**MODULE DESCRIPTION**

**5.1 PROJECT MODULES**

5.1.1 Identifying Leading Sessions

5.1.2 Ranking Based Evidences

5.1.3 Rating Based Evidences

5.1.4 Review Based Evidences

5.1.5 Evidence Aggregation

**5.1.1 IDENTIFYING LEADING SESSIONS**

* Ranking fraud usually happens in leading sessions Therefore, detecting ranking fraud of mobile Apps is actually to detect ranking fraud within leading sessions of mobile Products, a simple yet effective algorithm to identify the leading sessions of each Product based on its historical ranking records.
* Then, with the analysis of Products’ ranking ‘behaviors, we find that the fraudulent Products often have different ranking patterns in each leading session compared with normal Products. Mining Leading Sessions: There are two main steps for mining leading sessions.
  + First, we need to discover leading events from the Product’s historical, ranking records.
  + Second, we need to merge adjacent leading events for constructing leading sessions.

**5.1.2 RANKING BASED EVIDENCES**

* A leading session is composed of several leading events. Therefore, we should first analyze the basic characteristics of leading events for extracting fraud evidences.
* By analyzing the Products’ historical ranking records, we observe that Products’ ranking behaviors in a leading event always satisfy a specific ranking pattern, which consists of three different ranking phases, namely, rising phase, maintaining phase and recession phase.
* Specifically, in each leading event, an Product’s ranking first increases to a peak position in the leader board (i. e., rising phase), then keeps such peak position for a period (i. e., maintaining phase), and ﬁnally decreases till the end of the event (i. e., recession phase).

**5.1.3 RATING BASED EVIDENCES**

* The ranking based evidences are useful for ranking fraud detection. However, sometimes, it is not sufficient to only use ranking based evidences. Speciﬁcally, after an Product has been published, it can be rated by any user who downloaded it. Indeed, user rating is one of the most important features of Product advertisement.
* An Product which has higher rating may attract more users to download and can also be ranked higher in the leader board. Thus, rating manipulation is also an important perspective of ranking fraud. Intuitively, if an Product has ranking fraud in a leading session s, the ratings during the time period of s may have anomaly patterns compared with its historical ratings, which can be used for constructing rating based evidences.

**5.1.4 REVIEW BASED EVIDENCES**

* Besides ratings, most of the Product stores also allow users to write some textual comments as Product reviews. Such reviews can reflect the personal perceptions and usage experiences of existing users for particular mobile Products.
* Indeed, review manipulation is one of the most important perspectives of Product ranking fraud. Specifically, before downloading or purchasing a new mobile Product, users often firstly 5, read its historical reviews to ease their decision making, and a mobile Product contains more positive reviews may attract more users to download.
* Therefore, imposters often post fake reviews in the leading sessions of a specific Product in order to in ﬂate the Product downloads, and thus propel the Product’s ranking position in the leaderboard. Although some previous works on review spam detection have been reported in recent years, the problem of detecting the local anomaly of reviews in the leading sessions and capturing them as evidences for ranking fraud detection are still under-explored.

**5.1.5 EVIDENCE AGGREGATION**

* After extracting three types of fraud evidences, the next challenge is how to combine them for ranking fraud detection. Indeed, there are many ranking and evidence aggregation methods in the literature, such as permutation based models, score based models and Dempster-Shafer rules. However, some of these methods focus on learning global ranking for all candidates.
* This is not proper for detecting ranking fraud for new Products. Other methods are based on supervised learning techniques, which depend on the labeled training data and are hard to be exploited. Instead, we propose an unsupervised Product roach based on fraud similarity to combine these evidences.