

## РАНЖИРОВАНИЕ ПОНЯТИЙ, ИЗВЛЕКАЕМЫХ ИЗ ПОТОКОВ СЕТЕВЫХ НОВОСТЕЙ

*А.А. Снарский<sup>1,2</sup>, Д.В. Ланде<sup>2,1</sup>, Д.И. Зоринец<sup>2</sup>*

*<sup>1</sup>Национальный технический университет Украины им. И.  
Сикорского,*

*<sup>2</sup>Институт проблем регистрации информации НАН Украины*

Ранжирование – один из методов упорядочения объектов, как физических, так и информационных. В том случае, когда каждому объекту из совокупности можно приписать некоторое численное значение, задача ранжирования становится формально тривиальной, так как объекты можно ранжировать по величине этого значения. Сложность, однако, заключается в том, что, во-первых, не всегда понятно как определить такое численное значение, а, во-вторых, таких численных значений может быть много и не всегда ясен критерий, по которому нужно выбирать одно из них. Другими словами, наиболее сложной, плохо формализуемой частью задачи ранжирования является выбор критерия, по которому объекту приписывается численные значения (формализация объектов).

В работе, на примере группы ведущих мировых политиков (далее будем называть их персонажи), предложен новый метод ранжирования, позволяющий оценить наиболее «влиятельных» и наиболее «подверженных влиянию» политиков в данный период времени.

Метод состоит из двух этапов. На первом этапе для каждого персонажа определяется число его цитирований в некотором пуле печатных изданий на каждый день, т.е. формируется временной ряд. Для этого нами была использована система InfoStream. Далее каждый член этих временных рядов (соответствует дню) нормируется на полное число цитирований всех персонажей за этот день. Для полученных нормированных числовых рядов вычисляются взаимные временные корреляторы, задающие среднее значение по множеству полученных значений для совместного распределения двух процессов (персонажей). В зависимости от сдвига временного интервала при вычислении коррелятора, полученные корреляторы задают вероятность появления цитирования одного персонажа от того, был ли процитирован ранее другой. Полученные для всех персонажей корреляторы можно представить как набор сложных сетей. Каждая сложная сеть представляет собой набор узлов (персонажей), соединенных направленными связями, вес которых соответствует численному значению коррелятора. Каждой сети

соответствует свой временной сдвиг (запаздывание) при вычислении корреляторов – день, два и т.п. Таким образом получается полный направленный граф. Связь, направленная от одного персонажа к другому определяет прямое влияние (положительное или отрицательное, в зависимости от знака численного значения коррелятора).

На втором этапе для определения полного влияния одного персонажа на другого, учитывается не только непосредственное влияние (связь между персонажами), но и опосредованное, через цепочку. Для этого полученная сложная сеть представляется как когнитивная карта. Эта когнитивная карта может быть исследована различными методами, например, импульсным или фаззи-методом. Нами был использован К-метод [1].

В результате такого анализа вычисляются попарные влияния узлов, что позволяет рассчитать суммарное влияние каждого узла на все остальные. Именно это численное значение и было определено как критерий ранжирования.

Адекватность предложенного подхода была экспериментально проверена на группе ведущих мировых политиков, в частности, получен нормированный ряд цитирований для двух персонажей – Трампа и Клинтон для периода, соответствующего октябрю 2016 г., при этом ранговый коэффициент первого оказался выше. Также предложенным методом был получен ранжированный ряд для всех персонажей.

### *Литература*

1. Snarskii A.A., Zorinets D.I., Lande D.V., Levchenko A.V. K-method of cognitive mapping analysis // E-Preprint ArXiv:1605.08243, 2016.