

UNIVERSITY OF TARTU
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE
Institute of Computer Science
Software Engineering

Irene Teinemaa

**Community-Based Prediction of
Activity Decay
in a Social Network**

Master's thesis (30 ECTS)

Supervisor: Anna Leontjeva

Opponent: Luciano García-Bañuelos

TARTU 2014

Community-Based Prediction of Activity Decay in a Social Network

Abstract:

An important problem for facilitators of online social networks is to identify the users who are likely to decrease their level of activity in the near future. Such predictions are the basis for targeted campaigns aimed at sustaining or increasing the overall user engagement in the network. A common approach to this problem is to apply machine learning methods to make predictions at the level of individual users. The existing approaches, however, do not consider the social connections of the individuals to their full extent, leaving room for improvement.

In this context, we propose a new approach to the problem of activity decay prediction based on the idea of identifying groups of tightly inter-linked users (namely communities) where the level of social activity is likely to decay. We investigate two community detection methods and compare the resulting predictive accuracy against several baselines. We show that more individuals who are likely to decay can be reached by targeting communities instead of single users. Moreover, a bottom-up community detection method produces higher accuracy in this context than a top-down modularity-based approach. Additionally, a richer set of features related to user engagement can be used for prediction purposes, leading to more accurate predictions. The results pave the way for designing community-based approaches to analyze user engagement in social networks as well as associated community-based targeting methods.

Keywords: Social network analysis, community detection, machine learning, churn prediction

Kogukonnapõhine aktiivsuse languse ennustamine sotsiaalvõrgustikus

Lühikokkuvõte:

Virtuaalsete sotsiaalvõrgustike haldajate seisukohalt on oluline tuvastada kasutajaid, kes kaotavad suure tõenäosusega lähitulevikus huvi nende teenuse vastu. Selliste kasutajate ennustamine lubab suunata neile kampaaniaid, hoidmaks või suurendamaks aktiivsust võrgustikus. Nimetatud probleemi lahendatakse sageli masinõppemeetodite abil, tehes ennustusi üksikisiku tasandil. Olemasolevad lahendused ei kasuta aga maksimaalselt ära kasutajate omavahelisi suhteid.

Selles kontekstis tutvustame uut lähenemist, ennustamaks aktiivsuse langust kogukondade ehk omavahel tihedalt seotud kasutajate gruppide tasandil. Antud töös kasutame kahte meetodit kogukondade leidmiseks ning võrdleme tulemusi üksikkasutajate ja juhuslike kasutajate gruppidega. Analüüs näitab, et teenusest loobuda plaanivaid kasutajaid on lihtsam leida kogukondade kui üksikisiku tasandil. Tulemused näitavad, et ennustuste kvaliteet sõltub ka kasutatud kogukondade leidmise algoritmist. Meetod, mis leiab kogukonnad lokaalsel tasandil, lähtudes iga kasutaja otsesest suhtlusringkonnast, võimaldab paremaid ennustusi kui võrgustikule tervikuna orienteeritud meetod. Lisaks eelmainitule võimaldab kogukonnapõhine analüüs arvesse võtta täiendavaid tunnuseid, saamaks täpsemaid ennustusi. Saadud tulemused on aluseks uute kogukonnapõhiste meetodite väljatöötamisele, analüüsime kasutajate aktiivsust sotsiaalvõrgustikes ning tõstmaks turunduskampaaniate efektiivsust.

Võtmesõnad: Sotsiaalvõrgustiku analüüs, kogukondade leidmine, masinõpe, aktiivsuse ennustamine

Non-exclusive licence to reproduce thesis

I, Irene Teinemaa, (date of birth: 22.08.1990),

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for addition to the DSpace digital archives until expiry of the term of validity of the copyright

Community-Based Prediction of Activity Decay in a Social Network,
supervised by Anna Leontjeva,

2. I am aware of the fact that the author retains these rights.

3. This is to certify that granting the non-exclusive licence does not infringe the intellectual property rights or rights arising from the Personal Data Protection Act.

Tartu, 26.05.2014