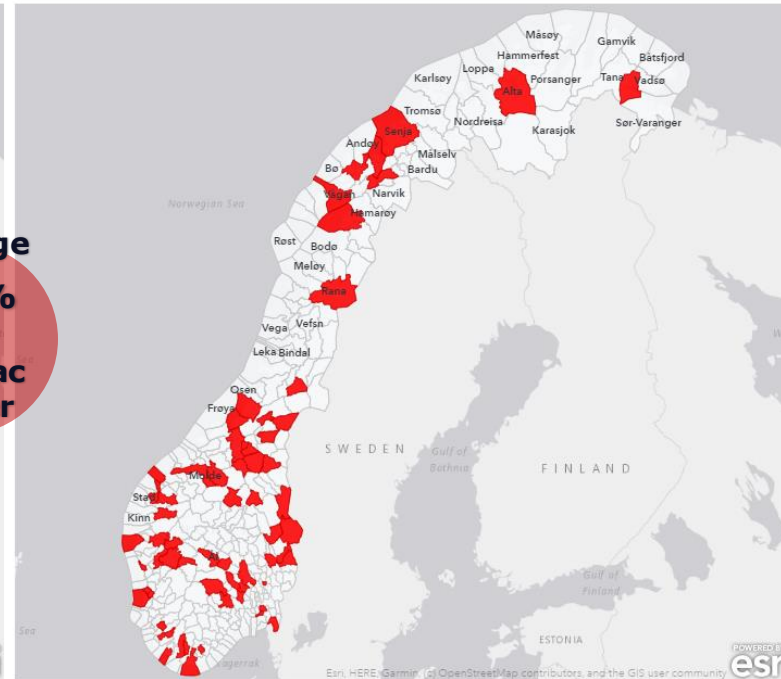
Distribution of cases



cases group

>100	0	1-10	11-25	26-50	51-100
------	---	------	-------	-------	--------

Distribution of classification



cm_new

TN	TP
----	----

average
95%
of
accurac
y over
time





Thank you

Hans de Wit, Telenor Mobile Norway, +47 48 29 1399