



Типы сетевого поведения пользователей социальной сети “ВКонтакте” на примере городов Вологодской области



Тип сетевого поведения – группа пользователей социальной сети, которая имеет склонность подписываться на ограниченный набор сообществ.

Задачи:

- выделение типов сетевого поведения;
- поиск связей между структурой типов сетевого поведения города и его социально-экономическими характеристиками.

Сбор данных (API «ВКонтакте»):

1. Поиск пользователей 15-70 лет по городам Вологодской области.
2. Фильтр фейков, 3 критерия:
 - был в сети более 2 недель назад;
 - слишком много подписчиков;
 - реклама в статусе.
3. Выгрузка списков сообществ пользователей с открытыми страницами.

Подготовка данных, удаление:

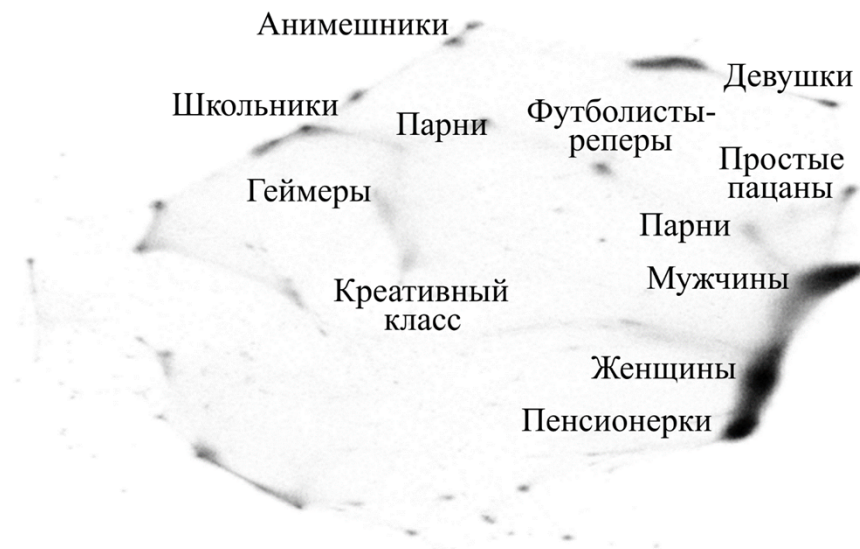
1. сообществ локальной повестки;
2. сообществ с одним подписчиком;
3. пользователей без подписок или с одной подпиской.

Исходные данные: граф в табличной форме:

вершина графа (386 тыс.) – пользователь (250 тыс.) или сообщество (136 тыс.), ребро графа – подписка (5 млн.).

Методика кластеризации графа:

- укладка графа – алгоритм OpenOrd (drl);
- случайная выборка 20 тыс. вершин графа;
- фильтрация шумовых вершин выборки – алгоритм dbscan;
- ядра кластеров выборки – иерархическая кластеризация;
- классификация неядерной части графа – алгоритм knn.



Два типа кластеров (типов сетевого поведения):

Половозрастные (нет основной темы сообществ, много юмора):

Пенсионерки, Женщины, Мужчины, Парни, Девушки, Школьники

Специализированные (есть одна-две основные темы): *Геймеры, Футболисты-реперы, Простые пацаны, Анимешники, Креативный класс*

Выводы:

Половозрастные кластеры – масса (95 % пользователей). Малые города – высока доля *Пенсионерок* и *Школьников*. Крупные города – высока доля *Парней*, *Девушек* и *Женщин*.

Специализированные кластеры – не элита, но девиации (5 %). Доля всех спец. кластеров, кроме *Простых пацанов*, повышена в крупнейших городах региона Вологде и Череповце.

Доля Креативного класса в Вологде на 14 % выше, чем в Череповце – столичность, ИТ, точное машиностроение против металлургии и минеральных удобрений.

Разнообразие типов сетевого поведения города зависит от его плотности и доли сектора услуг в структуре его экономики. При сопоставимой плотности разнообразие типов сетевого поведения, измеренное через энтропию Шеннона, выше там, где роль третичного сектора больше.

Все расчёты выполнены в R.