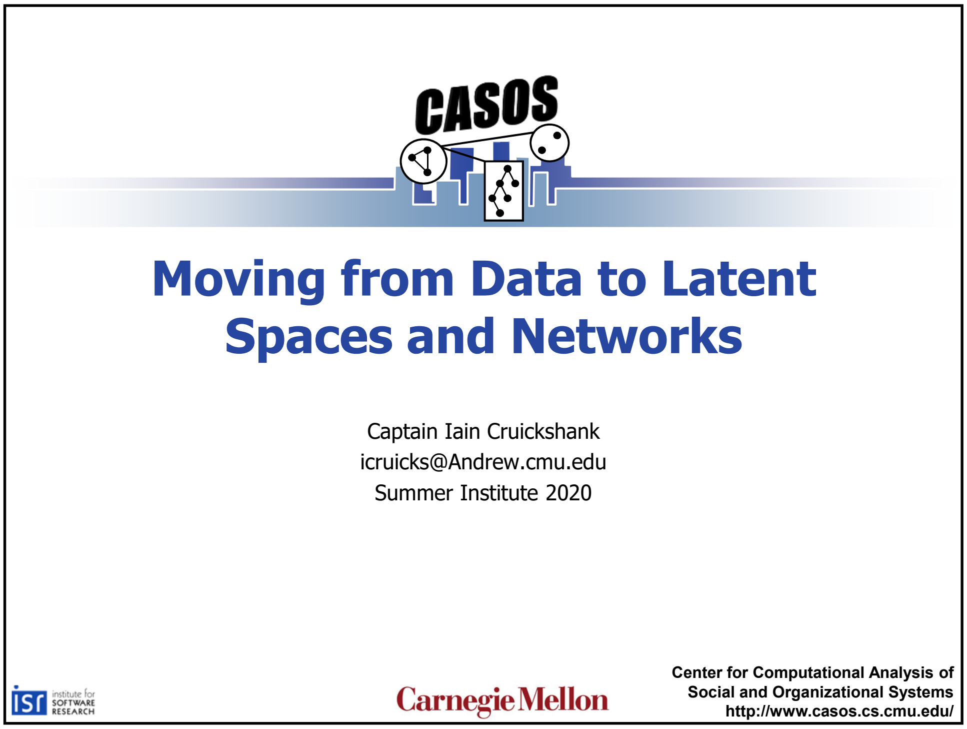
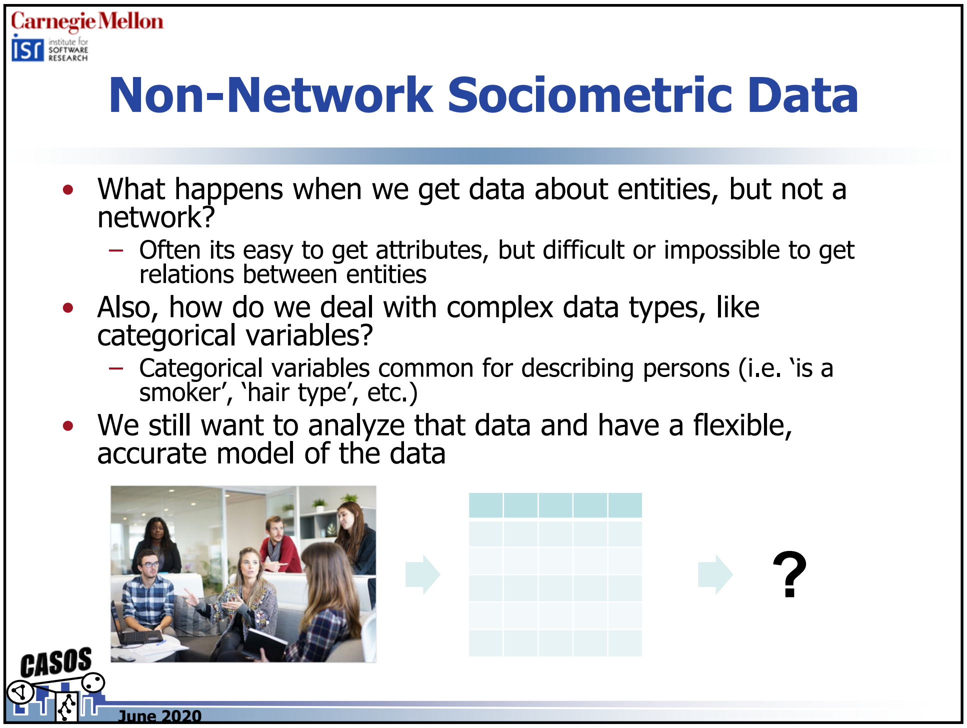


Moving from Data to Latent Spaces and Networks

Captain Iain Cruickshank
icruicks@Andrew.cmu.edu
Summer Institute 2020

Carnegie Mellon
Center for Computational Analysis of Social and Organizational Systems
http://www.casos.cs.cmu.edu/

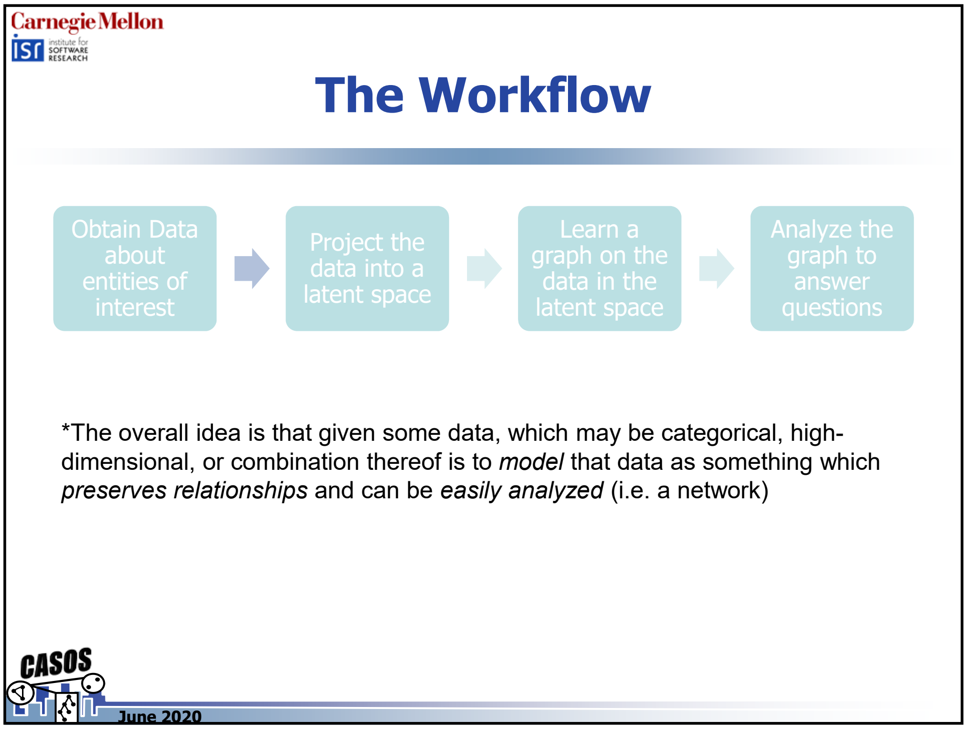


Non-Network Sociometric Data

- What happens when we get data about entities, but not a network?
  - Often its easy to get attributes, but difficult or impossible to get relations between entities
- Also, how do we deal with complex data types, like categorical variables?
  - Categorical variables common for describing persons (i.e. 'is a smoker', 'hair type', etc.)
- We still want to analyze that data and have a flexible, accurate model of the data

?

June 2020



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

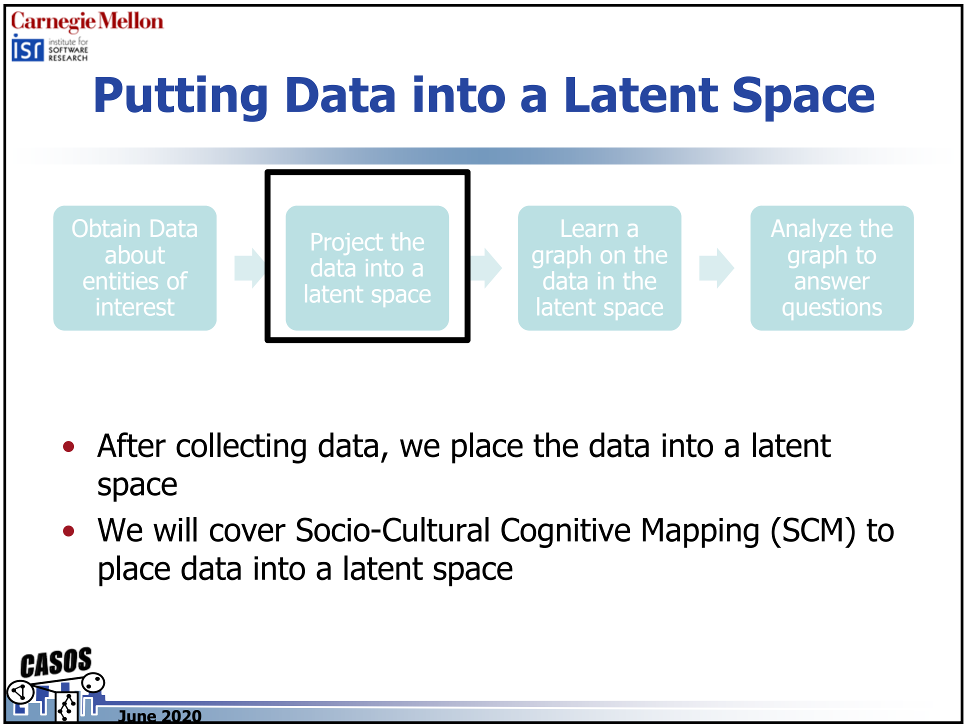
The Workflow

Obtain Data about entities of interest → Project the data into a latent space → Learn a graph on the data in the latent space → Analyze the graph to answer questions

*The overall idea is that given some data, which may be categorical, high-dimensional, or combination thereof is to *model* that data as something which *preserves relationships* and can be *easily analyzed* (i.e. a network)

CASOS

June 2020



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Putting Data into a Latent Space

Obtain Data about entities of interest → Project the data into a latent space → Learn a graph on the data in the latent space → Analyze the graph to answer questions

