IMPROVING ACTUARIAL PROCESSES WITH GRAPH THEORY AND NETWORK ANALYSIS

June 2020

Written by

Ankur Jain
Senior Engagement Manager

Pritam Banerjee
Engagement Manager

lookdeeper@exlservice.com
INTRODUCTION

Most traditional insurers find it overwhelming to transform the innumerable sensitive actuarial processes needed for day-to-day functioning. This problem is further amplified by most insurance actuaries spending most of their time on secondary activities, such as data reconciliation, rather than focusing on core actuarial tasks such as modelling, strategy development, and root cause analysis. These secondary activities are usually low value, repeatable, and time-consuming tasks.

It’s crucial to understand that unlike other insurance processes, actuarial processes are complex in nature, time consuming, and have a high number of touchpoints. Dynamic, frequently changing regulations can make these processes even more complicated. For instance, the New York Department of Financial Services (NYDFS) published its Circular Letter Number 1 in 2019 on the use of big data in underwriting life insurance.

In the document, the NYDFS states that “an insurer should not use external data sources, algorithms or predictive models in underwriting or rating unless the insurer has determined that the processes do not collect or utilize prohibited criteria and that the use of the external data sources, algorithms or predictive models are not unfairly discriminatory.”

This presents a need for full transparency to explain the variables computed and their effects, as well as a need for efficiency so that actuaries spend their time on analysis rather than data reconciliation. Other priorities will depend on the processes. For example, pricing and ALM modeling processes require greater flexibility and transparency, whereas valuation and economic projection models require more precision and prioritize governance over flexibility and transparency.

Irrespective of the modeling processes, legacy source systems, fragmented data, error-prone manual processes, and a lack of data standardization lead to problems within actuarial organizations. Analyzing actuarial processes is quite complex due to the interdependencies and relationships of its subtasks and files. With advancements in the field of artificial intelligence (AI) and machine learning (ML), copious amounts of data can be processed quite efficiently to identify hidden patterns. Network analysis is widely used in other domains to analyze different elements of a network. Within insurance, it can be applied for fraud detection and marketing. This paper describes an approach where network analysis is leveraged for actuarial process transformation.

An Upcoming Science: Graphs and Network Analysis

Graph and network analysis helps organizations gain a deep understanding of their data flows, process roadblocks, and other trends and patterns. The first step for graph and network analysis involves using tools to develop visual representations of data in order to better understand the data. The next step comprises of acting on this data, typically by carefully analyzing graph network parameters such as centrality, traversal, and cycles.

A graph is a data structure used to show pairwise relationships between entities. It consists of a set of vertices (V) and a set of edges (E). The vertices of a graph represent entities, such as persons, items, and files, and the edges represent relationships among vertices.
Graphs can be directed or undirected. An undirected graph (Figure 1) is where there is a symmetric relationship between nodes (A to B implies B to A), whereas a directed graph (Figure 2) is asymmetric in nature. In the case of process improvements, the dependencies of one task or file with the others in the process need to be modeled. The nature of the relationship is asymmetric, and therefore should be modeled through a directed graph.

Similarly, cycles, cliques, and paths can be traced along the network to optimize flow. Network analysis helps assess the current state of processes to identify gaps or redundancies, and determine which processes provide maximum value. Three key analyses are the most important in any process improvement framework:

1. Identifying process and data nodes which are crucial in the network
2. Tracing from the input to the output in the processes to identify touchpoints
3. Identifying cyclical references and dependencies in the network and linearize the flow

**1. Influential Nodes: Centrality**

Centrality measures the influence of a node in a network. As a node’s influence can be viewed differently, the right choice of centrality measures will depend on the problem statement.

**Degree Centrality:** Degree centrality measures influence based on the number of incoming and outgoing connections of a node. For a directed network, this can be further broken down into in-degree centrality for incoming connections, and out-degree centrality for outgoing connections.

