



ПОИСК НАИБОЛЕЕ ВЛИЯТЕЛЬНЫХ ОБЪЕКТОВ ВИРТУАЛЬНОЙ СОЦИАЛЬНОЙ СЕТИ НА ОСНОВЕ НЕЧЕТКОГО ГРАФА

Н.М. НОВИКОВА
А.В. БОРИСКИН

*Воронежский государственный
университет*

*e-mail:
nov.nelly@gmail.com
boriskinpost@gmail.com*

Представлены результаты моделирования автоматизированной системы управления кадровым потенциалом региона в строительной сфере, отвечающей следующим основным требованиям: обеспечение процесса эффективного и результативного управления кадровым потенциалом в выбранной сфере экономики, универсальность для любого уровня иерархии административно-территориального деления, адаптивность к текущим изменениям в социально-экономической сфере на рассматриваемой территории и во внешней среде.

Ключевые слова: поддержка принятия решений; кадровый потенциал строительного кластера региона; автоматизированная система управления; моделирование.

Введение

В современном мире отношения между людьми, кроме социального уровня, заняли еще один — цифровой. С распространением виртуальных социальных сетей появилась тенденция иметь собственную страницу с персональными данными, искать друзей по интересам, создавать группы. Количество информации в соцсетях постоянно увеличивается, и большая часть этой информации находится в необработанном виде. Такая информация не представляет интереса. Однако, обработав эти данные, можно получить информацию, которая может быть использована для практических целей. Например, данные о наиболее влиятельных объектах были бы полезны как для проведения различных виртуальных маркетинговых акций, так и для выявления пользователей с подозрительно высокой активностью.

Целью данной статьи является рассмотрение подхода на основе нечеткого графа для анализа влияния объектов внутри социальной сети.

Теоретический анализ

Социальную сеть принято представлять в виде графа, узлами которого являются люди [1]. Если объекты как-то связаны между собой (состоят в друзьях, либо переписываются), то связь между этими объектами обозначается в виде дуги. Полученный в этом случае граф обладает существенным недостатком: неважно насколько близко знакомы люди друг с другом, важно лишь, существует ли между ними хоть какая-то связь. В конечно счете он приравнивает отношения между случайными знакомыми к отношениям между людьми, которые тесно общаются уже не один десяток лет. Необходимо иметь модель, которая позволяла бы дифференцировать подобные случаи.

Для решения данной задачи может быть использовано понятие нечеткого графа [2]. В нашем случае под графом, представляющим модель данных, понимается пара (X, A) , где X — четкое множество вершин, представляющих людей в сети, A — нечеткое множество отношений между людьми. Более строго, A представляет собой множество пар (x_i, x_j) , для которых задана функция принадлежности $\mu_A(x, y)$, указывающая, в какой степени объект x напрямую влияет на объект y .

В дальнейшем, под *влиятельностью* будем понимать степень влияния объектов друг на друга. Следует заметить, что это понятие не сводится только к количеству связей с другими объектами. Должны учитываться степени влияния объектов, с которыми связан целевой объект (объект, для которого считается степень влияния). Таким образом, нужно так формализовать понятие влиятельности, чтобы учитывались как количество связей, так и влиятельность объектов, с которыми связан целевой объект.



Следует заметить, что существует понятие *центральности* [3] вершины графа, которое выражает понятие влиятельности и имеет различные методы измерения. Поскольку большинство алгоритмов расчета этой величины имеют квадратичную сложность, это препятствует их применению для больших объемов данных.

В статье [4] представлен один из алгоритмов решения данной задачи. Данный алгоритм обладает вышеуказанными недостатками: квадратичной сложностью и бинарностью в отношениях. Ниже представлена модель, которая позволяет обойти эти ограничения. Пусть $\mu_A(x, y)$ – функция принадлежности, указывающая в какой степени объект x напрямую влияет на объект y . Принимает значения из отрезка $[0,1]$.

В [4] вводится понятие *итерированной силы объекта i порядка k* (обозначается $p^i(k)$) через матрицу смежности графа [5]. Вместо матрицы смежности будем использовать функцию принадлежности. Получим

$$\begin{cases} p^i(k) = \sum_{j=1}^n \mu_A(x_i, x_j) \times p^j(k-1), & i = 1..n, k \geq 1 \\ p^i(0) = 1 & i = 1..n \end{cases} \quad (1)$$

Заметим, что итерированная сила первого порядка объекта i есть сумма значений функции принадлежности данного объекта с другими. Она еще не учитывает влиятельность других объектов. Начиная со второго порядка, в данную сумму включается влиятельность иных объектов.