- After collecting data, we place the data into a latent space
- We will cover Socio-Cultural Cognitive Mapping (SCM) to place data into a latent space

CASOS

June 2020

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Overview of SCM

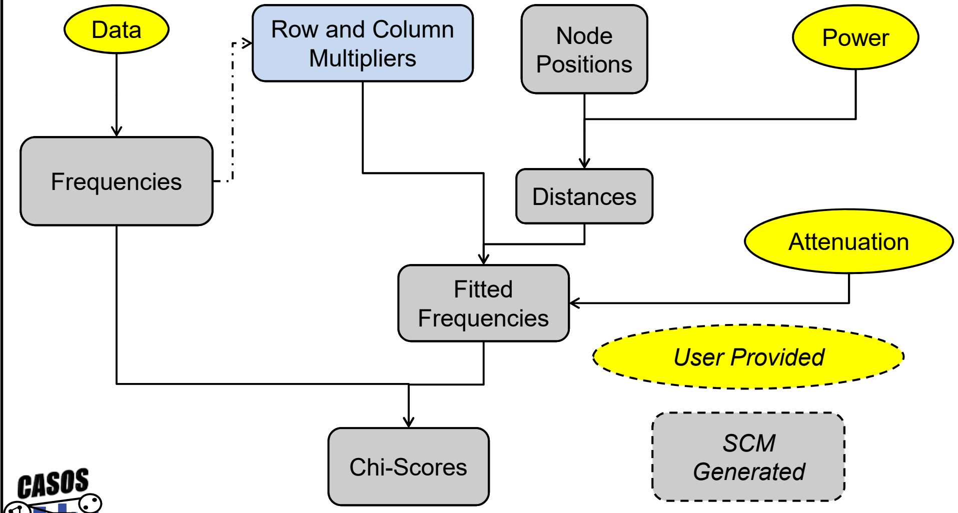
- Take a set of node attributes or network data and use the information to place nodes in space.
  - User defines the geometry of the space
  - User provides data
- Nodes that are highly similar will be near each other, while nodes that are quite different will be far apart.
- Overall goodness-of-fit is evaluated with a Chi-Squared Test

CASOS

June 2020

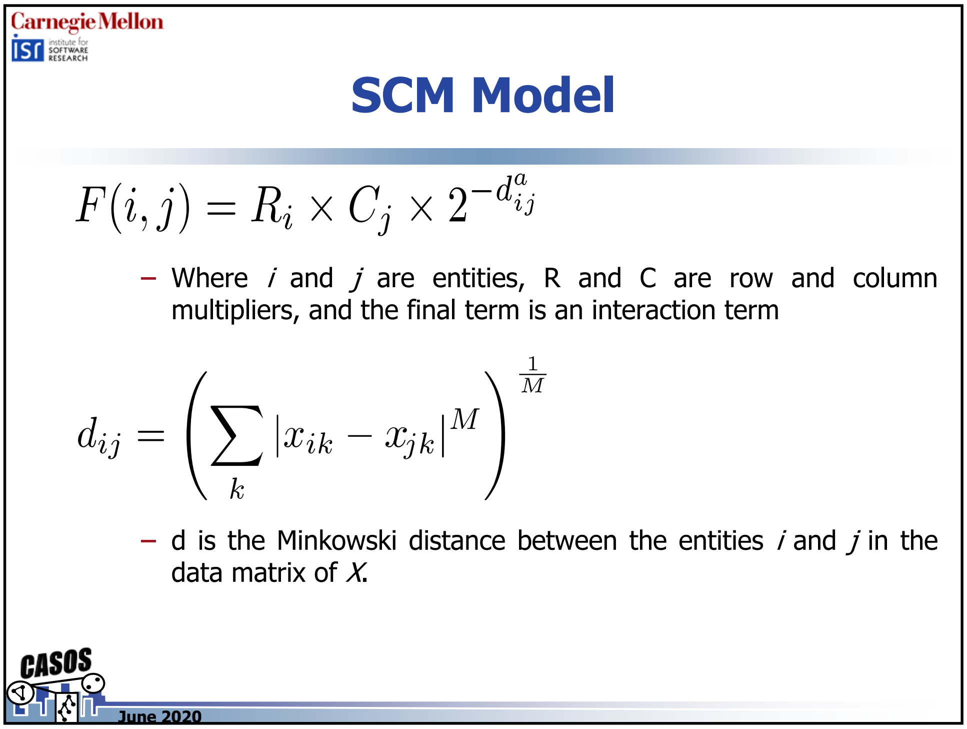
Carnegie Mellon
isr institute for SOFTWARE RESEARCH

SCM Process



CASOS

June 2020



Carnegie Mellon

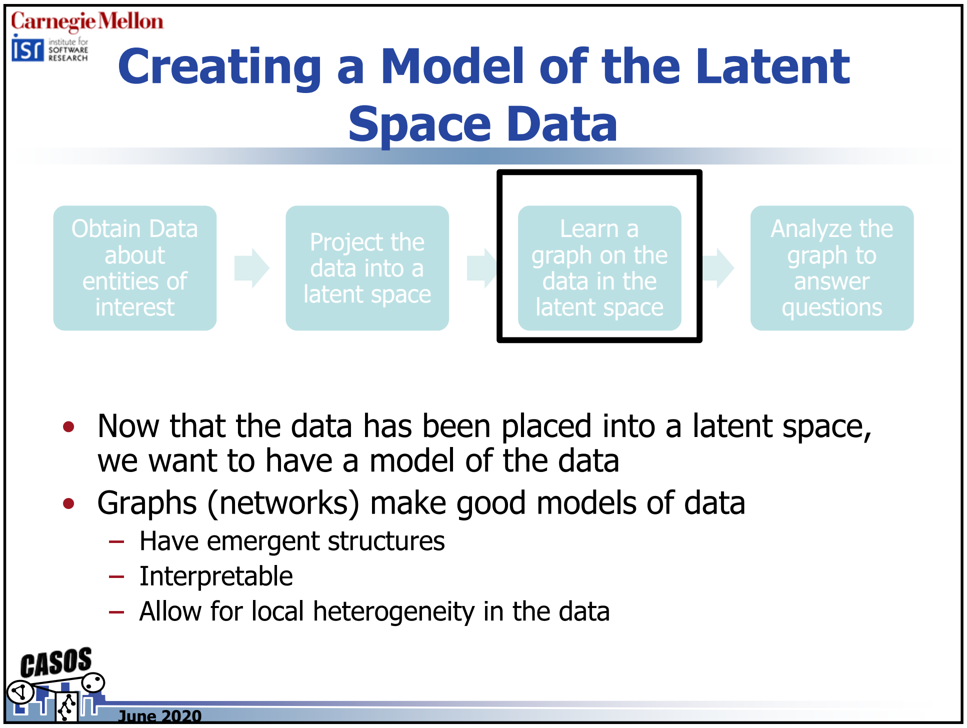
SCM Model

$$F(i,j) = R_i \times C_j \times 2^{-d_{ij}^a}$$

- Where *i* and *j* are entities, R and C are row and column multipliers, and the final term is an interaction term

$$d_{ij} = \left(\sum_k |x_{ik} - x_{jk}|^M\right)^{\frac{1}{M}}$$

- d is the Minkowski distance between the entities *i* and *j* in the data matrix of *X*.

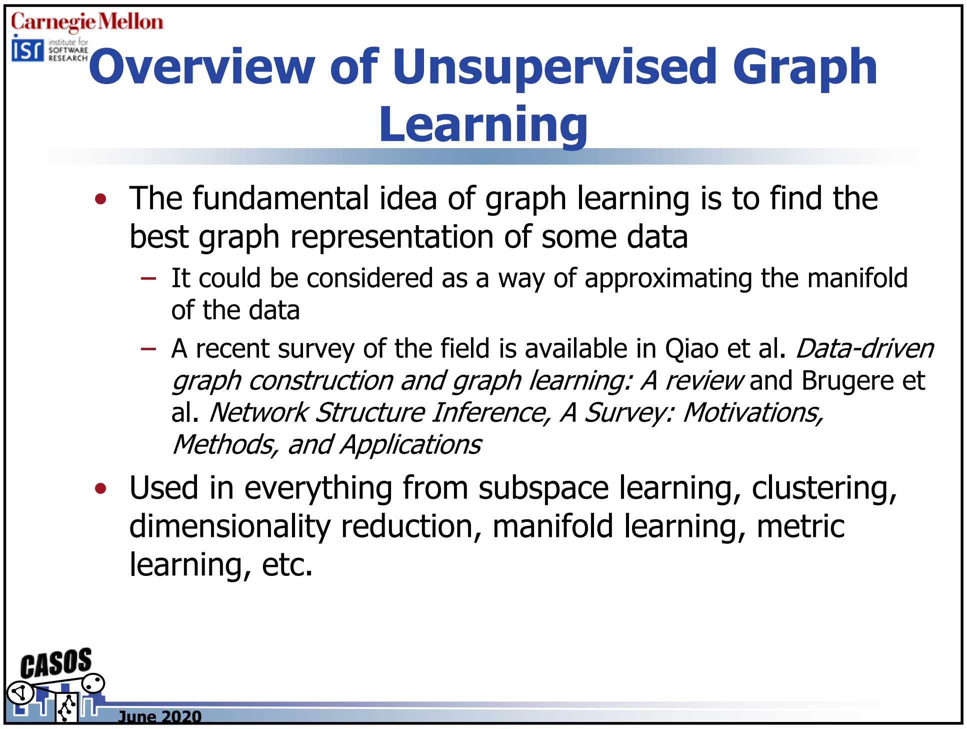


Carnegie Mellon

Creating a Model of the Latent Space Data

Obtain Data about entities of interest → Project the data into a latent space → Learn a graph on the data in the latent space → Analyze the graph to answer questions

