

УДК 316.77:[004.82+004.89]

О.О. Савельев, А.И.Шевченко

Донецкий национальный технический университет,
кафедра программного обеспечения интеллектуальных систем

ОПРЕДЕЛЕНИЕ ЕСТЕСТВЕННОГО ПЕРИОДА АКТИВНОСТИ АБОНЕНТА ТЕЛЕФОННОЙ СЕТИ

Аннотация

Савельев О.О., Шевченко А.И. Определение естественного периода активности абонента телефонной сети. Обоснован выбор метрик для класса эгоцентрических социальных сетей. Адаптирован существующий метод на основе корреляционного и спектрального анализа для предложенных метрик. Показана возможность практического применения.

Ключевые слова: естественный период, трафик телефонной сети, социальная сеть.

Постановка проблемы. Областью интересов авторов является проблематика предметной области анализа трафиков телефонных сетей (ТТС). В частности, в работе [1] выполнена постановка задачи прогнозирования связей и предложен подход к ее решению, основанный на методах машинного обучения. Одним из ключевых моментов подхода является моделирование сети абонентов динамическим графом (ДГ), который представляет дискретную по времени структуру, генерируемую по алгоритму [2]. Одним из исходных параметров алгоритма есть величина Δt – период квантования транзакционных данных ТТС. Выбор данного параметра может в значительной степени влиять на качество работы прогностической модели, поэтому синтез метода для определения его естественного значения Δt_{nat} является актуальной задачей.

Исследователями из MITN. Eagle и A. Pentland предложено использовать спектральный анализ для определения периода естественного ритма физических контактов субъектов организационной структуры [3]. В дальнейшем A. Clauset и N. Eagle развили и применили свой подход к распределенной сети субъектов [4], было предложено строить ДГ всей сети, измерять метрики для каждого из состояний ДГ: среднюю степень вершины, локальную плотность треугольников, коэффициент смежной корреляции состояний ДГ. Полученные последовательности метрик для каждого из состояний анализируются методами корреляционного и спектрального анализа, на основе результатов которых принимается экспертное решение о величине Δt_{nat} . В недавней работе [5] сотрудниками Ericsson R&DA. Polepally

и S. Mohan применялся аналогичный подход, но уже к ТТС. Отличие от работ [3, 4] составляет выбор иных метрик, в т.ч. средней длительности звонка.

Целью данной работы является увеличение качества прогноза связей в ТТС благодаря использованию естественного Δt_{nat} . Для достижения поставленной цели необходимо решить задачи: выбрать и обосновать метрики, и адаптировать метод [4] для определения естественного Δt_{nat} для исследуемого класса социальных сетей.

Выбор метрик и адаптация метода. Алгоритм построения ДГ [2] позволяет получить модель, представляющую как замкнутую сеть нескольких абонентов, так и социальный граф одного абонента. В этом состоит ключевое отличие предлагаемого подхода от работ [3-5]. Так как, в реальном мире сеть контактов – свободно растущая, то ограничиваться данными из одного источника для замкнутой группы абонентов и строить модель для усредненного абонента не совсем целесообразно – в модели не будут учтены абоненты извне группы. Поэтому будем рассматривать частный случай ДГ как модели социального графа одного абонента, т.е. эгоцентрической социальной сети, где анализируемый абонент есть центральная вершина, а остальные абоненты представляются висячими либо изолированными вершинами. Для полученной модели возможно использование следующих метрик.

1. Степень центральной(максимальная степень) вершины k_{max} . Данная метрика соответствует количеству уникальных связей в состоянии ДГ. Определяется как максимальное количество ребер инцидентных одной вершине. В большинстве случаев это будет вершина анализируемого абонента, кроме случаев, когда граф содержит одно ребро, тем не менее, это не влияет на саму метрику.

2. Инверсия коэффициента смежной корреляции соседних состояний ДГ \bar{r} . Данная метрика показывает насколько не похожи два состояния ДГ между собой. Определяется как разница единицы и модуля коэффициента корреляции Пирсона для последовательностей весов соответствующих ребер, где хотя бы одно ребро существует и имеет ненулевой вес. Для вырожденных графов данная метрика принимает нулевое значение.

Метод заключается в построении ДГ для минимально допустимого значения Δt , подсчете метрик k_{max} и \bar{r} , их корреляционном и спектральном анализе. Эксперт принимает личное решение о значении Δt_{nat} на основе полученных результатов анализа.

