Применение активного обучения к графовым моделям на примере оценки рисков распространения эпидемии

Антон Юрьевич Бишук

Московский физико-технический институт

Отчет по НИР/Группа 774, осень 2020
Научный руководитель: Зухба Анастасия Викторовна

Системы, описываемые графовыми моделями

- Предполагается, что каждая вершина графа может находиться в конечном количестве состояний.
- Состояние системы полностью определяется состоянием всех вершин.
- Текущее состояние известно только для части вершин.
- Возможно запрашивать состояние отдельных вершин.
- Задача заключается в том, чтобы наиболее точно оценить состояние всей системы.
- Состояние системы меняется во времени в соответствии с определенными правилами.

Цель исследования

Цель работы

Построить алгоритм, определяющий вершины, выяснение состояния которых позволят как можно точнее определить состояние всей системы с учетом изменения во времени.

Задачи

- Предложить возможные методы измерения количества информации о системе: в том числе примитивный (по количеству параметров) и энтропийный.
- Рассмотреть возможность использования предложенных методов для оценки полезности вершины с точки зрения информации о системе в целом.
- Изучить связь полученных методов с центральностями
- Выбрать подход к решению задачи

Гипотезы

Конкретизируем задачу: будем рассматривать графовые модели эпидемиологических процессов.

1. Поиск нужных людей можно переформулировать в задачу поиска вершин с наибольшей центральность. Поэтому необходимо определить какие центральности можно для этого использовать.
2. Поскольку мы имеем дело с не замкнутой системой, то наш алгоритм не должен быть детерминированным. Необходимо иногда делать случайные блуждания в графе.

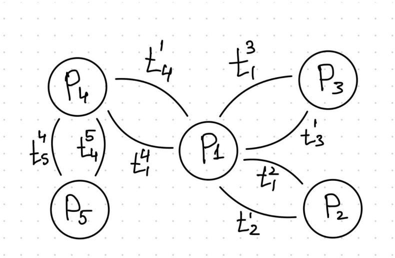
Модель: обозначения

1. i' – статус i-го человека (0 – здоров, 1 – болен);
2. p_i – вероятность того, что человек i – болен;
3. k_i – число контактов i-го человека;
4. t_i^j – время контакта i-го человека с j-ым.

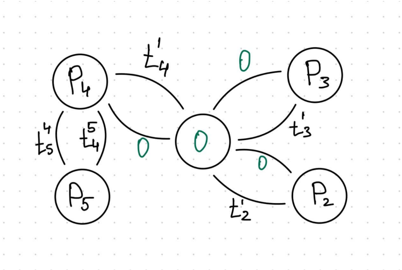
Тогда во всей системе общее число параметров:

$$n + 2\sum_{i=1}^{n} k_i + 2\sum_{i=1}^{n} [t_i]$$

Инициализируем каждого пользователя вероятностью того, что он был на текущий момент зараженным.
Такой подход позволяет указывать заведомо больных и здоровых людей просто ставя вероятность болезни 1 и 0 соответственно.



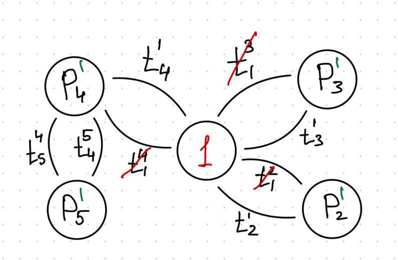
$$f(t_1^2) = f(t_1^3) = f(t_1^4) = 0$$



проверили человека и он оказался здоров, тогда граф на следующей итерации будет выглядеть следующим образом

$$p_1' = 1 \text{ далее всегда}$$

На этого человека теперь так же не влияют другие.



проверили человека и он оказался больным, тогда граф на следующей итерации будет выглядеть следующим образом

1. Если мы проверили i-го человека и он оказался здоров, то мы определили $k_i + [t_i]$ параметров системы;
2. Если мы проверили i-го человека и он оказался болен, то мы определили $1 + 2k_i + [t_i]$ параметров системы;

Тогда мы получаем, что:

$$\begin{cases} \text{с вероятностью } \mathrm{p}_i & \text{определим } 1{+}2\mathrm{k}_i + [t_i] \text{ параметров} \\ \text{с вероятностью } 1\text{-}\mathrm{p}_i & \text{определим } \mathrm{k}_i + [t_i] \text{ параметров} \end{cases}$$