The Use of Social Network Metrics in Model-Based Systems Engineering

Simon Perry

INCOSE UK MBSE IG

2023-07-20

SCARECROW CONSULTANTS

Motivation

The Metrics

Method

Results

Conclusions

Further Work

Summary

SCARECROW
CONSULTANTS

Motivation

- System models, e.g. SysML models, can be very complex
  - 10000s of elements & relationships
- Understanding this complexity is non-trivial…e.g.
- How well connected is my model?
  - Richness of connections?
  - Are there any unconnected parts?
- Which are most important elements?

SCARECROW
CONSULTANTS

Graph Theory – A Possible Solution

- A SysML model is a mathematical *graph*
  - *Vertices & edges*
- Graph theory, together with social network theory, defines *metrics* to address the questions
- Proof-of-concept test carried out to explore the metrics
  - Ease of generation
  - Usefulness

SCARECROW
CONSULTANTS

Motivation

The Metrics

Method

Results

Conclusions

Further Work

Summary

SCARECROW
CONSULTANTS

Overview of Metrics Used

Metric [1]	What Does it tell us about the model?
Degree Centrality (calculated per graph vertex)	Number of relationships each individual model element has
Average Degree (calculated for whole graph)	Average number of relationships each model element has
Density (calculated for whole graph)	How well connected, as a whole, the model is
Components (calculated for whole graph)	Whether the model contains parts that are unconnected
Closeness Centrality (calculated per graph vertex)	How close, on average, a model element is to all others
Betweenness Centrality (calculated per graph vertex)	The extent to which a model element lies on paths between other model elements

[1] Full details on how the metrics are calculated can be found in the accompanying paper

SCARECROW CONSULTANTS

Motivation
The Metrics
Method
Results
Conclusions
Further Work
Summary

SCARECROW CONSULTANTS

Method

- Anonymise data
  - Run VBScript across model, creating aliases
- Extract *vertices* & *edges*
  - Use SQL
- Format data
- Import to Mathematica®
- Evaluate Notebook

SCARECROW
CONSULTANTS

Extracting the Vertices - SQL

```
select alias from t_object
where object_type in ('Class', 'UseCase',
'Actor', 'Requirement')
```

```
A55,
A56,
A57,
A58,
A43,
A9,
A21,
A37,
```

SCARECROW CONSULTANTS

Extracting the Edges - SQL

```
select s.alias, t.alias from t_object as s, t_object as t, t_connector as c
where c.start_object_id in (select object_id from t_object
where object_type in ('Class', 'UseCase', 'Actor', 'Requirement'))
and c.end_object_id in (select object_id from t_object
where object_type in ('Class', 'UseCase', 'Actor', 'Requirement'))
and c.start_object_id = s.object_id
and c.end_object_id = t.object_id
```

```
U2028,U2018,
U2028,U2025,
U2027,U2025,
U2027,U2028,
C24,C37,
C32,C108,
```

SCARECROW
CONSULTANTS

Formatted Data

- Vertices

```
{A1,A10,A100,A101,A102, … }
```

- Edges

```
{A1<->A3,A1<->A39,A1<->A49,A1<-
>U160,A1<->U226,A10<->A1,A10<->A50, … }
```

