

1 Einleitung

Soziale Netzwerke wie Facebook, Instagram und Twitter sind in den letzten Jahren sehr stark gewachsen und haben somit einen großen Platz in unserer Gesellschaft eingenommen. Jede dieser Plattformen hat täglich mehrere Hunderte Millionen an aktiven Nutzern. Diese Nutzer kreieren oder konsumieren tagtäglich "Content" auf diesen Plattformen. Auch wenn man selber keinen "Content" erzeugt, interagiert man immer mit den anderen Nutzern in Form von Likes, Kommentaren oder impressionen also das Erreichen/Sehen des Posts.

Dadurch sind die Sozialen Netzwerke auch ein großes Gebiet in der Forschung geworden, da man sonst nirgendwo auf so viel Daten von Usern stößt. Es ist immer interessanter geworden, zu verstehen wie die Kommunikation auf solchen Plattformen in der Öffentlichkeit vonstattengeht. Denn man kann somit untersuchen und verstehen wie Menschen online mit anderen Menschen kommunizieren. Des Weiteren ist es möglich Konversationen und Diskussionen über bestimmte Themen zu analysieren. Es ist möglich das Verhalten der Menschen auf bestimmte Themen/Informationen zu verstehen und nachzuvollziehen wie diese Menschen damit umgehen. Und es ist möglich nachzuvollziehen wie sich Themen/Informationen in Sozialen Netzwerken ausbreiten. Wodurch sich dann ganze Netzwerke an User und Antworten auf die Themen/Informationen bilden.

Dadurch das Twitter eine sehr beliebte Plattform für den Themen-Austausch ist, beschäftigt sich diese Arbeit mit der "Microblogging"-Plattform Twitter. Denn auf Twitter finden Konversation nicht nur zwischen einzelnen Usern statt, sondern werden durch Öffentlich sichtbaren "Trends" von Millionen an Usern geführt. Deswegen ist es hier besonders spannend mit diesen Konversationen zu arbeiten, da hier besonders viele User aufeinander prallen und an den Konversationen teilnehmen können. Desweiteren bietet Twitter eine API Schnittstelle an, welches das Arbeiten mit den Konversation überhaupt erst möglich macht.

Mit der Twitter API wird in dieser Arbeit ein Verfahren entwickelt, ganze Konversationen von Twitter abzugreifen. Dazu müssen möglichst große Konversationen gefunden werden, diese müssen extrahiert werden und in einer analysierbaren Form gespeichert werden damit diese Konversationen in Zukunft weiter verwertbar sind. Dies gilt es zu schaffen, ohne die Twitter API an ihr Limit zu bringen, denn diese API ist für Drittsysteme nur begrenzt verfügbar. Diese begrenzte Verfügbarkeit ist dadurch zu Stande gekommen, das Soziale Medien in den letzten Jahren immer mehr unter Druck der Öffentlichkeit stehen um Nutzer Daten vor Drittsystemen zu schützen. Desweiteren sind Nutzer Daten für diese Plattformen das höchste Gut, da sie damit auch ihr Geld verdienen. Deswegen stellen Sie diese Dienste nur begrenzt zur Verfügung, was ein bestimmtes Abfrage Limit in einem bestimmten Zeitraum bedeutet.

1.1 Problembeschreibung

Um von Twitter ganze Konversationen zu extrahieren, und diese strukturiert darzustellen, gilt es einige Probleme auszumerzen. Das erste Problem was es zu lösen gilt, ist es Konversationen zu finden an denen möglichst viele Interaktionen stattfinden. Danach gilt es einen Algorithmus zu entwickeln und zu implementieren, welcher diese Konversationen abgreift. Dabei beinhaltet eine Konversation den "Haupt-" Tweet, welcher der Ankerpunkt der ganzen Konversation ist. Auf diesen "Haupt-" Tweet können unbegrenzt viele Antworten von unbegrenzt vielen Usern stattfinden. Dieser "Haupt-" Tweet und die Antworten bilden dann die gesamte Konversation.

Die Konversationen, also der "Haupt-" Tweet und die "Antwort-" Tweets werden durch die Twitter-API abgegriffen. Die nächste Hürde die es zu überwinden gilt, ist die Limitierung der Twitter-API. Denn diese Limitierung beinhaltet zum Beispiel die Regel, das innerhalb von 15 Minuten nur 300 Tweets abgefragt werden können. Dieses Limit ist bei großen Konversationen natürlich schnell erreicht. Deswegen gilt es in dieser Arbeit trotz der Einschränkungen ein verfahren zu finden, welches das abgreifen von den großen Konversationen möglich macht. Dabei gilt es außerdem zu berücksichtigen, das Twitter über die Twitter-API nicht direkt die gesamte Konversation ausliefert, sondern lediglich eine Id ausgibt, zu welcher Konversation ein Tweet gehört.

Dadurch das man nicht direkt an die gesamte Konversation durch die Twitter-API kommt, sondern nur an die Tweets mit einer Id, zu welcher Konversation diese Tweets gehören, gilt es ein Weg zu finden, um dennoch die gesamte Konversation, so wie diese auf Twitter stattgefunden hat, auch wieder darzustellen.

Des Weiteren sind nach dem einmaligen Abgreifen der Daten, die Konversationen nicht beendet. Also muss in dem Verfahren berücksichtigt werden, das Konversation im laufe der Zeit wachsen können. Somit müssen die Konversationen nicht nur einmalig, initial abgegriffen und wieder strukturiert dargestellt werden, sondern Sie müssen auch erweiterbar sein und erweiterbar bleiben. Deswegen bietet es sich an die Konversationen nach dem Sie abstrahiert wurden sind, als Graphen darzustellen. Denn somit ist die Konversation geordnet und immer erweiterbar sowohl in der Breite, als auch in der Tiefe.

1.2 Aufbau der Arbeit

In Kapitel 2 werden alle technischen und fachlichen sowie Konzepte und Definitionen vorgestellt.

In Kapitel 3 wird der Aufbau und die Durchführung des Algorithmusses nähergebracht, welche die Lösung auf die Problembeschreibung widerspiegelt.

In Kapitel 4 wird der Datensatz welcher sich aus der Visualisierung der gesammelten Daten, welche sich aus dem Algorithmus aus Kapitel 3 ergeben analysiert.

In Kapitel 5 wird die Analyse aus Kapitel 4 mit anderen Verfahren verglichen und bewertet.

In Kapitel 6 gibt es eine Zusammenfassung und ein Fazit zu den vorherigen Kapiteln. Desweiteren wird es einen Ausblick geben wozu die Datensätze verwendet werden können und wo es gegebenenfalls noch Verbesserungspotential gibt und welche Probleme für die Zukunft noch offen bleiben.

2 Grundlagen

In diesem Kapitel, werden die wichtigen Komponenten und Definitionen die zum Verstehen des Algorithmus und der Analyse des Datensatzes gebraucht werden. Außerdem wird die technische Implementierung erläutert um das Zusammenspiel der Komponenten zu verstehen.

2.1 Twitter

2.2 Graphen

2.3 Technische Implementierung