Результаты исследования. Для экспериментальной проверки был использован тот же набор данных [6], что и в работах [3, 4]. В качестве исследуемого абонента был выбран абонент с идентификатором 95, так как он является одним из участников с наибольшим количеством событий. В качестве выборки были взяты события голосовых звонков и текстовых сообщений за октябрь 2004 года (744 часа), всего 735 события.

Сначала построим ДГ для $\Delta t = \{1, 3, 8, 24\}$ часов и измерим метрики k_{\max} и \bar{r} (рис. 1 и 2). Нижняя граница в один час выбрана для соблюдения баланса между размером выборки и количеством вырожденных графов. При $\Delta t = 1$ наблюдается высокочастотная “модуляция” гладких кривых, с увеличением Δt кривые сглаживаются, но в виду дискретности приобретают прямоугольный вид. На отдельных графиках можно хорошо различить паттерны для промежутков утро-день-вечер-ночь или даже будний-выходной день.

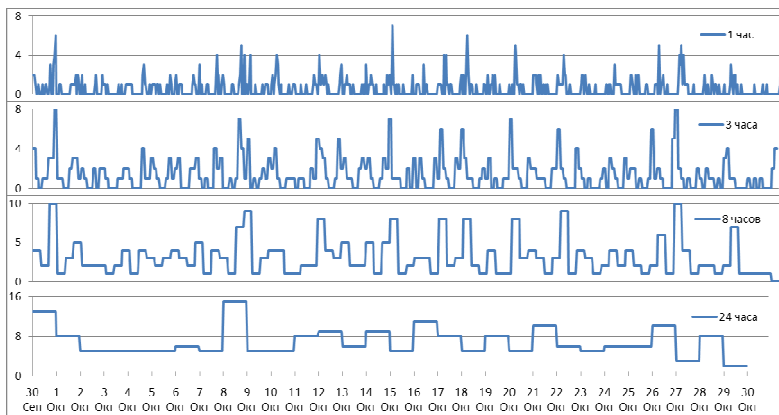


Рисунок 1 – Максимальная степень вершины

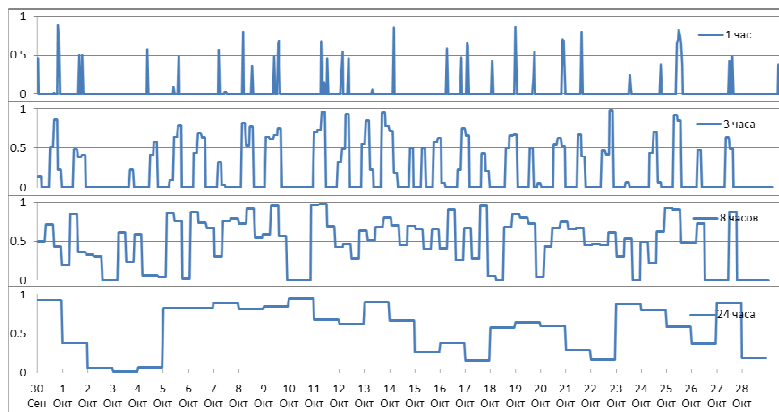


Рисунок 2 – Инверсия коэффициента смежной корреляции

На рис. 3 показаны усредненные значения метрик $\langle k_{\max} \rangle$ и $\langle \bar{r} \rangle$ как функций от Δt . Видно, что с увеличением Δt имеет место монотонная

тенденция к насыщению. Причем, $\langle k_{\max} \rangle$ достигнет насыщения при $\Delta t = 744$, а $\langle \bar{r} \rangle$ уже при $\Delta t = 12$. На рис. 4 показана автокорреляционная функция для обеих метрик k_{\max} и \bar{r} , вычисленных при $\Delta t = 1$. Анализ данного графика свидетельствует о периодической зависимости. Автокорреляция для k_{\max} и \bar{r} устанавливается в ноль при времени разделения ≈ 6 .