SCARECROW
CONSULTANTS

Motivation
The Metrics
Method
Results
Conclusions
Further Work
Summary

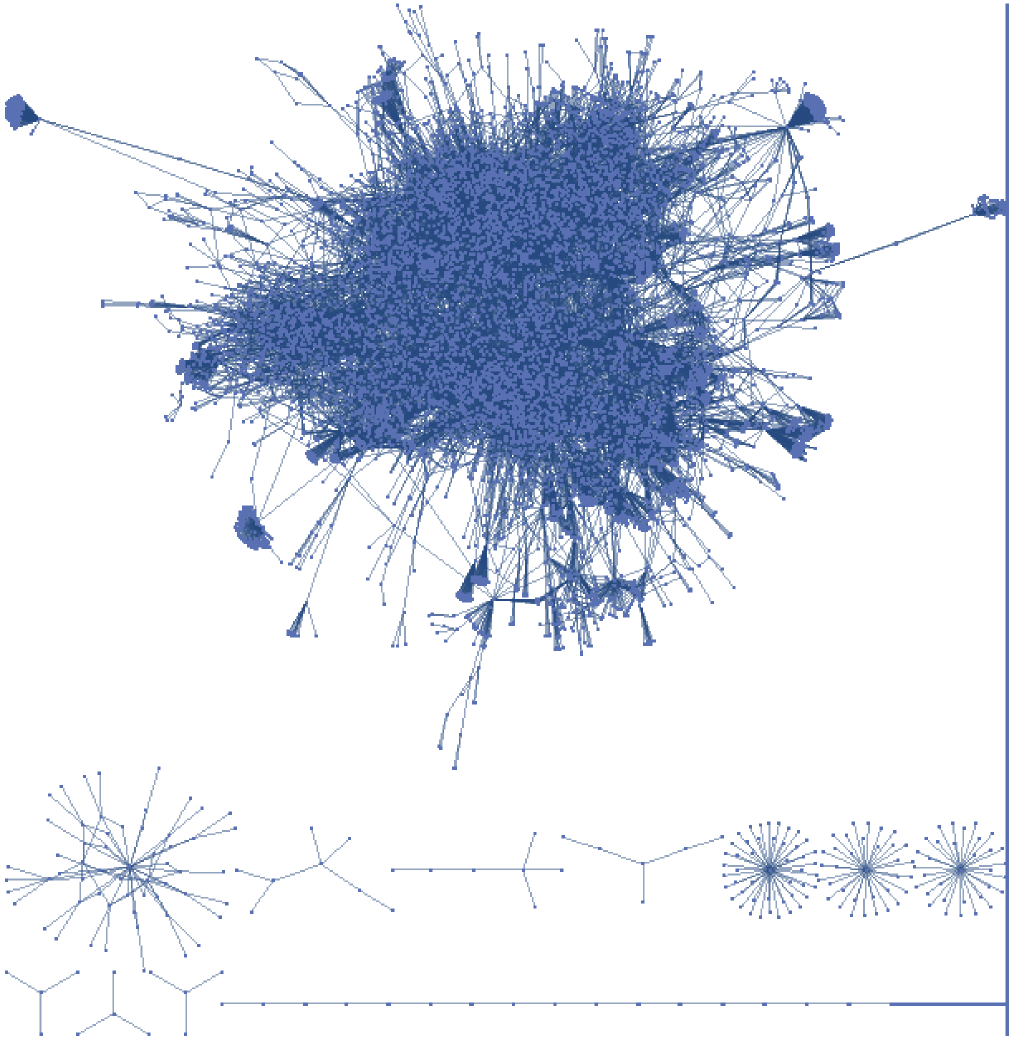
SCARECROW
CONSULTANTS

Results – Approximate timings

- Model size:
  - ≈ 11k elements & ≈ 22k relationships
- Anonymise data
  - 15 minutes
- Extract *vertices* & *edges*
  - < 10 seconds
- Format data
  - 5 minutes
- Import to Mathematica® & Evaluate Notebook
  - < 20 seconds

SCARECROW CONSULTANTS

The Generated Graph



SCARECROW
CONSULTANTS

Simple Metrics

Each model element is connected, on average, to 4 others

Vertex Count	Edge Count	Average Degree	Density	Number of Components
10 763	21 971	4.08269	0.000379362	966

The model is sparsely connected

The model is very disjointed, with 966 unconnected components

SCARECROW
CONSULTANTS

Components

89% of the model elements form a single component

The rest form components containing between 1 and 58 elements

Component Sizes

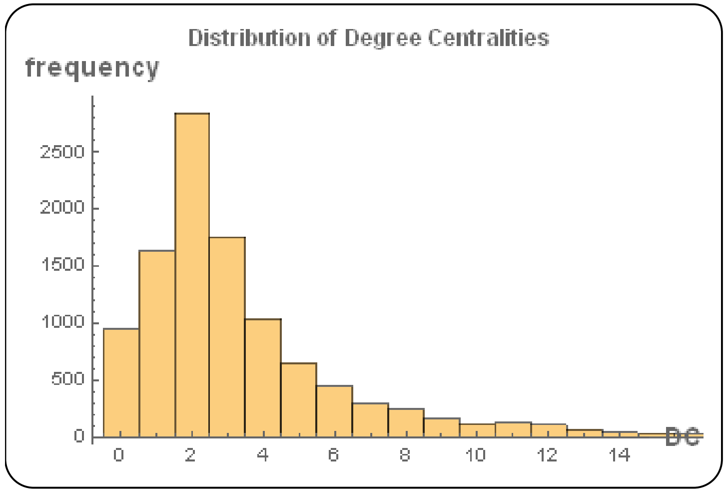
Component Size	Count
9587	1
58	1
47	1
34	1
33	1
8	1
7	1
6	1
4	3
3	5
2	6
1	944

