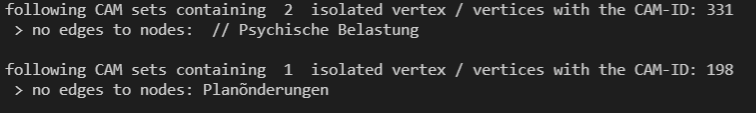
Folgende Probleme habe ich in den Valence Daten gefunden:

Das Grundproblem ist, das durch eine automatisierte Behebung der Probleme nicht mehr die Wortlisten / aggregierte CAMs usw. nicht mehr den Rohdaten entsprechen.

**Geisterknoten**

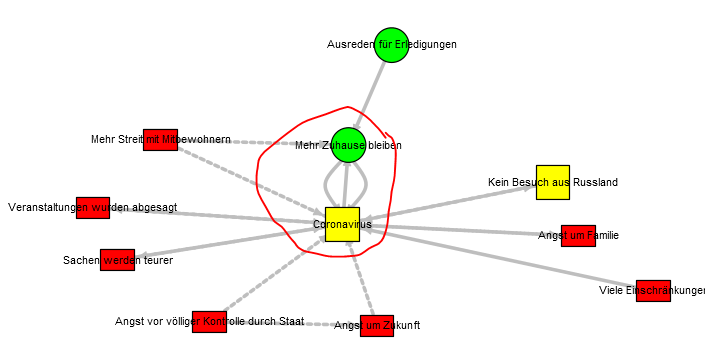
Dies sind Knoten (Vertex), welche mit keinerlei anderem Knoten verbunden sind in der CAM (also auch kein kleinstmögliches Netzwerk bestehend aus zwei Knoten bilden). In dem Beispieldatensatz anbei betrifft dies 3 Knoten in 2 CAM Datensätzen:



* Diese kann ich einfach automatisiert löschen?

**vielfache Verbindungen**

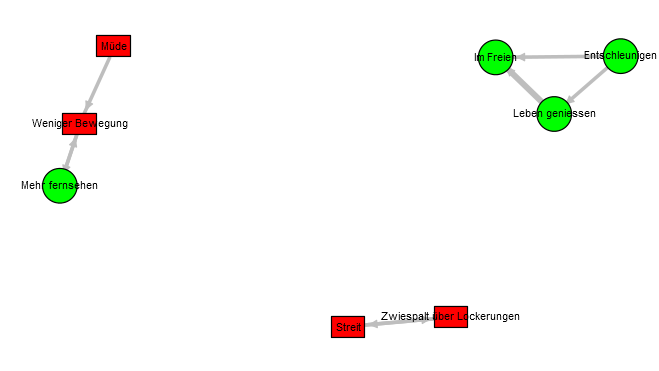
In den Valence Daten werden fälschlicherweise vielfache Verbindungen zwischen Knoten gezeichnet. Dies führt dazu das nicht mehr ein Einfaches, sondern in den Daten ein komplex Netzwerk (*multigraph*) vorliegt. Dies sieht graphisch wie folgt aus:



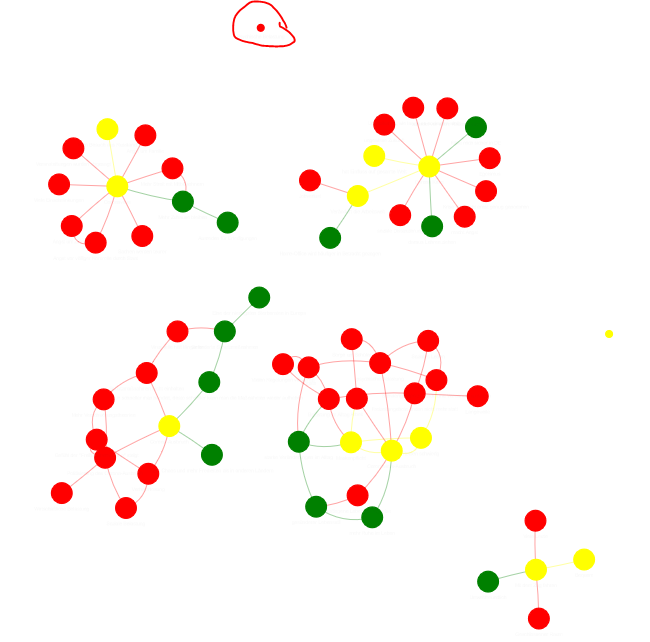
* Alle mehrfachen Verbindungen bis auf die letzte (Verwendung timestamp) löschen?

**unverbundene Netzwerke**

Valence zwingt die TeilnehmerInnen nicht alle Knoten zu einem großen Netzwerk zu verbinden und dies führt dazu, dass in einzelnen CAMs mehrere Komponenten (Netzwerkteile) vorliegen:



Wenn wir solche Daten aggregieren würden entsteht folgendes (hierbei rot umkreist ein Geisterknoten):



* Alle Komponenten bis auf die größte löschen und wenn es wie im Bild oben zwei gleichgroße Komponenten gibt Auswahl einer zufälligen Netzwerkkomponente?
* Wenig Knoten
* Evtl. leere Knoten