31st Annual INCOSE international symposium
virtual event
July 17 - 22, 2021

Developing a Topic Network of Published Systems Engineering Research

Dr. Rudolph Oosthuizen

www.incose.org/symp2021

Scope

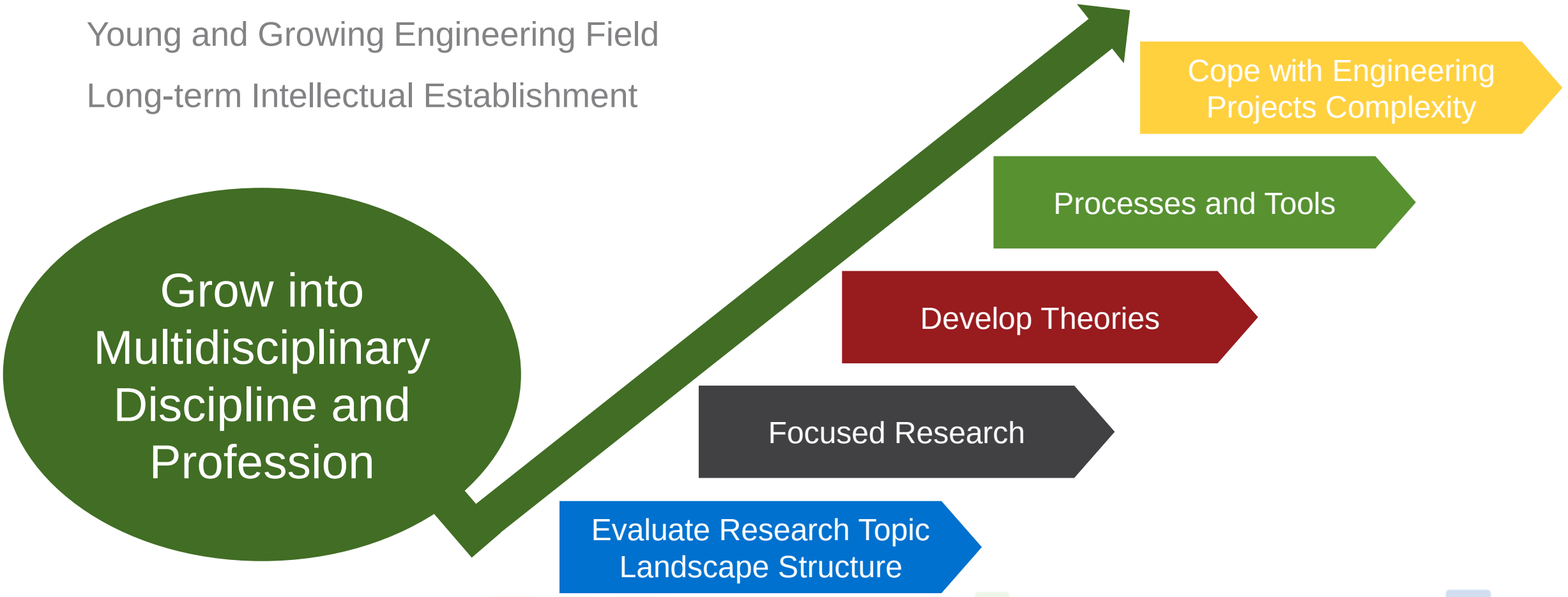
- Introduction
- Bibliometrics
- Natural Language Processing
- Network Analysis
- Method
- Results and Discussion