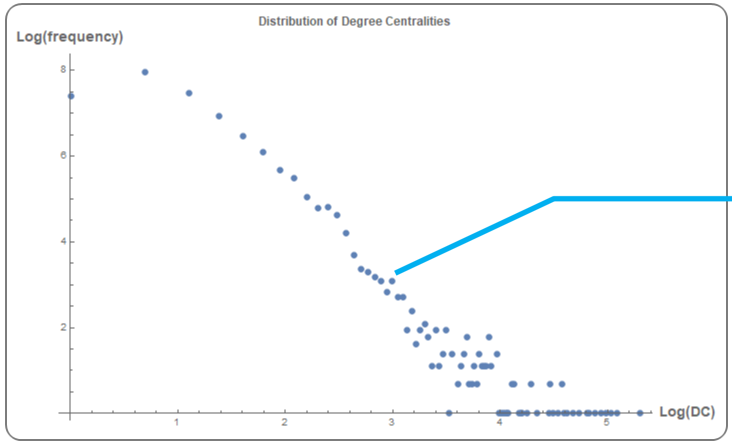
There are 944 model elements that are completely unconnected to anything

SCARECROW CONSULTANTS

Degree Centralities

A small number of model elements have LOTS of connections – an indication of their importance





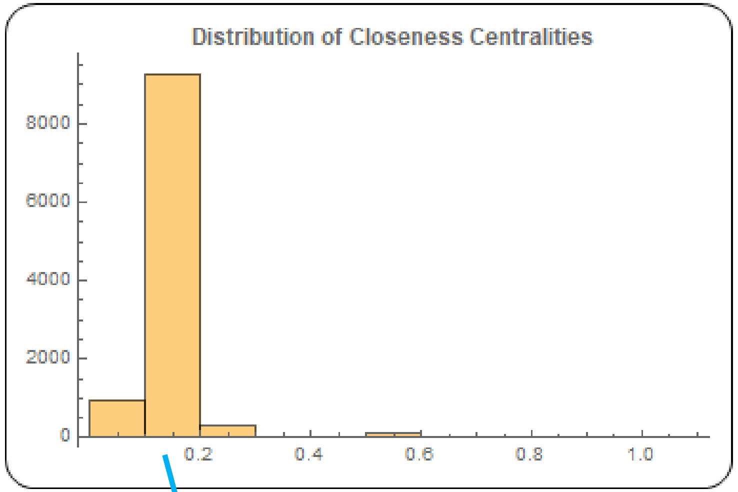
Plotting distribution as a LogLog plot suggests a “scale-free network” - a large fraction of edges connected to a small fraction of vertices

Top 20 Degree Centralities (DC)

Vertex	DC
C2968	201
C858	162
C17	153
C841	146
C37	140
C9	132
A66	125
R54	122
C851	114
C827	108
C856	102
C8	99
C813	97
C1020	97
C823	94
R425	89
C994	87
C985	87
C825	86
C819	77

SCARECROW CONSULTANTS

Closeness Centralities



The majority of elements have low closeness centrality (89% had $C_i < 0.2$)

On reflection, not unexpected; most models more like trees than webs

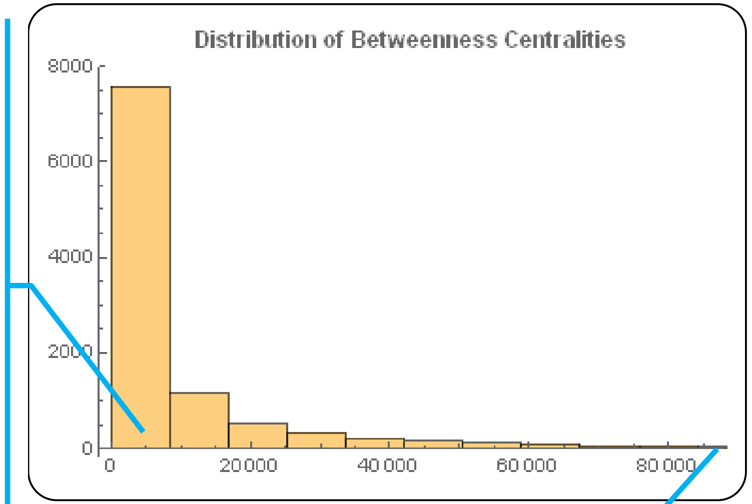
Top 20 Closeness Centralities (CC)