**Between-ness Centrality:** Between-ness centrality measures the influence of a node over the information flow of a network. It assumes that the information flows through the shortest path, and captures the number of times a particular node appears in that path.

These different centrality measures can be used to derive insights about a network. While degree centrality defines strength as the number of neighbors, betweenness centrality defines strength as control over information passing between other neighbors through the node. Nodes that are high in both degree are the influential nodes in the network.

**Network Analysis Basics and Process Improvements**

Graphs provide a better way of dealing with the dependencies in the various data files, data systems, and processes. Once any process is represented as a graph, there are multiple operations and analyses can be performed. For instance, influencer nodes can be easily identified using centrality measures.
2. Graph Traversal
Graph traversals are used to understand the flow within the network. It is used to search for nodes within a network by passing through each of the nodes of the graph. Traversals can be made to identify the shortest path, or to search for connected vertices in a graph. The latter is of particular importance for making actuarial process improvements. Understanding the path of data throughout the process can help evaluate the process holistically and identify improvement opportunities.

3. Cliques and Cycles
A clique is a set of vertices in an undirected graph where every two distinct vertices are connected to each other. It is used to find communities in a network, and has varied applications in social network analysis, bioinformatics, and other areas. For process improvement, it finds an application in identifying local communities of processes and data. For directed graphs, finding cycles are of great importance in process improvement, as insights mined from investigating cyclical dependencies can be quite useful.

Step Approach for an Actuarial Transformation using Graph Theory

1. Understanding the Scope of Transformation
Understanding the scope of transformation is of key importance. The number of outputs touchpoints and files used by the organization is often significantly less than the number of files produced. Moreover, due to evolving regulations, actuarial processes can undergo changes. Some of the key questions to answer at this stage include:

- Which processes are in the scope of the transformation?
- Will these processes undergo changes in the near future due to regulations (US GAAP LDTI/IFRS 17)?
- Are all the tasks and files for the chosen process actually required, or is there a scope for rationalization?

2. Understanding Data Flows
Once the scope of the transformation is defined, data dependencies need to be traced. Excel links, database queries, and existing data models need to be analyzed. In some cases, manually copying and pasting the data creates breaks in the data flow. In such cases, the analyst needs to fill in the gaps and create the end to end flow of the data. Some key aspects to consider at this stage are:

- What are the data dependencies in the process?
- Are there breaks in the data flow due to manual adjustment?
- What are the inputs, outputs, and intermediate files?

3. Implementing the Network of Files
After mapping the data flow, the graph network can be constructed. The network can then be analyzed to identify potential opportunities, identify key files, linearize data flows, and create the goal state for the process. The key analysis to perform at this stage are:

- Identifying important nodes in the network through degree measures
- Capturing redundant intermedia files in the system
- Capturing cyclical-references and patterns in the process

Based on the analysis of the network, bottlenecks and inconsistencies can be easily identified. This information can lead to process reengineering and end-to-end data-based process transformation.

The results can be validated with business users, and changes can be made. The figures below show some of the patterns which can be captured using network graphs. The input, intermediate, and output nodes are color-coded as blue, grey, and red respectively.
The Benefits of Actuarial Process Transformation Using Graph Theory

Due to the inherent complexity of actuarial processes, decomposing process and data flows can be difficult to achieve. While analyzing any actuarial sub-process at the lowest level of granularity, it is quite possible to discover multiple interrelated files with lots of interrelated calculations. Moreover, a major challenge quite common in actuarial processes is a lack of data documentation. Graph theory enables insurers to overcome these challenges.

- **Creating a Data Lineage from Source System to Output**: Graph networks help improve the quality of data feeding into subsequent sub-processes. This benefits actuaries, as higher quality data produces better models regardless of the techniques being employed.

- **Improved Visualization and Bottleneck Identification**: Graph networks help visualize the relationship between various databases. It also helps build a foundation for a data factory that not only creates a 360-degree view of useful information, enables data visualization, and enables future self-service analytics. Moreover, several analyses can identify process bottlenecks which can be investigated further.