Systems Engineering Research

Formalization after the Second World War

Young and Growing Engineering Field

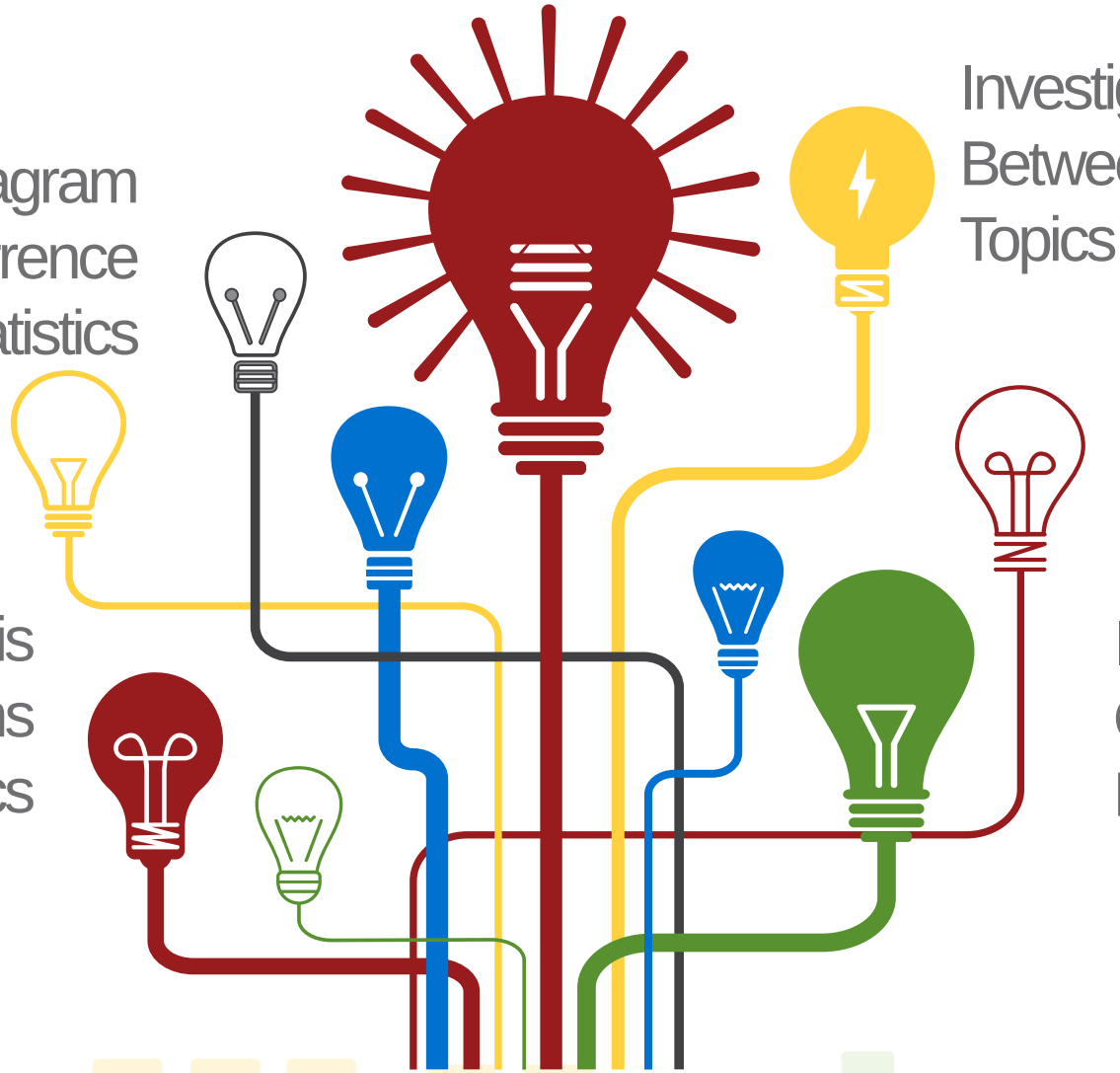
Long-term Intellectual Establishment



Analysis of Research Field

Construct Network Diagram of Topic Co-occurrence Statistics

Investigate Relationships Between Research Topics



Network Structure Analysis Reveal Connections between Research Topics

Identify Possible Opportunities and Research Focus Areas

Bibliometrics

- Quantitative and Systematic Statistical Analysis of Words in Text to Visualize Patterns and Trends
- Scientists Codify Research Outputs in Publications
  - Building Blocks of Science
  - Peer Reviews Provide Validation
- Suited to most research fields

Manual Topic Modeling

- Extract Topics from Published Text in Papers
- Assign to a Predetermined List
- Information Overload of Researcher Textual Data
- Bias and Miss Underlying Latent Topics

Automated Topic Modeling

- Quantitative, Statistical and Unsupervised Machine-learning
- Automated Semantic Extraction of Latent Topics from an Extensive set of Document
- Requires no Training Data or Labels

Natural Language Processing

- Machine-learning Algorithms Enables Automated Text Processing
- Extract Keywords, Relationships between Words and Text Clusters
- Categorize, Summarize or Classify Documents