Vertex	CC
U1434	1.
U1242	1.
R950	1.
R556	1.
R555	1.
R548	1.
R547	1.
R545	1.
R543	1.
C34	1.
C2320	1.
C2319	1.
C2318	1.
C2047	1.
C1602	1.
C1600	1.
C1595	1.
C1581	1.
C1575	1.
C1441	1.

SCARECROW CONSULTANTS

Betweenness Centralities

32% of model elements have betweenness centrality of 0; influenced by the number of disconnected components



Histogram has a long tail; maximum betweenness centrality is 8.4×10^6

Top 20 Betweenness Centralities (BC)

Vertex	BC
C858	8.41982×10^6
A66	3.88063×10^6
C586	3.67475×10^6
A4	3.05368×10^6
C827	3.02099×10^6
C851	2.93601×10^6
C589	2.81934×10^6
C813	2.40895×10^6
C2968	2.35731×10^6
C1090	2.23806×10^6
C856	1.62568×10^6
C825	1.59879×10^6
C32	1.59238×10^6
C819	1.47725×10^6
C17	1.42726×10^6
C938	1.40557×10^6
C845	1.39625×10^6
C815	1.38806×10^6
C37	1.2662×10^6
R425	1.26397×10^6

SCARECROW
CONSULTANTS

Common Elements in Top 20 DCs & BCs

Top 20 Degree Centralities (DC)

Vertex	DC
C2968	201
C858	162
C17	153
C841	146
C37	140
C9	132
A66	125
R54	122
C851	114
C827	108
C856	102
C8	99
C813	97
C1020	97
C823	94
R425	89
C994	87
C985	87
C825	86
C819	77

Top 20 Betweenness Centralities (BC)

Vertex	BC
C858	8.41982×10^6
A66	3.88063×10^6
C586	3.67475×10^6
A4	3.05368×10^6
C827	3.02099×10^6
C851	2.93601×10^6
C589	2.81934×10^6
C813	2.40895×10^6
C2968	2.35731×10^6
C1090	2.23806×10^6
C856	1.62568×10^6
C825	1.59879×10^6
C32	1.59238×10^6
C819	1.47725×10^6
C17	1.42726×10^6
C938	1.40557×10^6
C845	1.39625×10^6
C815	1.38806×10^6
C37	1.2662×10^6
R425	1.26397×10^6

Vertices Common to both Top 20 Degree & Betweenness Centralities

Vertex
A66
C17
C2968
C37
C813
C819
C825
C827
C851
C856
C858
R425

The most important elements?

SCARECROW CONSULTANTS

Motivation

The Metrics

Method

Results

Conclusions

Further Work

Summary

SCARECROW CONSULTANTS

Conclusions

- Method works
  - Now all set up, can analyse a model in < 1 hour
- Reveals useful information
  - Disconnected model elements
  - Identifies "important" model elements as measured by *degree centrality* (they are connected to lots of other elements) and *betweenness centrality* (they are essential links in the traceability of the model)
- Some metrics (e.g. *closeness centrality*) perhaps not that useful

SCARECROW
CONSULTANTS

Motivation
The Metrics
Method
Results
Conclusions
Further Work
Summary

SCARECROW
CONSULTANTS

Topics for Further Research

- Relationship direction
- Relationship types
  - What types should be considered?
  - Weighting of types?
- Multiple relationships
- Applying separately to structural & behavioural aspects
- Other metrics
  - Additional types of centrality
  - Cores

SCARECROW
CONSULTANTS

Motivation
The Metrics
Method
Results
Conclusions
Further Work

Summary

SCARECROW CONSULTANTS

Summary

- As a proof of concept, approach works
  - Want to reduce amount of manual intervention
- Initial results suggest that idea has some use
- More data needed
  - Please contact Simon Perry if you have models you would like analysed

SCARECROW
CONSULTANTS