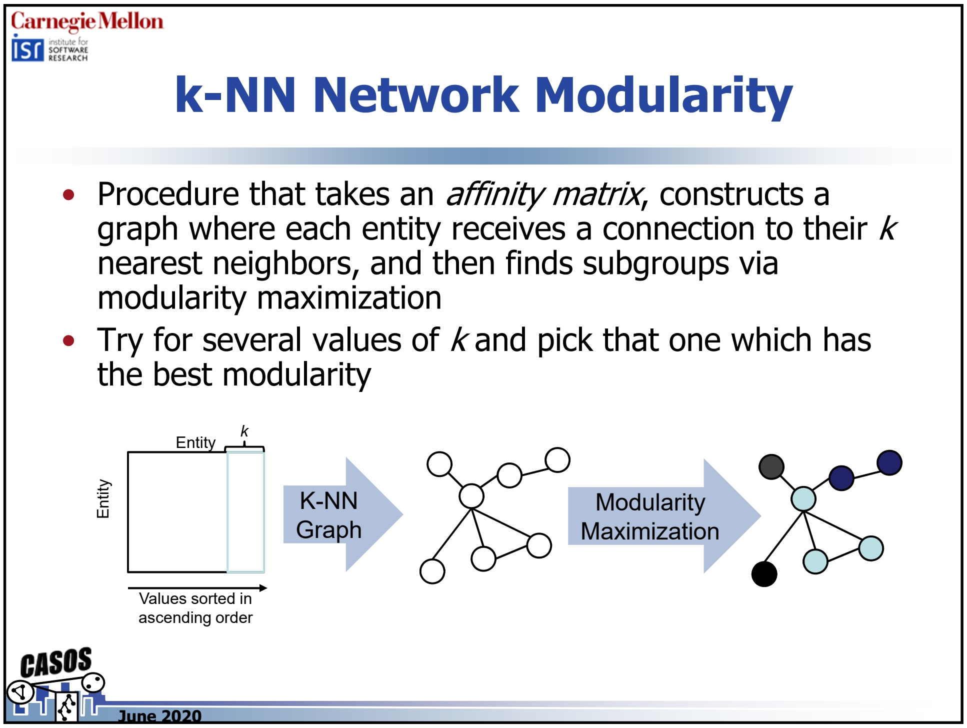
- Now that the data has been placed into a latent space, we want to have a model of the data
- Graphs (networks) make good models of data
  - Have emergent structures
  - Interpretable
  - Allow for local heterogeneity in the data



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Overview of Unsupervised Graph Learning

- The fundamental idea of graph learning is to find the best graph representation of some data
  - It could be considered as a way of approximating the manifold of the data
  - A recent survey of the field is available in Qiao et al. *Data-driven graph construction and graph learning: A review* and Brugere et al. *Network Structure Inference, A Survey: Motivations, Methods, and Applications*
- Used in everything from subspace learning, clustering, dimensionality reduction, manifold learning, metric learning, etc.



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

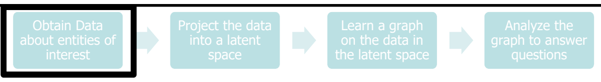
k-NN Network Modularity

- Procedure that takes an *affinity matrix*, constructs a graph where each entity receives a connection to their *k* nearest neighbors, and then finds subgroups via modularity maximization
- Try for several values of *k* and pick that one which has the best modularity

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Time for an Example!

Carnegie Mellon
isr institute for SOFTWARE RESEARCH



Obtain Data about entities of interest

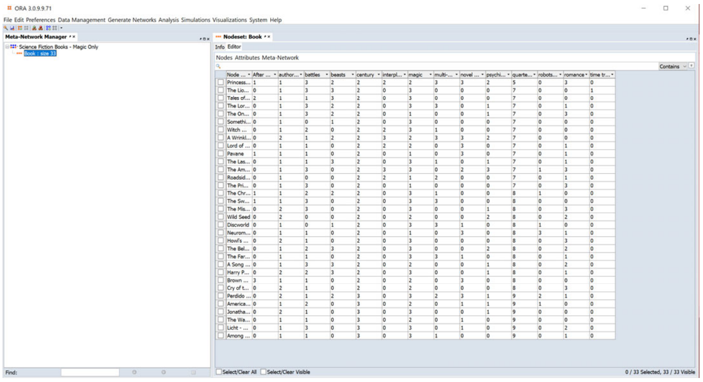
Project the data into a latent space

Learn a graph on the data in the latent space

Analyze the graph to answer questions

Step 1: Find the Data

- Read in “Science Fiction Books – Magic Only.xml”
- 33 Books
- Set of Attributes for each Book



ORA 3.0.9.9.71
File Edit Preferences Data Management Generate Networks Analysis Simulations Visualizations System Help
Meta-Network Manager
Science Fiction Books - Magic Only
Book : size 33
Nodeset: Book
Info Editor
Nodes Attributes Meta-Network
Contains
Find:

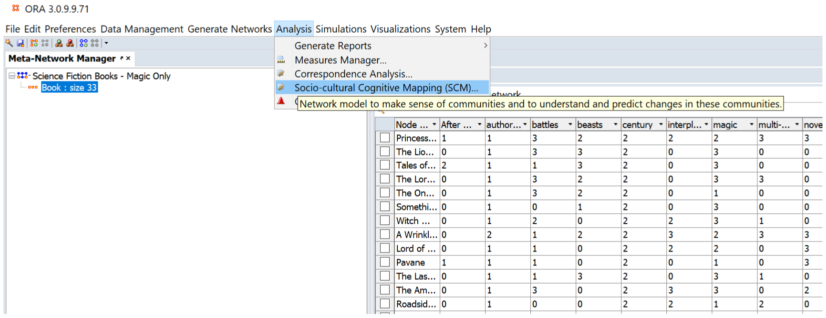
	Node ...	After ...	author...	battles	beasts	century	interpl...	magic	multi-...	novel ...	psychi...	quarte...	robots...	romance	time tr...
□	Princess...	1	1	3	2	2	2	2	3	3	2	5	0	3	0
□	The Lio...	0	1	3	3	2	0	3	0	0	0	7	0	0	1
□	Tales of...	2	1	1	3	2	0	3	0	0	0	7	0	0	0
□	The Lor...	0	1	3	2	2	0	3	3	0	1	7	0	1	0
□	The On...	0	1	3	2	2	0	1	0	0	1	7	0	3	0
□	Somethi...	0	1	0	1	2	0	3	0	0	0	7	0	0	0
□	Witch ...	0	1	2	0	2	2	3	1	0	0	7	0	0	0
□	A Wrinkl...	0	2	1	2	2	3	2	3	3	2	7	0	0	0
□	Lord of ...	0	1	1	0	2	2	2	0	3	0	7	0	1	0
□	Pavane	1	1	1	0	2	0	1	0	3	0	7	0	1	0
□	The Las...	0	1	1	3	2	0	3	1	0	1	7	0	1	0
□	The Am...	0	1	3	0	2	3	3	0	2	3	7	1	3	0
□	Roadsid...	0	1	0	0	2	2	1	2	0	0	7	0	1	0
□	The Pri...	0	1	3	0	2	0	1	0	0	0	7	0	3	0
□	The Chr...	1	1	2	2	2	0	3	1	0	0	8	1	0	0
□	The Sw...	1	1	3	0	2	0	3	3	0	0	8	0	0	0
□	The Mis...	0	2	3	0	2	0	3	0	0	1	8	0	3	0
□	Wild Seed	0	2	0	0	2	0	2	0	0	2	8	0	2	0
□	Discworld	0	1	0	1	2	0	3	3	1	0	8	1	0	0
□	Neurom...	0	1	1	0	2	0	1	0	3	0	8	3	1	0
□	Howl's ...	0	2	1	0	2	0	3	0	0	0	8	0	3	0
□	The Bel...	0	1	2	3	2	0	3	0	0	2	8	0	2	0
□	The Far...	0	1	1	0	2	0	3	3	1	0	8	0	1	0
□	A Song ...	0	1	3	3	2	0	2	0	0	1	8	0	2	0
□	Harry P...	0	2	2	3	2	0	3	0	0	1	8	0	1	0
□	Brown ...	3	1	1	0	2	0	2	0	3	0	8	0	0	0
□	Cry of t...	0	2	1	0	2	0	2	0	0	0	8	0	3	0
□	Perdido ...	0	2	1	2	3	0	3	2	3	1	9	2	1	0
□	America...	0	1	2	0	3	0	2	0	1	1	9	1	0	0
□	Jonatha...	0	2	1	0	3	0	3	0	0	1	9	0	0	0
□	The Wa...	0	1	1	0	3	0	2	0	1	0	9	0	0	0
□	Licht - ...	0	1	3	0	3	0	3	0	1	0	9	0	2	0
□	Among ...	0	1	1	0	3	0	3	1	0	0	9	0	1	0

□ Select/Clear All □ Select/Clear Visible 0 / 33 Selected, 33 / 33 Visible

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Obtain Data about entities of interest → **Project the data into a latent space** → Learn a graph on the data in the latent space → Analyze the graph to answer questions

Step 2: Start the SCM


ORA 3.0.9.9.71
File Edit Preferences Data Management Generate Networks Analysis Simulations Visualizations System Help
Meta-Network Manager
Science Fiction Books - Magic Only
Book : size 33
Generate Reports
Measures Manager...
Correspondence Analysis...
Socio-cultural Cognitive Mapping (SCM)...

Network model to make sense of communities and to understand and predict changes in these communities.

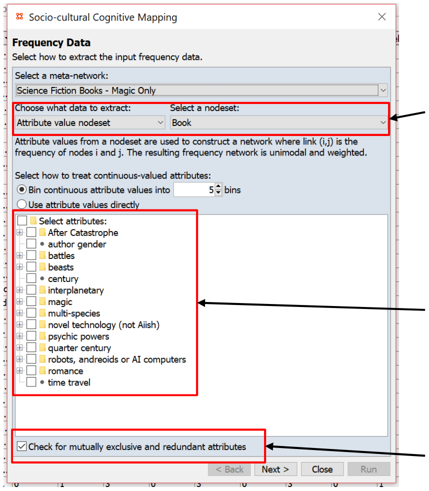
Node ...	After ...	author...	battles	beasts	century	interpl...	magic	multi-...	nove
Princess...	1	1	3	2	2	2	2	3	3
The Lio...	0	1	3	3	2	0	3	0	0
Tales of...	2	1	1	3	2	0	3	0	0
The Lor...	0	1	3	2	2	0	3	3	0
The On...	0	1	3	2	2	0	1	0	0
Somethi...	0	1	0	1	2	0	3	0	0
Witch ...	0	1	2	0	2	2	3	1	0
A Wrinkl...	0	2	1	2	2	3	2	3	3
Lord of ...	0	1	1	0	2	2	2	0	3
Pavane	1	1	1	0	2	0	1	0	3
The Las...	0	1	1	3	2	0	3	1	0
The Am...	0	1	3	0	2	3	3	0	2
Roadsid...	0	1	0	0	2	2	1	2	0

CASOS
June 2020

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Obtain Data about entities of interest → **Project the data into a latent space** → Learn a graph on the data in the latent space → Analyze the graph to answer questions

Select the Frequency Data



Frequency Data

Select how to extract the input frequency data.