- **Enabling Flexibility and Governance**: On the surface, flexibility and governance may sound like competing priorities. Increased flexibility makes it difficult to control what is happening in the process and leads to increased security risks. However, graph theory helps manage governance by visualizing complicated data relationships and helps in maintaining data integrity.

- **Speed of Analysis**: Traditionally, most of the time spent producing model is used to gather, clean, and manipulate data. Graph theory helps in driving dependencies, enabling efficient processes, and providing quicker results for a given problem. Graph theory can be used to rationalize non-value adding files or processes, leading to streamlined and automated process flows. By linking the data elements from outputs to source systems, organizations can analyze processes in depth through backpropagation.
Case Example – How EXL Leveraged Network Analysis to Identify Hidden Patterns in the Annuities Valuation Process for a Major US Life Insurance Carrier

A major life insurance player in the US engaged EXL to examine its annuities valuation process and identify process improvement opportunities. There were multiple interfaces in the annuities valuation process, and many stakeholders were involved. Regulatory frameworks, a high number of touchpoints, actuarial judgment, and manual adjustments made the annuities valuation process complex. Moreover, the client had multiple source systems from which data were pulled. Data came to the actuarial team through SQL servers, data warehouses, Excel, Access databases, and flat files. As a result of the data fragmentation, a significant amount of effort was spent on data reconciliation, data validation, and data pulls. While some aspects of these steps were automated, many of the processes were manually intensive, wasting actuarial bandwidth.

EXL deployed a two-speed approach, tackling the problem from a short term local optimization as well as from a long-term process improvement perspective. The local optimization approach focused on understanding the standard operating procedures for the individual tasks in order to automate the manual efforts. These optimizations generated quick wins, but did not address the overall efficiency and improvement goals per se.

Knowing that there was a possibility of finding multiple tasks that can be rationalized, EXL prioritized and balanced the local and long-term improvements. This included speaking to multiple stakeholders to identify the regulatory GAAP processes for deferred annuities that needed to be focused on in the long term, and what the other processes could be addressed through local optimization.

For the deferred annuities GAAP process, EXL leveraged network analysis to analyze the file dependencies. Each of the hundreds of process files and tasks were categorized into pure inputs, outputs, and intermediates. These files were modeled as nodes in the network, while the data flows were modeled as edges. In order to capture the data linkages, a Visual Basic Macro (VBM)-based tool was deployed which automatically identified the Excel links and formulae to capture dependencies. Centrality measures were calculated for each of the files and then attached to the node attributes. The centrality measures showed important sub-processes and communities of files. For example, the topside sub-processes ingested more than 20 files and were high on degree centrality. Annual reporting sub-processes were high on degree centrality.

The team also found 11 avoidable cyclical references for data flows. These data flows were linearized to create the goal process state. Moreover, it was also observed that some of the intermediate files were merely being used to stage the data. These files had basic data checks embedded, but did not add a lot of value. These files were rationalized. Network analysis helped in providing an understanding of the data flows and creating the to-be state for process improvement. Moreover, the time required to analyze hundreds of tasks and files was reduced significantly. The team was able to identify an over 30% reduction in effort through a combination of automation and data-based solutions.

References
EXL (NASDAQ: EXLS) is a leading operations management and analytics company that helps our clients build and grow sustainable businesses. By orchestrating our domain expertise, data, analytics and digital technology, we look deeper to design and manage agile, customer-centric operating models to improve global operations, drive profitability, enhance customer satisfaction, increase data-driven insights, and manage risk and compliance. Headquartered in New York, EXL has more than 32,600 professionals in locations throughout the United States, the UK, Europe, India, the Philippines, Colombia, Australia and South Africa. EXL serves multiple industries including insurance, healthcare, banking and financial services, utilities, travel, transportation and logistics, media and retail, among others.

For more information, visit www.exlservice.com.

© 2020 ExlService Holdings, Inc. All Rights Reserved.