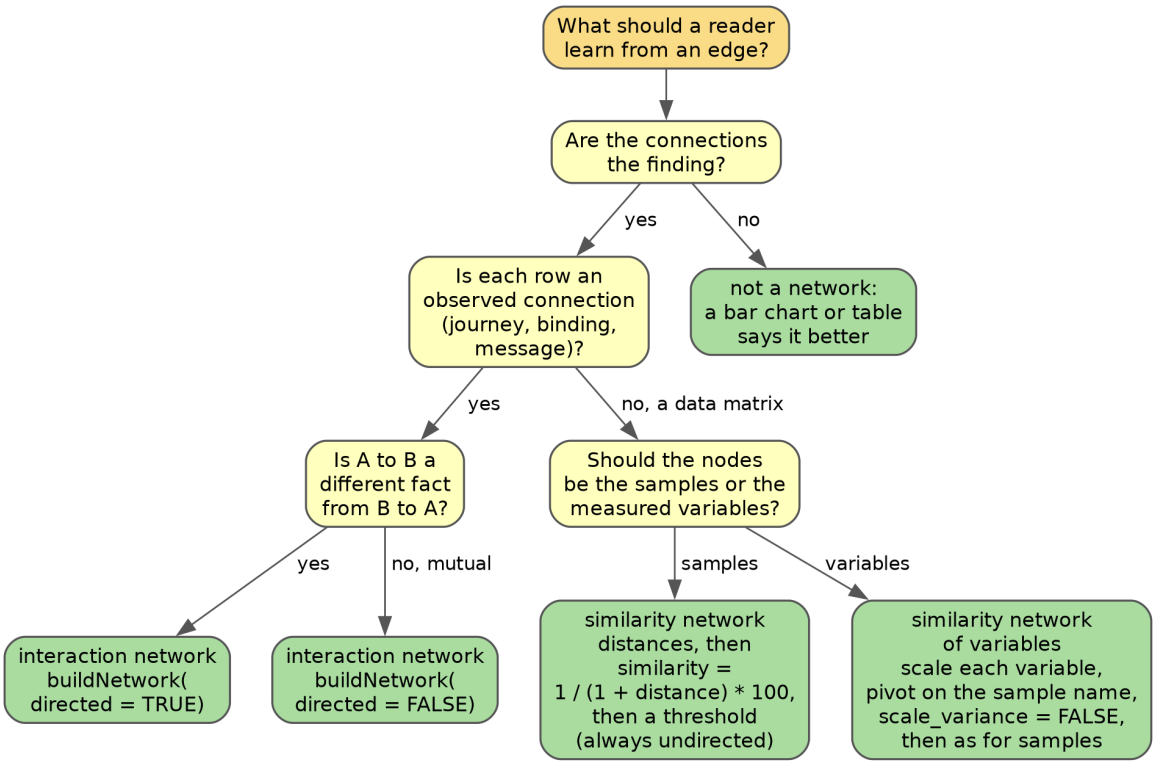
networks: summary

Integrated Bioanalytical Chemistry · thebustalab.github.io/integrated_bioanalytics



The first decision that must be made when working with networks is determining what an edge means. A **similarity network** is built: distances become similarities, and a chosen threshold decides which pairs are joined, so the network is always undirected and its shape depends on that threshold. An **interaction network** is given by the data themselves: each row is an observed connection, there is no threshold, and the edges may be directed. We use a metric called 'degree' to count a node's connections and a metric called 'strength' adds up a node's traffic. However, neither degree nor strength accounts for whether a node holds different portions of a network together (an articulation point) - only removing the node and counting the pieces that remain can tell us about that. Importantly, the positions of nodes in a force-directed layout carry no quantitative information.

Step	Code from this chapter	What to set
distances, one row per pair	`runMatrixAnalysis(analysis = "dist", output_format = "long")`	`scale_variance`, `columns_w_values_for_single_analyte`, `columns_w_sample_ID_info`, `distance_method`
distance to similarity	`mutate(similarity = 1 / (1 + distance) * 100)`	
threshold	`filter(similarity > 20)`	the cutoff (compare several)
build the network	`buildNetwork()`	`edgelist` (node names in columns 1 and 2, optional weight in 3), `node_attributes`, `directed`
draw it	`geom_segment()` on `net$edges`, `geom_point()` on `net$nodes`	`linewidth` and `alpha` mapped to the weight
degree	`group_by(airport = origin) %>% summarize(out_degree = n())`	origin for out-degree, destination for in-degree
strength	`summarize(passengers_out = sum(passengers))`	
node removal	`filter(origin != "KTN", destination != "KTN")`, then rebuild	the node to remove