Select a meta-network: Science Fiction Books - Magic Only

Choose what data to extract: Attribute value nodeset

Select a nodeset: Book

Attribute values from a nodeset are used to construct a network where link (i,j) is the frequency of nodes i and j. The resulting frequency network is unimodal and weighted.

Select how to treat continuous-valued attributes:

● Bin continuous attribute values into 5 bins

○ Use attribute values directly

- □ Select attributes:
  - □ After Catastrophe
  - □ author gender
  - □ battles
  - □ beasts
  - □ century
  - □ interplanetary
  - □ magic
  - □ multi-species
  - □ novel technology (not Aiish)
  - □ psychic powers
  - □ quarter century
  - □ robots, andreoids or AI computers
  - □ romance
  - □ time travel

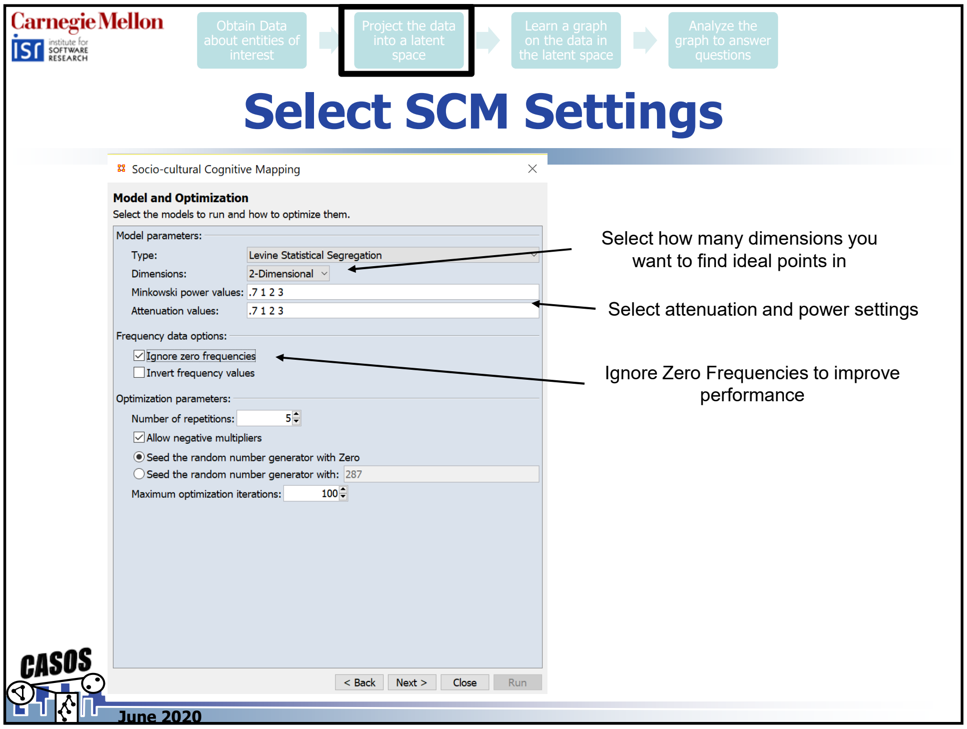
☑ Check for mutually exclusive and redundant attributes

Can select a network or attributes of a node, which creates a frequency network

Can select different attributes and different levels of attributes

Check for mutually exclusive and redundant attributes. Generally you always want to do this to improve performance

CASOS
June 2020



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Obtain Data about entities of interest → Project the data into a latent space → Learn a graph on the data in the latent space → Analyze the graph to answer questions

Select SCM Settings

Socio-cultural Cognitive Mapping

Model and Optimization

Select the models to run and how to optimize them.

Model parameters:

Type: Levine Statistical Segregation

Dimensions: 2-Dimensional

Minkowski power values: .7 1 2 3

Attenuation values: .7 1 2 3

Frequency data options:

☑ Ignore zero frequencies

□ Invert frequency values

Optimization parameters:

Number of repetitions: 5

☑ Allow negative multipliers

● Seed the random number generator with Zero

○ Seed the random number generator with: 287

Maximum optimization iterations: 100

< Back | Next > | Close | Run

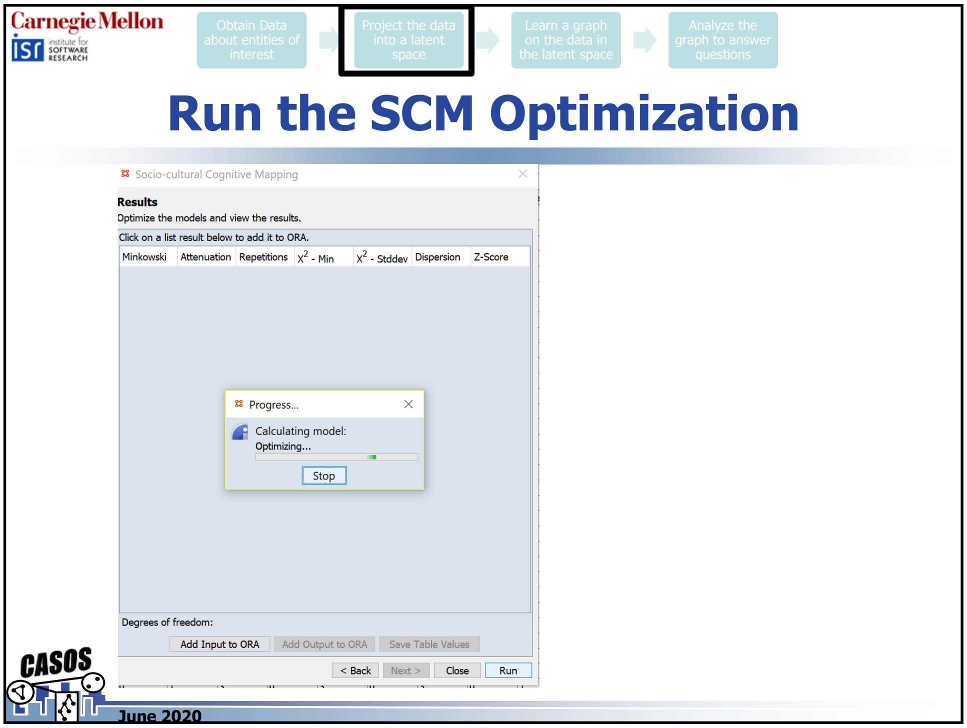
Select how many dimensions you want to find ideal points in

Select attenuation and power settings

Ignore Zero Frequencies to improve performance

CASOS

June 2020



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Obtain Data about entities of interest → Project the data into a latent space → Learn a graph on the data in the latent space → Analyze the graph to answer questions

Run the SCM Optimization

Socio-cultural Cognitive Mapping

Results

Optimize the models and view the results.

Click on a list result below to add it to ORA.

Minkowski	Attenuation	Repetitions	X^2 - Min	X^2 - Stddev	Dispersion	Z-Score

Progress...

Calculating model:

Optimizing...

Stop

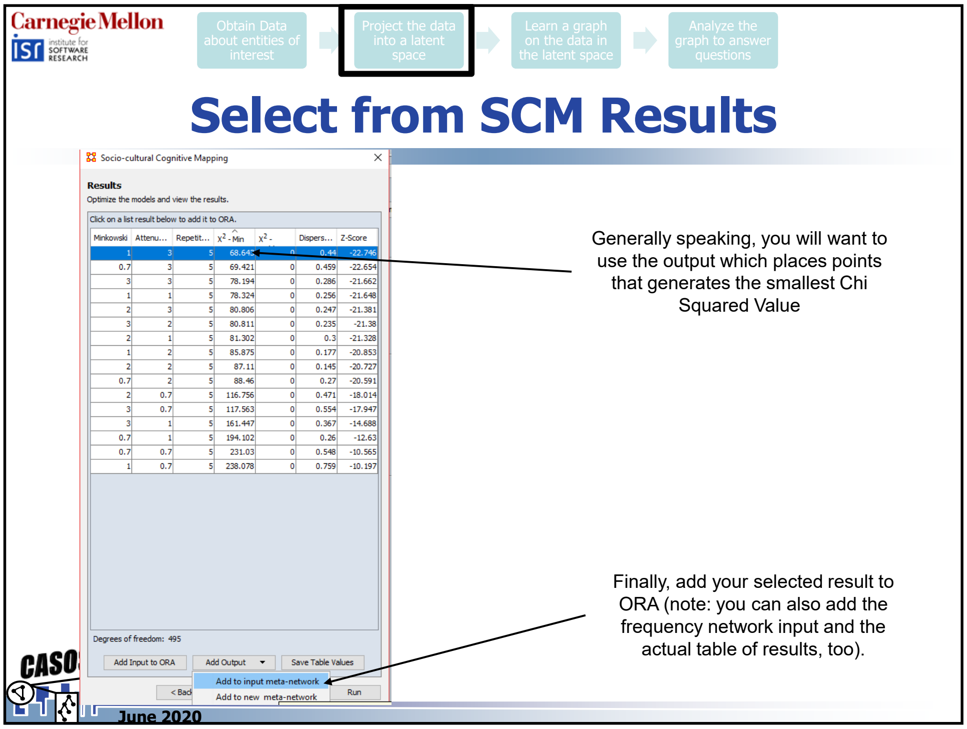
Degrees of freedom:

Add Input to ORA | Add Output to ORA | Save Table Values

< Back | Next > | Close | Run

CASOS

June 2020



Obtain Data about entities of interest → Project the data into a latent space → Learn a graph on the data in the latent space → Analyze the graph to answer questions

Select from SCM Results

Socio-cultural Cognitive Mapping

Results

Optimize the models and view the results.

Click on a list result below to add it to ORA.