Latent Dirichlet Allocation (LDA)

- Assumes Mixture of Words Contains Latent Topics
- Topic has a Probability Distribution over Words
- Identifies Thematic Clusters from the Text with Probability Distribution
- More Comprehensive and Faster than Manual Method

Network Analysis

- Investigate Network Structure (Nodes and Edges)
- Interconnection amongst Semantic Research Topics
- Visualize for the Qualitative Assessment
- Quantify Network Parameters using Graph Theory Concepts
  - Centrality, Distance between Nodes, Number of Interactions, Ranking of Nodes

Method

Capture Bibliometric Data

Prepare Text

Topic Modelling

Process Results

Analyze Network

Data Capture

- Basic Search Terms: "Systems Engineering“
- Retrieve 68 572 Documents from 1938 - 2021
- Limit to 1990 Onwards
- Remove Documents with Incomplete Information
- 62 136 Papers Remained for Processing
- Guideline: 1000–2000 docs of 100–200 words

Data Preparation

- Combine titles and abstracts
- Clean Text
- Prepare Text
  - Abbreviations
  - Lowercase
  - Tokenize and Lemmatize
  - Remove Stop Words
  - Remove Irrelevant Words

Topic Modelling

- Vectorize Text into a Document Term Matrix
- LDA Algorithm
  - Required Number of Topics (with Terms)
  - Term Importance per Topic
  - Probability of each Topic per Document
  - Perplexity of the Topic Model
- Topic Number set at 25 for this Research

Post Processing

- Identify Topics
- Distribution Probability of Each Topic per Paper
  - Total Distribution per Topic over Text Corpus
- Network:
  - Topics are Nodes
  - Topic Co-Occurrence Probabilities are Edges

Previous Work

- INCOSE Systems Engineering Journal (622)
- INCOSE International Symposiums (3 694)
- Controlled Sample: 20 Systems Engineering Topics
- Experts Identified and Validated the Topics

Repeated Topics

- Architecting
- Modelling
- Integration
- SE Methodology
- System Operation
- SE Capability
- Project management
- SE Practice
- System Cost
- SE Processes
- Decision Support
- System Life Cycle
- Software Engineering
- Complexity
- Requirements
- Product development
- Design
- Education