Возникает вопрос: до какого порядка считать вектор итерированных сил? Покажем, что из практических соображений достаточно считать до второго либо до третьего порядка. Предполагается, что итерированная сила объекта i порядка k выражает степень влияния объекта i , учитывая, что он свое влияние может распространить не более чем в радиусе k . Это вытекает из определения итерированной силы. Например, если вы просите своего друга попросить кого-то сделать что-то для вас – то это влиятельность в радиусе 2, и выражается она итерированной силой второго порядка. Если учитываются цепи влияния с еще одним промежуточным участником, то нужно считать вектор итерированных сил 3-го порядка. Считать итерированные силы больших порядков кажется нецелесообразным ввиду ничтожно малой вероятности возникновения столь длинных цепочек в реальной жизни.

Таким образом, если анализируется влиятельность объектов в масштабе крупного города и выше, то нужно считать до 3-го порядка. Для меньших масштабов целесообразно использовать второй порядок.

Следует заметить, что важно не численное значение итерированной силы, а то, как силы для разных объектов соотносятся между собой. Поэтому после расчета вектора итерированной силы очередного порядка целесообразно нормировать данный вектор.

Из формулы (1) может показаться, что алгоритм имеет квадратичную сложность, однако, следует принять во внимание 2 факта:

- в реализации алгоритма функция принадлежности задается не в виде матрицы, а в виде списка. Это имеет те же самые преимущества, что и списки смежности в детерминированном графе [5];
- количество связей одного объекта с другими в социальных сетях обычно ограничивается некоторой константой.

Принимая во внимание вышеуказанные факты, получаем, что расчет вектора итерированных сил имеет линейную сложность. После расчета следует сортировка элементов данного вектора, поэтому предложенный алгоритм имеет сложность $O(n \cdot \ln(n))$. Следовательно, данный алгоритм можно применять к большим объемам данных.



Методика эксперимента

Для проведения эксперимента был написан программный комплекс, состоящий из двух приложений:

- приложения для сбора данных;
- приложения для анализа данных.

Первое приложение собирает информацию из социальной сети «ВКонтакте». Сначала выбирается объект, с которого будет начинаться сбор данных. Назовем его *центром выборки*. Для данного объекта приложение загружает 2 набора друзей: до 2-го и до 3-го уровней. Под уровнем n подразумевается множество объектов, удаленных от центра выборки посредством n связей. Помимо самих друзей для каждого объекта загружались следующие данные:

- список фотографий, на которых пользователь поставил отметку «мне нравится»;
- список видеозаписей, на которых пользователь поставил отметку «мне нравится»;
- список записей, на которых пользователь поставил отметку «мне нравится».

Было исследовано [6], что если загружать 4 уровня, то будет собрана информация о более чем 90% пользователей сети. Существует теория шести рукопожатий, которая утверждает, что между любыми двумя людьми на Земле не более 5 уровней общих знакомых. В силу того, что VK (VKontakte) распространена в основном в СНГ, то эта цифра должна быть меньшей.

Теперь остался открытым вопрос о вычислении функции принадлежности $\mu_A(x, y)$. Ее можно определить по-разному в зависимости от количества предоставляемой социальной сетью информации. Ключевым моментом при определении должно стать следующее правило: чем выше значение функции, тем большее прямое влияние оказывает объект x на объект y . В данном приложении функция принадлежности определяется следующим образом:

$$\mu_A(x, y) = 0.25 \cdot F(x, y) + 0.25 \cdot P(x, y) + 0.25 \cdot V(x, y) + 0.25 \cdot M(x, y),$$

где

$$F(x, y) = \begin{cases} 1, & \text{объекты } x \text{ и } y \text{ являются друзьями} \\ 0, & \text{иначе} \end{cases}$$

$$P(x, y) = \begin{cases} 1, & \text{объект } y \text{ пометил более 10 фотографий} \\ & \text{объекта } x \text{ как "мне нравится"} \\ 0.1 \cdot k, & \text{объект } y \text{ пометил } k \text{ (} 1 \leq k \leq 10 \text{) фотографий} \\ & \text{объекта } x \text{ как "мне нравится"} \\ 0, & \text{иначе} \end{cases}$$