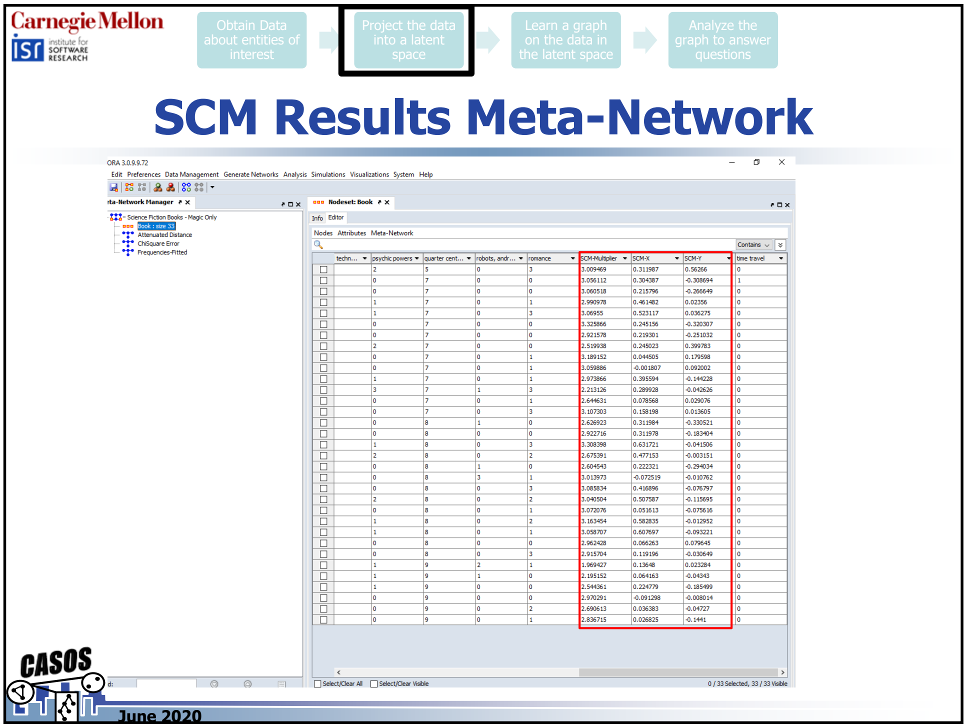
Minkowski	Attenu...	Repetit...	X^2 - Min	X^2 -	Dispers...	Z-Score
1	3	5	68.643	0	0.44	-22.746
0.7	3	5	69.421	0	0.459	-22.654
3	3	5	78.194	0	0.286	-21.662
1	1	5	78.324	0	0.256	-21.648
2	3	5	80.806	0	0.247	-21.381
3	2	5	80.811	0	0.235	-21.38
2	1	5	81.302	0	0.3	-21.328
1	2	5	85.875	0	0.177	-20.853
2	2	5	87.11	0	0.145	-20.727
0.7	2	5	88.46	0	0.27	-20.591
2	0.7	5	116.756	0	0.471	-18.014
3	0.7	5	117.563	0	0.554	-17.947
3	1	5	161.447	0	0.367	-14.688
0.7	1	5	194.102	0	0.26	-12.63
0.7	0.7	5	231.03	0	0.548	-10.565
1	0.7	5	238.078	0	0.759	-10.197

Degrees of freedom: 495

Generally speaking, you will want to use the output which places points that generates the smallest Chi Squared Value

Finally, add your selected result to ORA (note: you can also add the frequency network input and the actual table of results, too).

June 2020



Obtain Data about entities of interest → Project the data into a latent space → Learn a graph on the data in the latent space → Analyze the graph to answer questions

SCM Results Meta-Network

ORA 3.0.9.9.72
Edit Preferences Data Management Generate Networks Analysis Simulations Visualizations System Help
eta-Network Manager
Science Fiction Books - Magic Only
Book : size 33
Attenuated Distance
ChiSquare Error
Frequencies-Fitted
Nodeset: Book
Info Editor
Nodes Attributes Meta-Network
Contains

	techn...	psychic powers	quarter cent...	robots, andr...	romance	SCM-Multiplier	SCM-X	SCM-Y	time travel
☐		2	5	0	3	3.009469	0.311987	0.56266	0
☐		0	7	0	0	3.056112	0.304387	-0.308694	1
☐		0	7	0	0	3.060518	0.215796	-0.266649	0
☐		1	7	0	1	2.990978	0.461482	0.02356	0
☐		1	7	0	3	3.06955	0.523117	0.036275	0
☐		0	7	0	0	3.325866	0.245156	-0.320307	0
☐		0	7	0	0	2.921578	0.219301	-0.251032	0
☐		2	7	0	0	2.519938	0.245023	0.399783	0
☐		0	7	0	1	3.189152	0.044505	0.179598	0
☐		0	7	0	1	3.059886	-0.001807	0.092002	0
☐		1	7	0	1	2.973866	0.395594	-0.144228	0
☐		3	7	1	3	2.213126	0.289928	-0.042626	0
☐		0	7	0	1	2.644631	0.078568	0.029076	0
☐		0	7	0	3	3.107303	0.158198	0.013605	0
☐		0	8	1	0	2.626923	0.311984	-0.330521	0
☐		0	8	0	0	2.922716	0.311978	-0.183404	0
☐		1	8	0	3	3.308398	0.631721	-0.041506	0
☐		2	8	0	2	2.675391	0.477153	-0.003151	0
☐		0	8	1	0	2.604543	0.222321	-0.294034	0
☐		0	8	3	1	3.013973	-0.072519	-0.010762	0
☐		0	8	0	3	3.085834	0.416896	-0.076797	0
☐		2	8	0	2	3.040504	0.507587	-0.115695	0
☐		0	8	0	1	3.072076	0.051613	-0.075616	0
☐		1	8	0	2	3.163454	0.582835	-0.012952	0
☐		1	8	0	1	3.058707	0.607697	-0.093221	0
☐		0	8	0	0	2.962428	0.066263	0.079645	0
☐		0	8	0	3	2.915704	0.119196	-0.030649	0
☐		1	9	2	1	1.969427	0.13648	0.023284	0
☐		1	9	1	0	2.195152	0.064163	-0.04343	0
☐		1	9	0	0	2.544361	0.224779	-0.185499	0
☐		0	9	0	0	2.970291	-0.091298	-0.008014	0
☐		0	9	0	2	2.690613	0.036383	-0.04727	0
☐		0	9	0	1	2.836715	0.026825	-0.1441	0

☐ Select/Clear All ☐ Select/Clear Visible

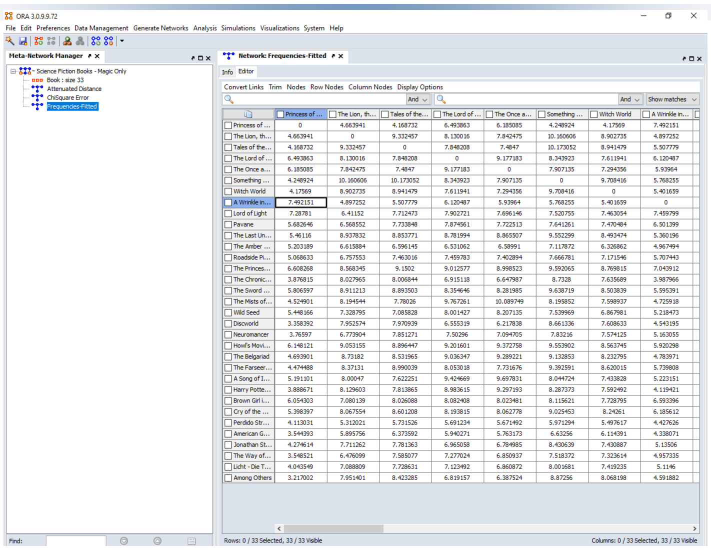
0 / 33 Selected, 33 / 33 Visible

June 2020

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Obtain Data about entities of interest → Project the data into a latent space → Learn a graph on the data in the latent space → Analyze the graph to answer questions

SCM Results Meta-Network


ORA 3.0.9.9.72
File Edit Preferences Data Management Generate Networks Analysis Simulations Visualizations System Help
Meta-Network Manager
Network: Frequencies-Fitted
Info Editor
Convert Links Trim Nodes Row Nodes Column Nodes Display Options
And
And
Show matches
Find:

- Science Fiction Books - Magic Only
  - Book : size 33
  - Attenuated Distance
  - ChiSquare Error
  - Frequencies-Fitted