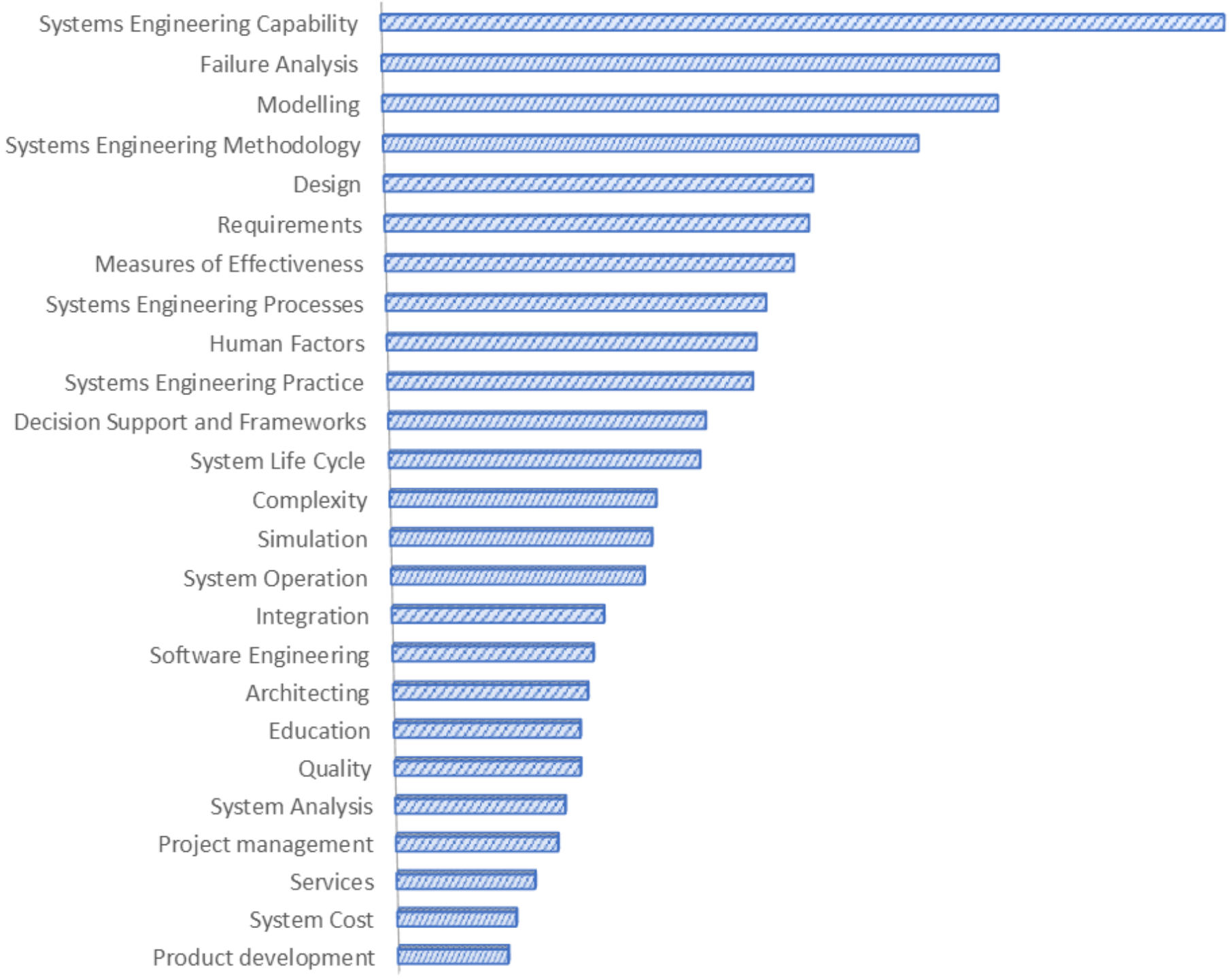
New Topics

- Quality
- Failure Analysis
- Measures of Effectiveness
- Human Factors
- Services
- System Analysis
- Simulation

