

Методы и технологии анализа данных социальных сетей

Власенко Д.М., федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики»

Научный руководитель – Иванов С.Е., доцент университета ИТМО

Введение

Анализ социальных сетей — сравнительно молодое явление, но резко возросший интерес к этой области породил множество методов и технологий. Популярность социальных сетей породила огромные массивы пользовательских данных и, тем самым, подготовив почву для исследований в самых разных областях. Несмотря на важность этой темы, сегодня имеется очень мало исследований, которые могли бы в достаточной мере осветить существующие подходы. Анализ социальной сети можно проводить различными способами, самые распространенные из них — путем визуализации, анализа социальных графов, анализа тональности текста или анализа медиафайлов.

Цель работы

Изучение существующих методов применяемых для анализа социальных сетей. Практическое сравнение методов визуализации, анализа социальных графов и анализа тональности текста. Обзор существующих решений. Изучение возможностей сбора данных для проведения исследования.

Описание предполагаемого подхода

Первым шагом исследования является сбор данных. В исследовании будут использоваться несколько произвольных выборок данных заданного размера, чтобы минимизировать шанс на расхождение полученных результатов с результатами на больших данных. Перед этим следует изучить возможные методы сбора публичных данных социальных сетей и сравнить их с уже имеющимися готовыми решениями. Реальные данные понадобятся для дальнейшего сравнения методов анализа. После сбора следует обзор существующих инструментов и методов визуализации, а также обзор сильных и слабых сторон методов после практического анализа на их основе.

Таким же способом будут рассмотрены с практической точки зрения различные решения и методы для анализа социальных графов и анализа тональности текста на основе выборки публичных данных. Таким образом, будут выявлены лучшие методы для каждой подзадачи анализа и сопоставлены с уже готовыми решениями.