	□ Princess of ...	□ The Lion, th...	□ Tales of the...	□ The Lord of ...	□ The Once a...	□ Something ...	□ Witch World	□ A Wrinkle in...
□ Princess of ...	0	4.663941	4.168732	6.493863	6.185085	4.248924	4.17569	7.492151
□ The Lion, th...	4.663941	0	9.332457	8.130016	7.842475	10.160606	8.902735	4.897252
□ Tales of the...	4.168732	9.332457	0	7.848208	7.4847	10.173052	8.941479	5.507779
□ The Lord of ...	6.493863	8.130016	7.848208	0	9.177183	8.343923	7.611941	6.120487
□ The Once a...	6.185085	7.842475	7.4847	9.177183	0	7.907135	7.294356	5.93964
□ Something ...	4.248924	10.160606	10.173052	8.343923	7.907135	0	9.708416	5.768255
□ Witch World	4.17569	8.902735	8.941479	7.611941	7.294356	9.708416	0	5.401659
□ A Wrinkle in...	7.492151	4.897252	5.507779	6.120487	5.93964	5.768255	5.401659	0
□ Lord of Light	7.28781	6.41152	7.712473	7.902721	7.696146	7.520755	7.463054	7.459799
□ Pavane	5.682646	6.568552	7.733848	7.874561	7.722513	7.641261	7.470484	6.501399
□ The Last Un...	5.46116	8.937832	8.853771	8.781994	8.865507	9.552299	8.493474	5.360196
□ The Amber ...	5.203189	6.615884	6.596145	6.531062	6.58991	7.117872	6.326862	4.967494
□ Roadside Pi...	5.068633	6.757553	7.463016	7.459783	7.402894	7.666781	7.171546	5.707443
□ The Princes...	6.608268	8.568345	9.1502	9.012577	8.998523	9.592065	8.769815	7.043912
□ The Chronic...	3.876815	8.027965	8.006844	6.915118	6.647987	8.7328	7.635689	3.987966
□ The Sword ...	5.806597	8.911213	8.893503	8.354646	8.281985	9.638719	8.503839	5.595391
□ The Mists of...	4.524901	8.194544	7.78026	9.767261	10.089749	8.195852	7.598937	4.725918
□ Wild Seed	5.448166	7.328795	7.085828	8.001427	8.207135	7.539969	6.867981	5.218473
□ Discworld	3.358392	7.952574	7.970939	6.555319	6.217838	8.661336	7.608633	4.543195
□ Neuromancer	3.76597	6.773904	7.851271	7.50296	7.094705	7.83216	7.574125	5.163055
□ Howl's Movi...	6.148121	9.053155	8.896447	9.201601	9.372758	9.553902	8.563745	5.920298
□ The Belgariad	4.693901	8.73182	8.531965	9.036347	9.289221	9.132853	8.232795	4.783971
□ The Farseer...	4.474488	8.37131	8.990039	8.053018	7.731676	9.392591	8.620015	5.739808
□ A Song of I...	5.191101	8.00047	7.622251	9.424669	9.697831	8.044724	7.433828	5.223151
□ Harry Potte...	3.888671	8.129603	7.813865	8.983615	9.297193	8.287373	7.592492	4.119421
□ Brown Girl i...	6.054303	7.080139	8.026088	8.082408	8.023481	8.115621	7.728795	6.593396
□ Cry of the ...	5.398397	8.067554	8.601208	8.193815	8.062778	9.025453	8.24261	6.185612
□ Perdido Str...	4.113031	5.312021	5.731526	5.691234	5.671492	5.971294	5.497617	4.427626
□ American G...	3.544393	5.895756	6.373592	5.940271	5.763173	6.63256	6.114391	4.338071
□ Jonathan St...	4.274614	7.711262	7.781363	6.965058	6.784985	8.430639	7.430887	5.13506
□ The Way of...	3.548521	6.476099	7.585077	7.277024	6.850937	7.518372	7.323614	4.957335
□ Licht - Die T...	4.043549	7.088809	7.728631	7.123492	6.860872	8.001681	7.419235	5.1146
□ Among Others	3.217002	7.951401	8.423285	6.819157	6.387524	8.87256	8.068198	4.591882

Rows: 0 / 33 Selected, 33 / 33 Visible

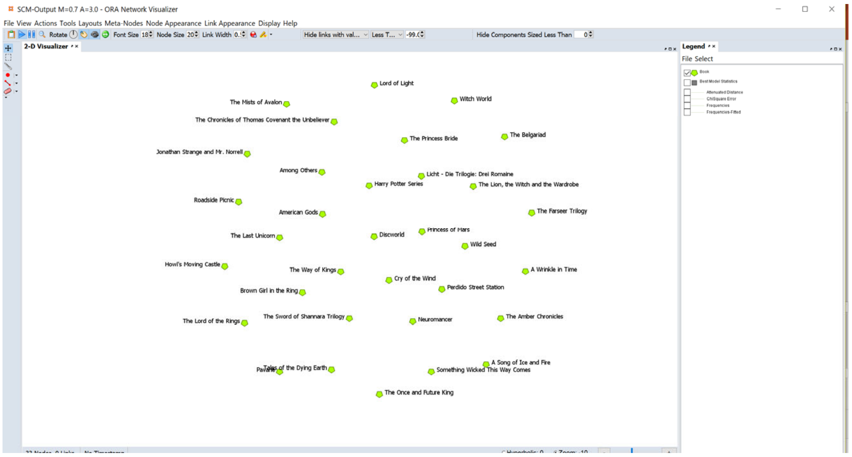
Columns: 0 / 33 Selected, 33 / 33 Visible

June 2020

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Obtain Data about entities of interest → Project the data into a latent space → Learn a graph on the data in the latent space → Analyze the graph to answer questions

Visualizing SCMs

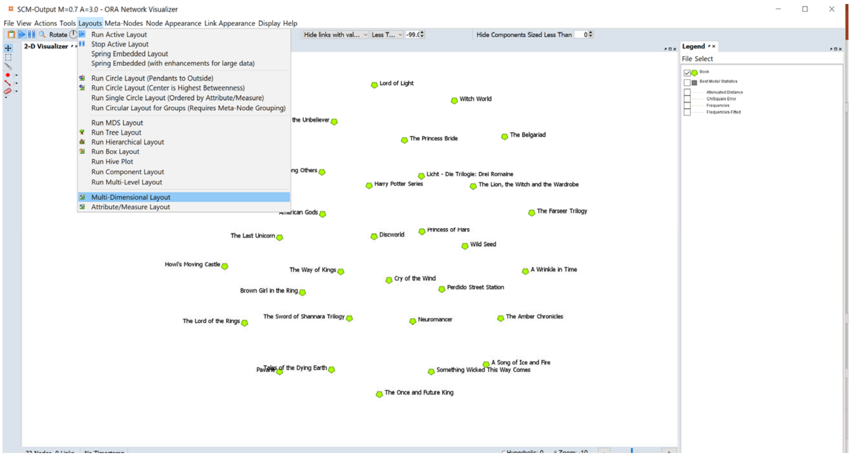


June 2020

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Obtain Data about entities of interest → **Project the data into a latent space** → Learn a graph on the data in the latent space → Analyze the graph to answer questions

Go to "Multi-Dimensional Layout"



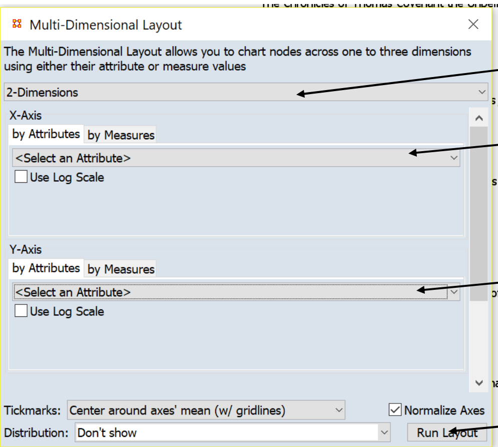
CASOS

June 2020

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Obtain Data about entities of interest → **Project the data into a latent space** → Learn a graph on the data in the latent space → Analyze the graph to answer questions

Configure the Layout



Multi-Dimensional Layout

The Multi-Dimensional Layout allows you to chart nodes across one to three dimensions using either their attribute or measure values

2-Dimensions

X-Axis
by Attributes | by Measures
<Select an Attribute>
□ Use Log Scale

Y-Axis
by Attributes | by Measures
<Select an Attribute>
□ Use Log Scale

Tickmarks: Center around axes' mean (w/ gridlines)
☑ Normalize Axes
Distribution: Don't show
Run Layout

We will visualize in 2-d, since we found spatial points in 2-d

Select 'SCM-X'

Select 'SCM-Y'

Select 'Run Layout'

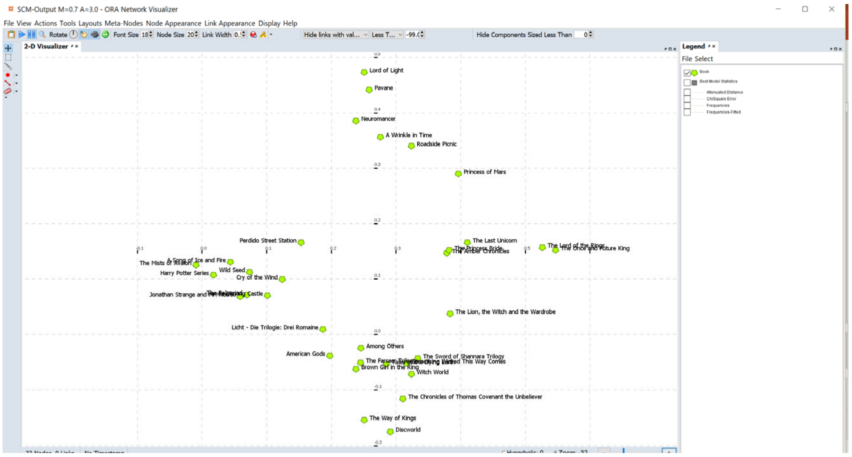
CASOS

June 2020

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Obtain Data about entities of interest → **Project the data into a latent space** → Learn a graph on the data in the latent space → Analyze the graph to answer questions

See Layout!