$$V(x, y) = \begin{cases} 1, & \text{объект } y \text{ пометил более 10 видеозаписей} \\ & \text{объекта } x \text{ как "мне нравится"} \\ 0.1 \cdot k, & \text{объект } y \text{ пометил } k \text{ (} 1 \leq k \leq 10 \text{) видеозаписей} \\ & \text{объекта } x \text{ как "мне нравится"} \\ 0, & \text{иначе} \end{cases}$$



$$M(x, y) = \begin{cases} 1, & \text{объект } y \text{ пометил более 10 записей} \\ & \text{объекта } x \text{ как "мне нравится"} \\ 0.1 \cdot k, & \text{объект } y \text{ пометил } k \text{ (} 1 \leq k \leq 10 \text{) записей} \\ & \text{объекта } x \text{ как "мне нравится"} \\ 0, & \text{иначе} \end{cases}$$

После того, как этап сбора информации был завершен, данные были проанализированы вторым приложением.

Обсуждение результатов

Итак, было получено 2 набора данных:

- для 2 уровней друзей – порядка 40 тыс. пользователей;
- для 3 уровней друзей – порядка 4,2 млн. пользователей.

В обоих случаях центром выборки был студент Воронежского государственного университета (ВГУ).

На рис. 1 представлено приложение по анализу данных. Проанализируем первый набор.