Missing Topics

- Risk management
- SE Management

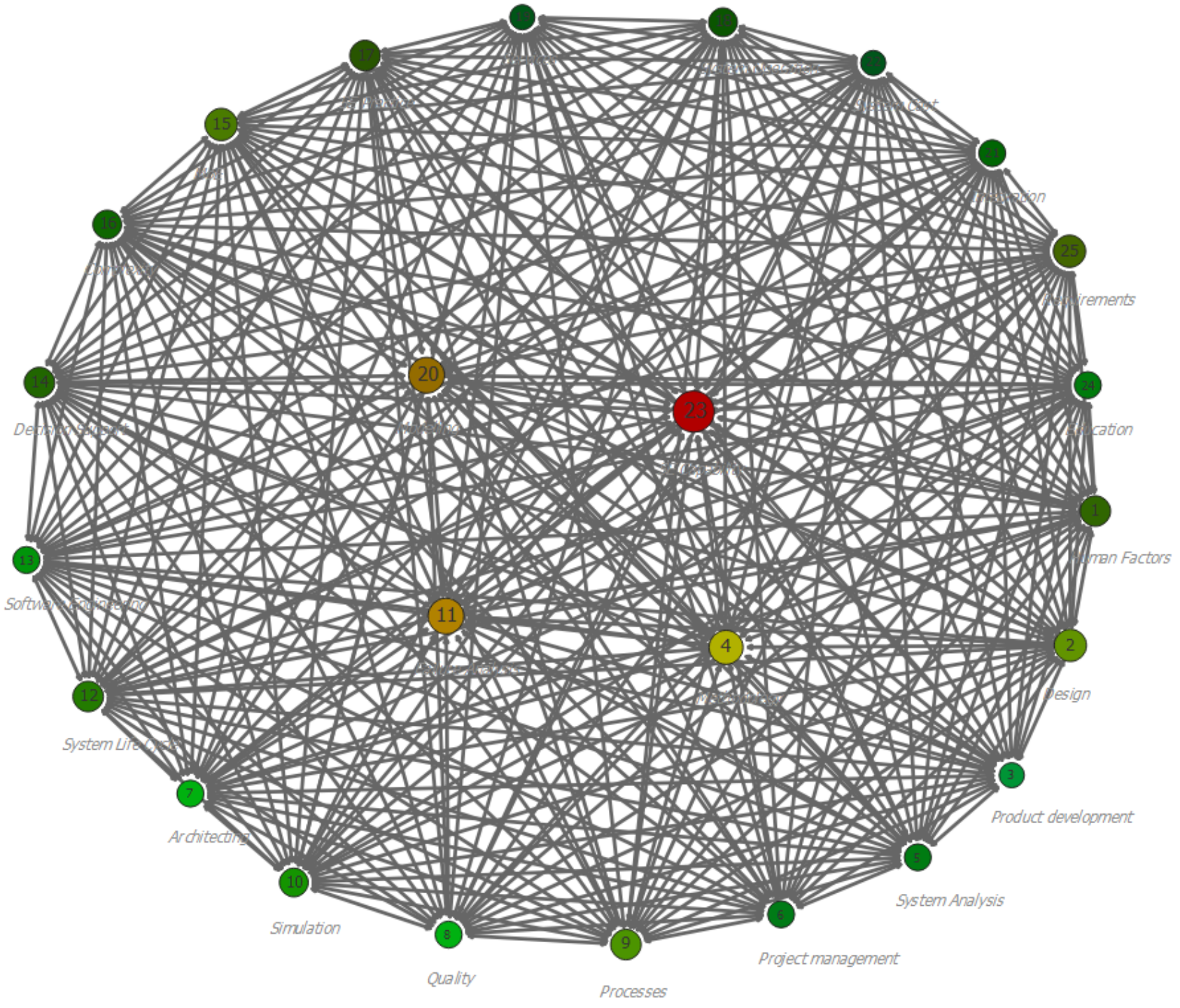
Topic Promi-nence



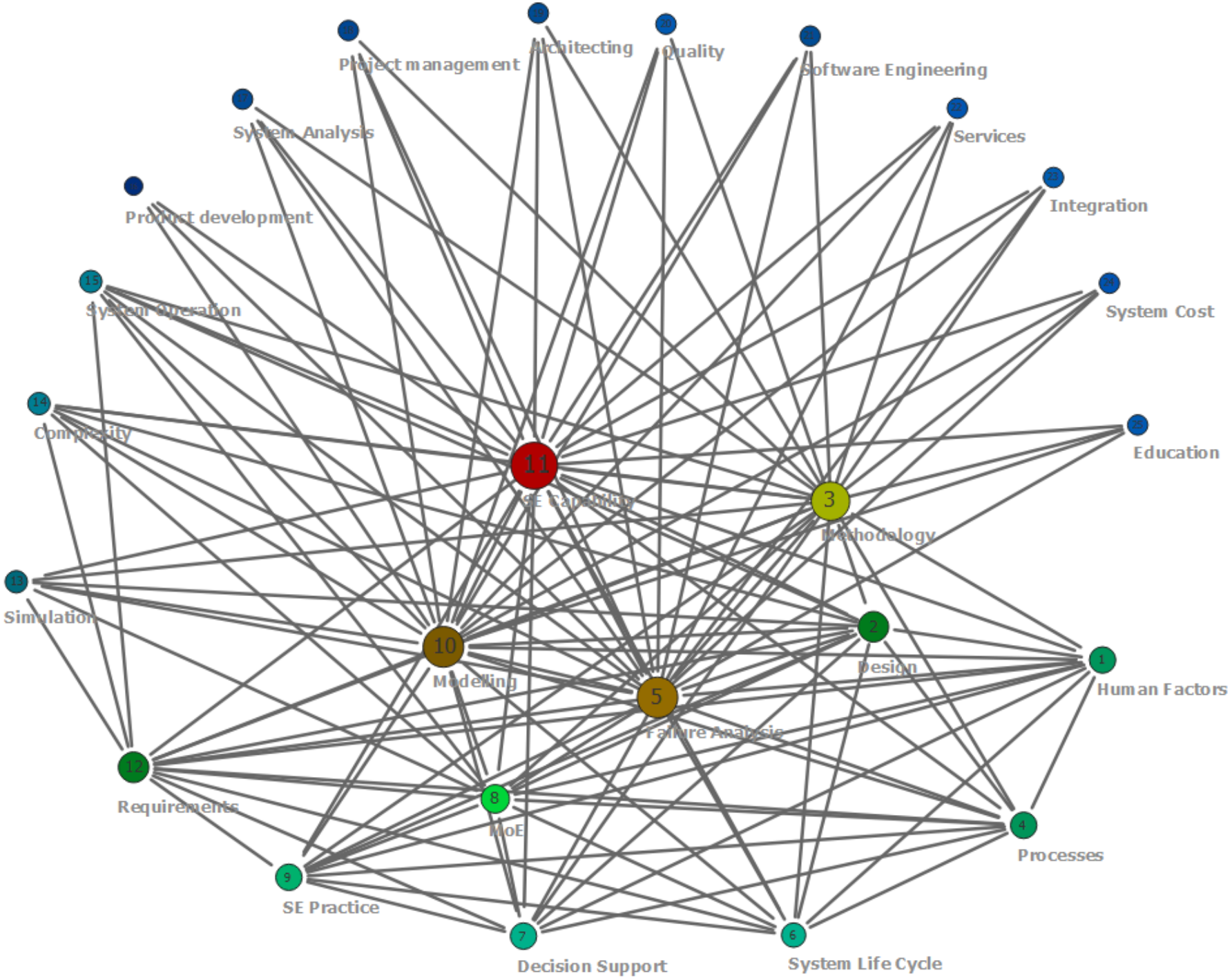
Topic Network

- Sum Co-occurrence Probabilities of Topics per Paper
- N2 Matrix of Nodes and Edges
- Every Node is Connected to all the Other Nodes in the Network
- Fully Connected or Mesh Network Topology
- Interdisciplinary Nature of Systems Engineering

Full Topic Network



50% Topic Network



Topic Network Analysis

- Node Strength
  - Sum of all Connections' Weighted Number and Strength
  - Indicates a Node's Strength Relative to other Nodes
- Degree Centrality
  - Sum of the Edges' Weights Attached to that Node
- Clusters and Cliques
  - Subset of Nodes Closer Connected to one Another than other Nodes

Topic Network Analysis

Node Name	Node Strength	Degree Centrality
SE Capability	6167	205515
System Analysis	4520	167412
Modelling	4517	167732
Methodology	3935	149002
Design	3161	90461
Requirements	3131	89884
MoE	3019	88458
Processes	2814	71368
Human Factors	2740	70283
SE Practice	2715	70330