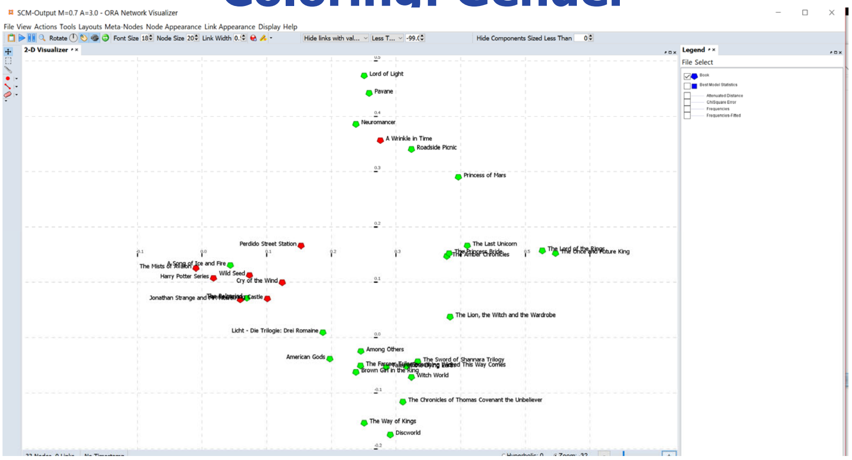


June 2020

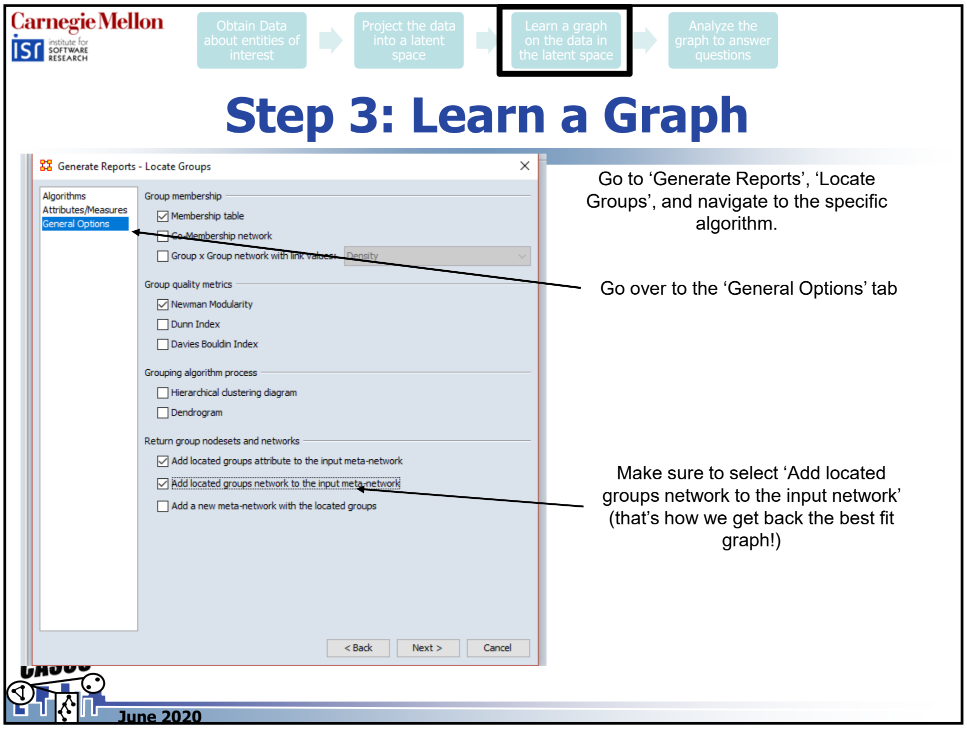
Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Obtain Data about entities of interest → **Project the data into a latent space** → Learn a graph on the data in the latent space → Analyze the graph to answer questions

Explore the Layout with Node Coloring: Gender



June 2020



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Obtain Data about entities of interest
Project the data into a latent space
Learn a graph on the data in the latent space
Analyze the graph to answer questions

Step 3: Learn a Graph

Generate Reports - Locate Groups

Algorithms
Attributes/Measures
General Options

Group membership

- ☑ Membership table
- □ Co-Membership network
- □ Group x Group network with link values: Density

Group quality metrics

- ☑ Newman Modularity
- □ Dunn Index
- □ Davies Bouldin Index

Grouping algorithm process

- □ Hierarchical clustering diagram
- □ Dendrogram

Return group nodesets and networks

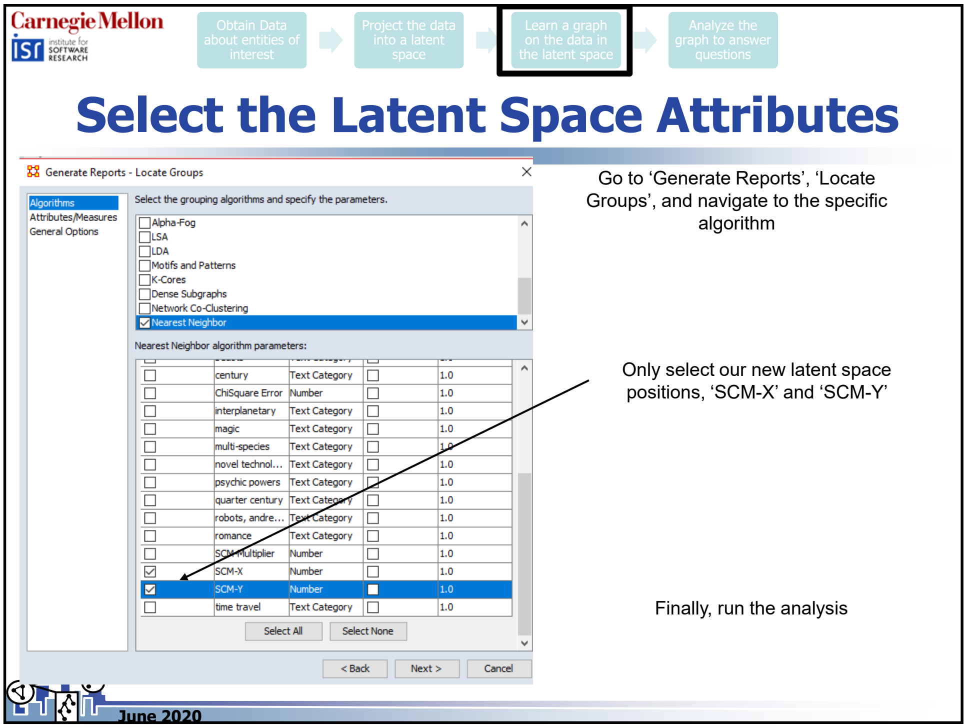
- ☑ Add located groups attribute to the input meta-network
- ☑ Add located groups network to the input meta-network
- □ Add a new meta-network with the located groups

Go to 'Generate Reports', 'Locate Groups', and navigate to the specific algorithm.

Go over to the 'General Options' tab

Make sure to select 'Add located groups network to the input network' (that's how we get back the best fit graph!)

June 2020



Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Obtain Data about entities of interest
Project the data into a latent space
Learn a graph on the data in the latent space
Analyze the graph to answer questions

Select the Latent Space Attributes

Generate Reports - Locate Groups

Algorithms
Attributes/Measures
General Options

Select the grouping algorithms and specify the parameters.

- □ Alpha-Fog
- □ LSA
- □ LDA
- □ Motifs and Patterns
- □ K-Cores
- □ Dense Subgraphs
- □ Network Co-Clustering
- ☑ Nearest Neighbor

Nearest Neighbor algorithm parameters:

□	century	Text Category	□	1.0
□	ChiSquare Error	Number	□	1.0
□	interplanetary	Text Category	□	1.0
□	magic	Text Category	□	1.0
□	multi-species	Text Category	□	1.0
□	novel technol...	Text Category	□	1.0
□	psychic powers	Text Category	□	1.0
□	quarter century	Text Category	□	1.0
□	robots, andre...	Text Category	□	1.0
□	romance	Text Category	□	1.0
□	SCM-Multiplier	Number	□	1.0
☑	SCM-X	Number	□	1.0
☑	SCM-Y	Number	□	1.0
□	time travel	Text Category	□	1.0

Go to 'Generate Reports', 'Locate Groups', and navigate to the specific algorithm

Only select our new latent space positions, 'SCM-X' and 'SCM-Y'

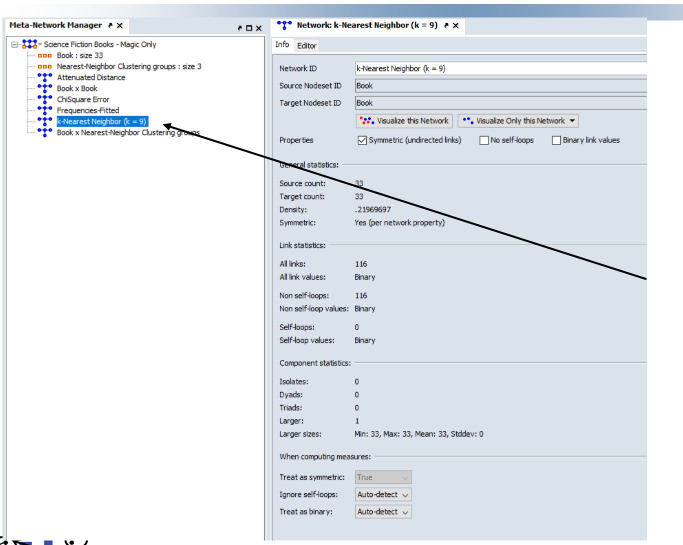
Finally, run the analysis

June 2020

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Obtain Data about entities of interest → Project the data into a latent space → Learn a graph on the data in the latent space → Analyze the graph to answer questions

Step 4: Analyze the Results



Meta-Network Manager

- Science Fiction Books - Magic Only
  - Book : size 33
  - Nearest-Neighbor Clustering groups : size 3
  - Attenuated Distance
  - Book x Book
  - ChiSquare Error
  - Frequencies-Fitted
  - k-Nearest Neighbor (k = 9)
  - Book x Nearest-Neighbor Clustering groups

Network: k-Nearest Neighbor (k = 9)

Info | Editor

Network ID: k-Nearest Neighbor (k = 9)
Source Nodeset ID: Book
Target Nodeset ID: Book

Visualize this Network | Visualize Only this Network

Properties: ☑ Symmetric (undirected links) ☐ No self-loops ☐ Binary link values

General statistics:

Source count: 33
Target count: 33
Density: .21969697
Symmetric: Yes (per network property)

Link statistics:

All links: 116
All link values: Binary

Non self-loops: 116
Non self-loop values: Binary

Self-loops: 0
Self-loop values: Binary

Component statistics:

Isolates: 0
Dyads: 0
Triads: 0
Larger: 1
Larger sizes: Min: 33, Max: 33, Mean: 33, Stddev: 0

When computing measures:

Treat as symmetric: True
Ignore self-loops: Auto-detect
Treat as binary: Auto-detect

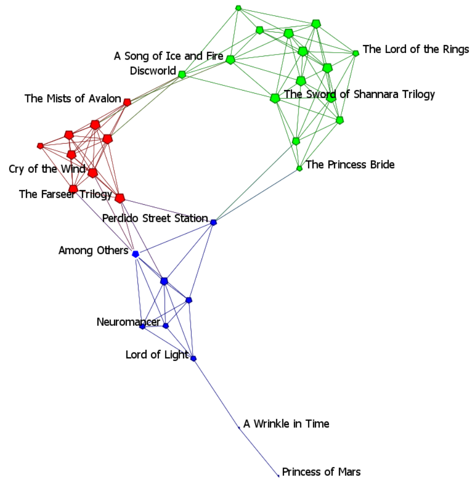
Now, we have learned the best fit k-NN graph for our data, using modularity as the means of determining the goodness of fit.