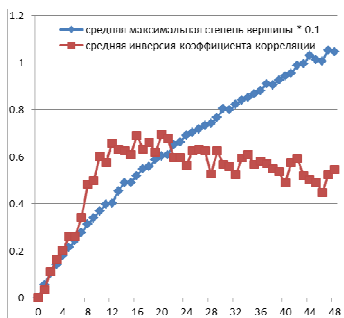


Рисунок 3 – Усредненные метрики $\langle k_{\max} \rangle$ и $\langle \bar{r} \rangle$ как функции

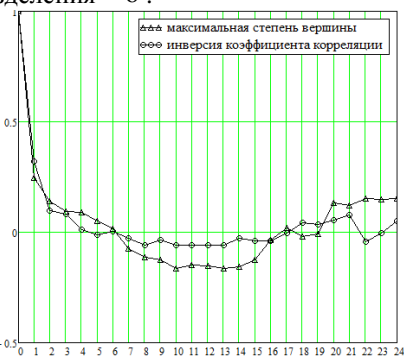


Рисунок 4 – Автокорреляционные функции

На рис. 5 показаны спектры мощности для метрик k_{\max} и \bar{r} . Можно увидеть сильные пики на обоих графиках для частот 1, 2, 3, 4 раз в день и менее сильные для 6, 7 раз в день. Соответственно в качестве Δt допустимо выбрать любое значение, удовлетворяющее этим частотам.

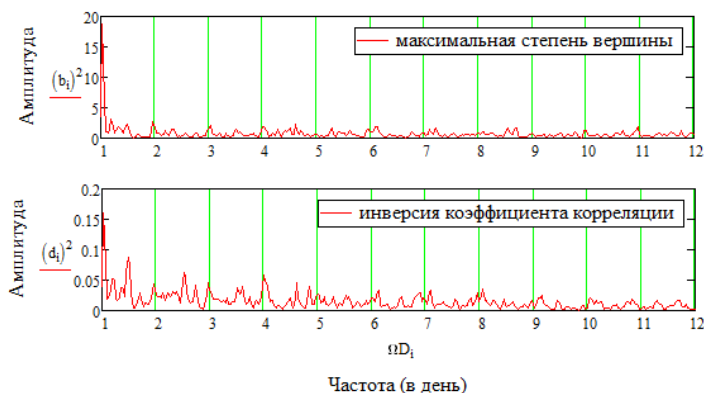


Рисунок 5 – Спектры мощности для метрик k_{\max} и \bar{r}

Принимая во внимание все результаты, сделаем вывод, что наиболее целесообразно выбрать $\Delta t_{nat} = 3$ часа, соответственно средние метрики будут иметь значения $\langle k_{max} \rangle = 1.4$ и $\langle \bar{r} \rangle = 0.164$.

Выводы. Предложены новые метрики максимальной степени вершины и инверсии коэффициента смежной корреляции соседних состояний ДГ для класса эгоцентрических социальных сетей, адаптирован существующий метод для определения естественного периода активности абонента по предложенным метрикам. Приведены результаты практического применения, показывающие высокую степень адекватности предложенного подхода.

Список литературы

1. Савельев О. О. Постановка задачи исследования прогнозирования связей в трафиках телефонных сетей [Текст] / О. О. Савельев, А. И. Шевченко // Восточно-Европейский журнал передовых технологий. – 2012. – № 6/3 (60). – С. 51-60.
2. Савельев О. О. Построение динамического социального графа по транзакционным данным трафиков телефонных сетей [Текст] / О. О. Савельев // Матеріали доповідей VI міжнародної науково-практичної конференції молодих учених, аспірантів, студентів «Сучасна інформаційна Україна: інформатика, економіка, філософія», 26 квітня 2012 р. – Донецьк : Наука і освіта. – 2012. – С. 79-83.
3. Eagle N. Reality mining: sensing complex social systems [Текст] / Nathan Eagle, Alex (Sandy) Pentland // Journal Personal and Ubiquitous Computing. – 2006. – Volume 10, Issue 4. – PP. 255-268.
4. Clauset A. Persistence and periodicity in a dynamic proximity network [Электронный ресурс] / Aaron Clauset, Nathan Eagle // DIMACS/DyDAN Workshop on Computational Methods for Dynamic Interaction Networks, September 24 - 25, 2007. – Rutgers, N.J. : DIMACS Center. – 2007. – 5 PP. – Режим доступа : <http://reality.media.mit.edu/pdfs/Clauset.pdf>.
5. Polepally A. Behavior Analysis of Telecom Data Using Social Network Analysis [Текст] / Avinash Polepally, Saravanan Mohan // Behavior Computing: Modeling, Analysis, Mining and Decision [ed. Longbing Cao, Philip S. Yu]. – London : Springer-Verlag, 2012. – PP. 291-303.
6. Eagle N. Reality mining [Электронный ресурс] / Nathan Eagle. – Cambridge : Massachusetts Institute of Technology Media lab. – 2009. – Режим доступа : <http://reality.media.mit.edu/>