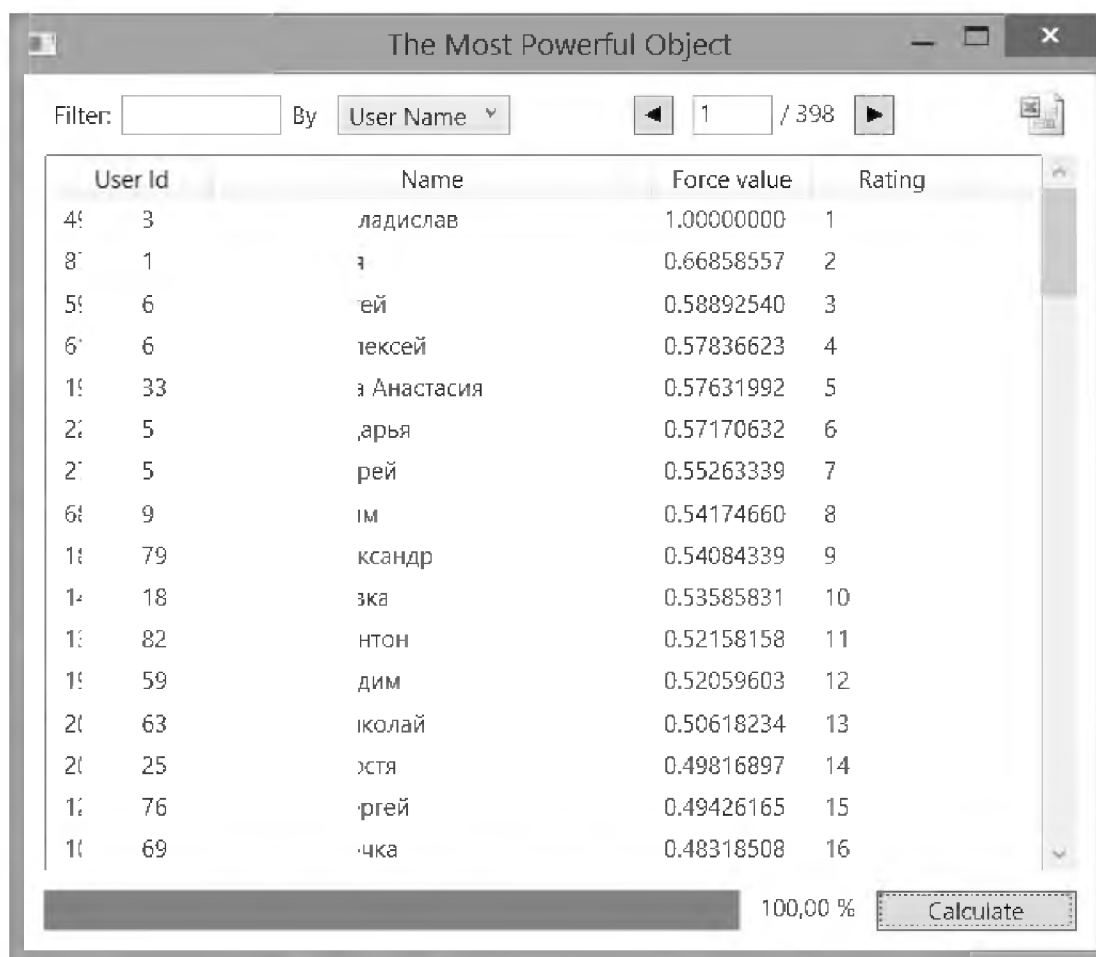


Рис. 1. Анализ первого набора данных

Здесь центр выборки занял первое место по влиятельности. Объяснить это можно тем, как загружались данные. Дело в том, что весь последний уровень друзей содержит лишь по одной связи. Для этих объектов список друзей пуст. Поэтому влиятельность

друзей 1-го уровня (которая в значительной степени базируется на друзьях 2-го уровня) оценивается низко. Из-за этого полученные результаты не являются объективными. Тем не менее, можно наблюдать интересные закономерности: в первых 30 строках располагаются около половины активистов ВГУ.

Анализ второго набора данных представлен на рис. 2. Здесь центр выборки по рейтингу занимает 3248 место. Оно и понятно: тут влияние рассматривается уже в рамках города.

Rank	Value	User Name	Score	Order
30		mka	0.94246038	3
29	0	а	0.93167087	4
1	39	эннадий	0.91494713	5
8	2	эдим	0.91044472	6
1	91	й	0.90298056	7
1	43	зел	0.89092538	8
4		ев Константин	0.89065200	9
1	99	ня	0.88892969	10
150808972		Ручьев Сто	0.86363781	11
49		згения	0.85986040	12
10	1	эман	0.85727470	13
31	5	трина	0.85696980	14
14	3	димир	0.85518377	15
13	17	islav	0.85330297	16
88	7	иниловая	0.85107801	17
18	26	а	0.84823910	18
120706211		Vrn Арт-реальность	0.84632015	19

Рис. 2. Анализ второго набора данных

В начале данной таблицы прослеживаются следующие объекты:

- Кинотеатр Спартак
- Арт-реальность Vrn
- Воронеж nighparty.ru
- Вrn экскурсии
- Ручьев Сто
- Воронеж Афиша

Что касается людей в первых строчках данного списка, можно сказать, что часть из них действительно известные люди Воронежа. Но активность некоторых объектов вызывает подозрение: слишком большая степень влияния при достаточно скупой информации о личности. Данные объекты могут быть рассмотрены администраторами сети на соответствие информации действительности.

Выводы

В работе предложен алгоритм поиска наиболее влиятельных объектов социальной сети, который может быть применен к большим объемам данных. Продемонстрировано



применение данного алгоритма к известной социальной сети «ВКонтакте». Анализ результатов показывает, что расчет влияемостей объектов в крупных социальных сетях может предоставить информацию, которая была бы полезной как в виртуальном маркетинге, так и для повышения безопасности внутри социальной сети.

Список литературы

1. Давыдов А.А. Системная социология / А.А. Давыдов. – М.: ИС РАН, 2009. – 192 с.
2. Ибрагимов В.А. Элементы нечеткой математики / В.А. Ибрагимов. – М.: Баку, АГНА, 2010. – 394 с.
3. Freeman L.C. Centrality in social networks: Conceptual clarification / L.C. Freeman // *Social Networks*. –1978. – № 1. – С. 215-239.
4. Web и Social Mining: [сайт]. – (URL: http://www.basegroup.ru/library/web_mining/) (дата обращения 20.05.2014)
5. Берж К. Теория графов и ее применения / К. Берж. – М.: Иностранная литература, 1962. – 318 с.
6. Facebook cuts six degrees of separation to four – Telegraph: [сайт]. – (URL: <http://www.telegraph.co.uk/technology/facebook/8906693/Facebook-cuts-six-degrees-of-separation-to-four.html>) (дата обращения 03.05.2014)

SEARCH THE MOST INFLUENTIAL OBJECTS OF VIRTUAL SOCIAL NETWORK BASED ON FUZZY GRAPH

N. NOVIKOVA
A. BOBISKIN

Voronezh State University

e-mail:
nov.nelly@gmail.com
boriskinpost@gmail.com

In this article one of possible methods for searching the most powerful objects in a virtual social network is considered. The algorithm to solve the problem for huge social networks is suggested. It's demonstrated how to apply the algorithm to the famous social network «VK».

Keywords: virtual social networks, data mining, fuzzy graphs.