June 2020

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

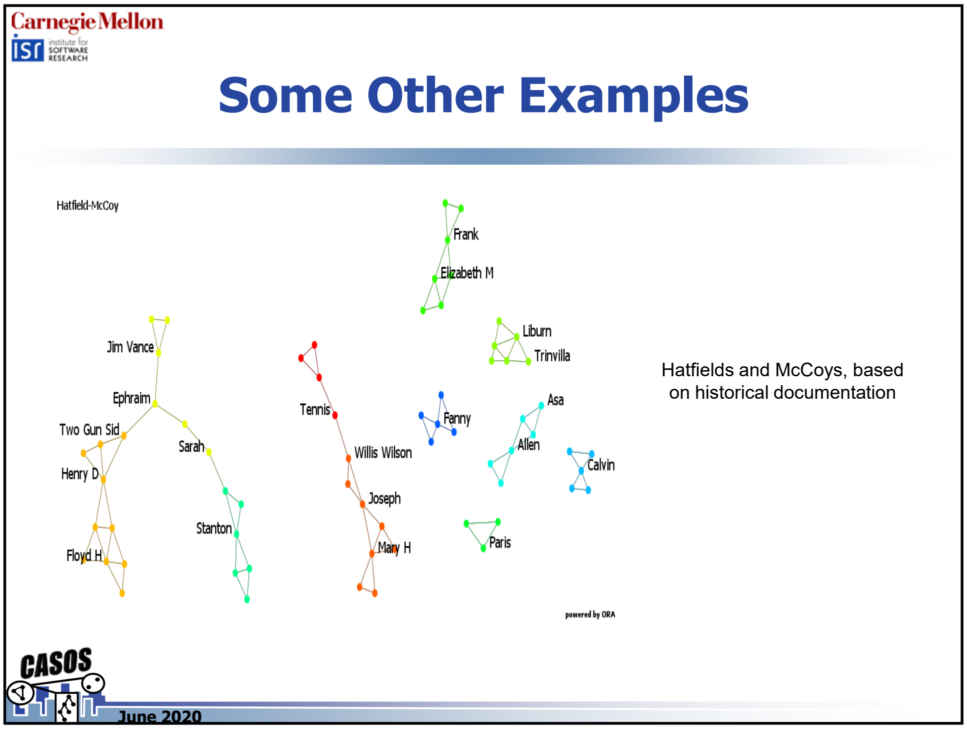
Obtain Data about entities of interest → Project the data into a latent space → Learn a graph on the data in the latent space → Analyze the graph to answer questions

Step 4: Analyze the Results



Node coloring by sub group.
Node size by degree centrality

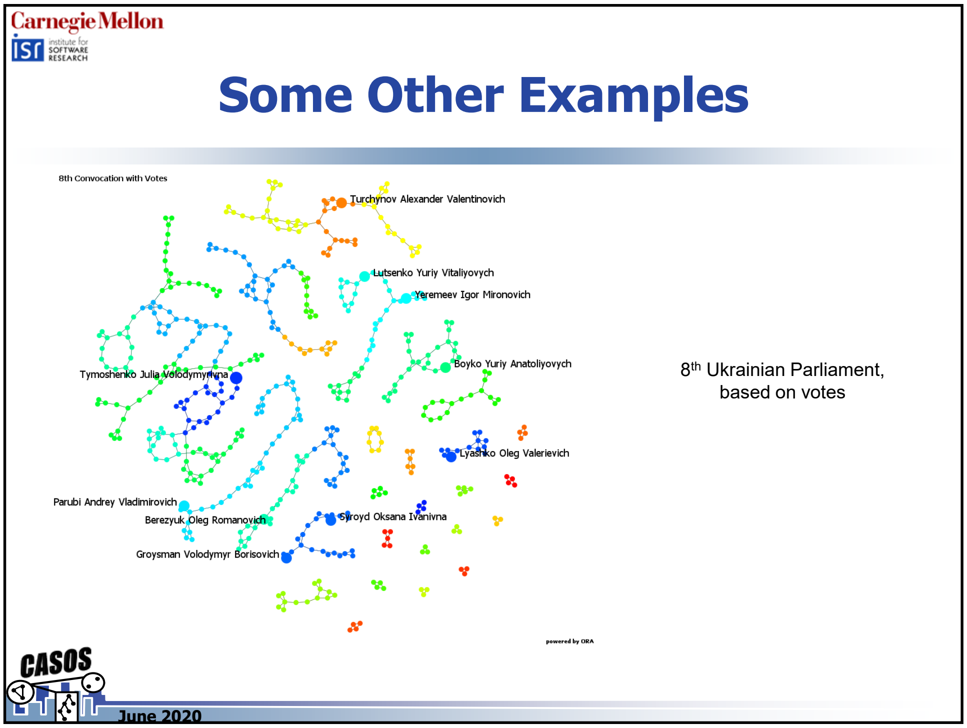
June 2020



Some Other Examples

Hatfields and McCoys, based on historical documentation

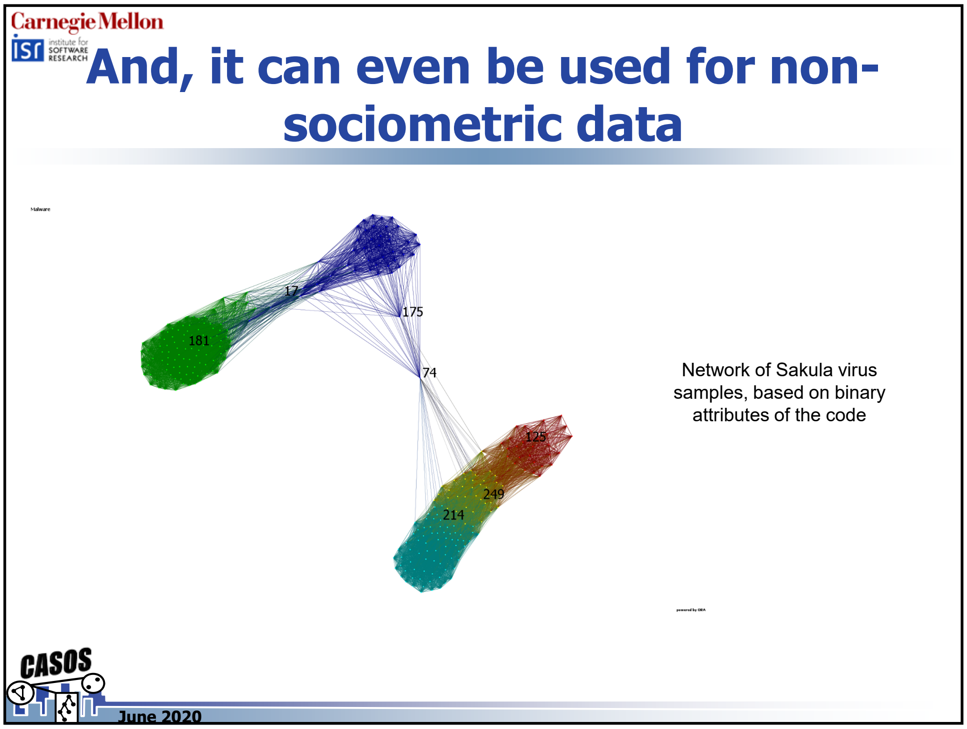
June 2020


Carnegie Mellon
isr institute for SOFTWARE RESEARCH
8th Convocation with Votes
Turchynov Alexander Valentinovich
Lutsenko Yuriy Vitaliyovych
Yeremeev Igor Mironovich
Boyko Yuriy Anatoliyovych
Tymoshenko Julia Volodymyrivna
Lyashko Oleg Valerievich
Parubi Andrey Vladimirovich
Berezyuk Oleg Romanovich
Syroyd Oksana Ivanivna
Groysman Volodymyr Borisovich
powered by ORA
CASOS

Some Other Examples

8th Ukrainian Parliament, based on votes

June 2020



Carnegie Mellon

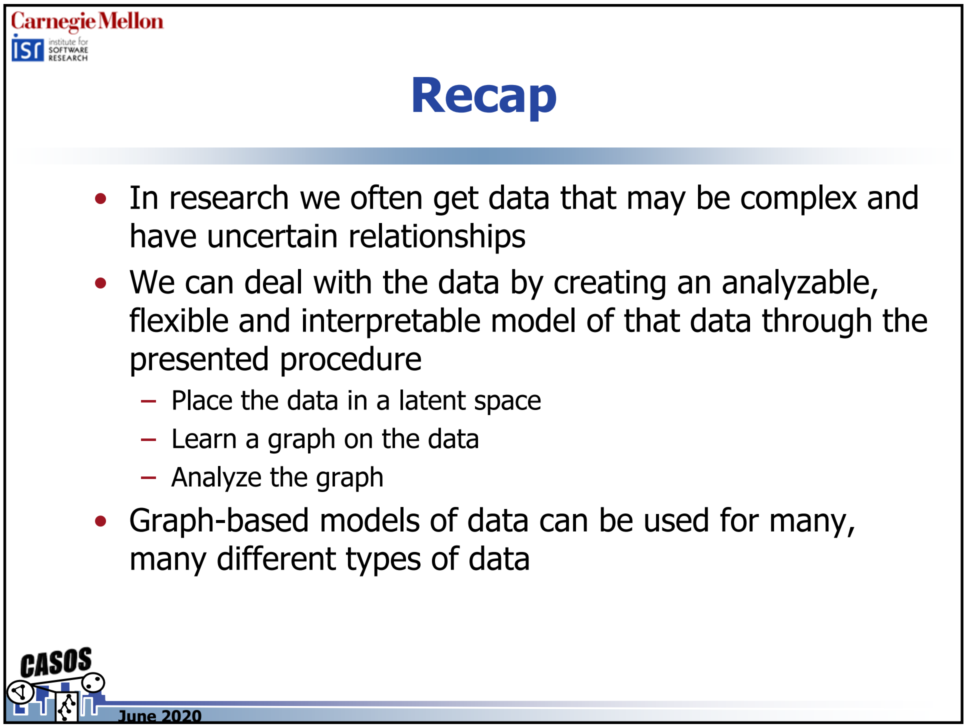
isr institute for SOFTWARE RESEARCH

And, it can even be used for non-sociometric data

Network of Sakula virus samples, based on binary attributes of the code

CASOS

June 2020



Carnegie Mellon

isr institute for SOFTWARE RESEARCH

Recap

- In research we often get data that may be complex and have uncertain relationships
- We can deal with the data by creating an analyzable, flexible and interpretable model of that data through the presented procedure
  - Place the data in a latent space
  - Learn a graph on the data
  - Analyze the graph
- Graph-based models of data can be used for many, many different types of data

CASOS

June 2020