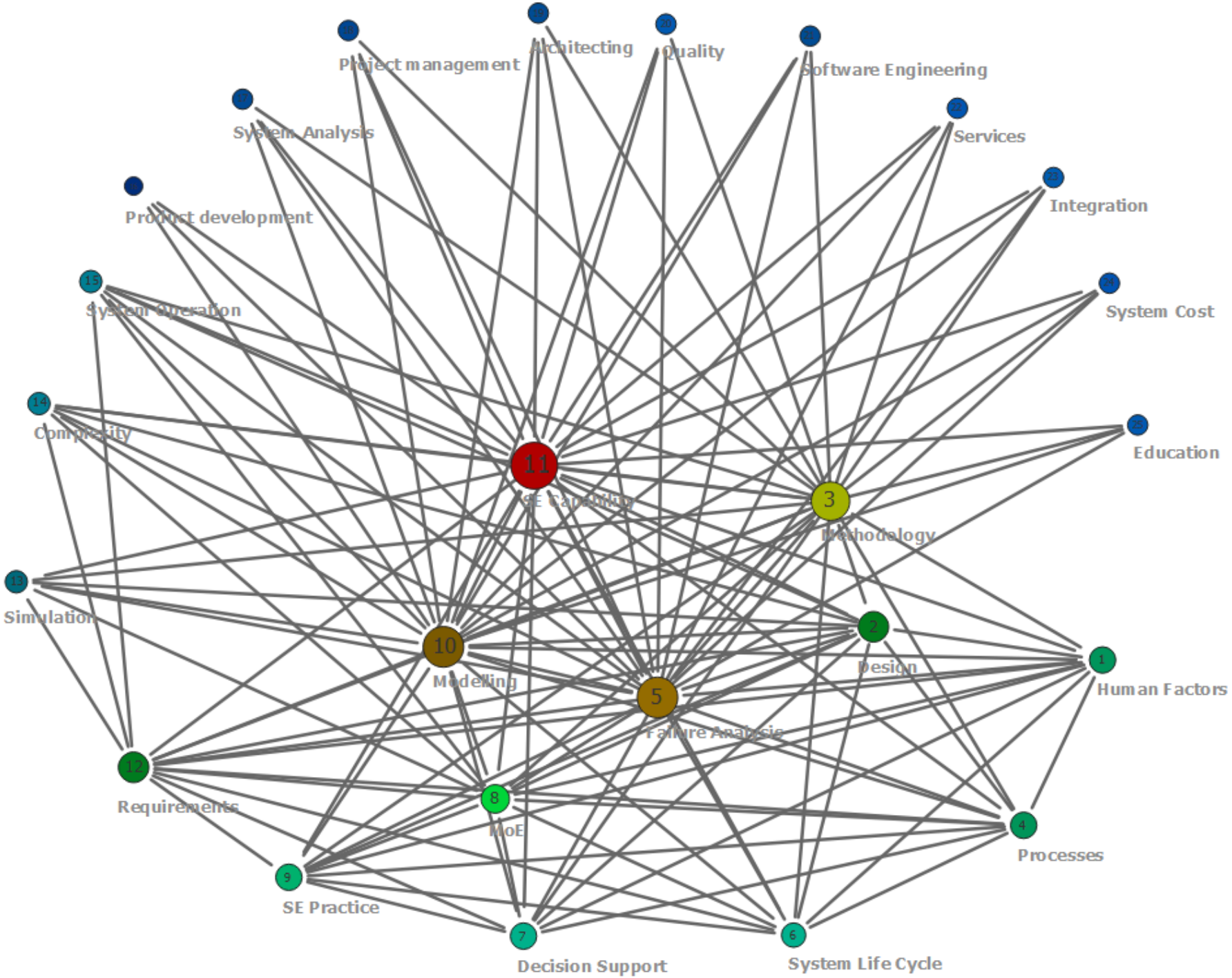
Topic Network Analysis

Node Name	Node Strength	Degree Centrality
Decision Support	2363	62038
System Life Cycle	2320	61226
Complexity	1990	43329
Simulation	1956	43274
System Operation	1897	42892
Integration	1591	26144
Software Engineering	1510	25816
Architecting	1465	25704
Education	1405	25417
Quality	1405	25417

Topic Network Cliques

- Product development, Modelling, SE Capability, Analysis
- System Analysis,
  - Modelling, SE Capability, Analysis, Methodology
- Architecting,
  - Modelling, SE Capability, Analysis, Methodology
- Project management,
  - Modelling, SE Capability, Analysis, Methodology

50% Topic Network



Topic Network Cliques

- Software Engineering
  - Modelling, SE Capability, Analysis, Methodology
- Quality
  - Modelling, SE Capability, Analysis, Methodology
- Integration
  - Modelling, SE Capability, Analysis, Methodology
- Services
  - Modelling, SE Capability, Analysis, Methodology

Topic Network Cliques

- Education
  - Modelling, SE Capability, Analysis, Methodology
- System Cost
  - Modelling, SE Capability, Analysis, Methodology
- Requirements, Simulation, MoE, Design
  - Modelling, SE Capability, Analysis, Methodology
- System Operation
  - Modelling, SE Capability, Analysis, Methodology
  - Requirements, MoE, Design

Topic Network Cliques

- Complexity
  - Modelling, SE Capability, Analysis, Methodology
  - Requirements, MoE, Design
- Decision Support, SE Practice, Processes, Human Factors
  - Modelling, SE Capability, Analysis, Methodology
  - Requirements, MoE, Design
- System Life Cycle, SE Practice, Processes, Human Factors
  - Modelling, SE Capability, Analysis, Methodology
  - Requirements, MoE, Design

Conclusion

- Good Comparison of Topics to Previous Work
- Developed Topic Network Diagram
- Simplification Improved Visual Interpretation
- Network Analysis Identified Centrality Similarities
- SE Researchers Need to Identify Topic Building Blocks and Interdependencies

31st Annual INCOSE international symposium

virtual event

July 17 - 22, 2021

www.incose.org/symp2021