# Generating Synthetic Event Datasets for Root Cause Analysis Algorithm Tuning

D. Gueorguiev 7/19/23

## Notation

- directed graph

- the vertex set of the directed graph

– the arc set of the directed graph

– event type for which we would like to do root cause analysis

– set of event types associated with the event subject to analysis

– set of all event types

- the set of all instances of associated events of types in

Let us denote by the set of the event types which are relevant in root cause analysis of specific event type .

That is, the event types and the event type will form causal pairs for which we want to calculate causal significance factor and construct Directed Causal Graph (DCG).

We will assume that we can have multiple instances of each event type . Here denote different sets of arguments for the same event type . We will impose *directly follows* () constraint to a subset of event types. We will impose reachable from () constraint to a subset of event types. Another relevant constraint is the multiplicity constraint: max sibling count, max total count within the dataset of associated events[[1]](#footnote-1).

We would like to construct

1. The subscript is from *association* (Greek **σ**υνεταιρισμός) [↑](#footnote-ref-1)