Proceedings of the Twenty-Ninth International
Florida Artificial Intelligence Research Society Conference

Nuking Item-Based Collaborative Recommenders with Power Items and Multiple Targets

Carlos E. Seminario and David C. Wilson
University of North Carolina at Charlotte
Charlotte, North Carolina, USA
cseminar@uncc.edu davils@uncc.edu

Abstract

Attacks on Recommender Systems (RS) tend to bias predictions and corrupt datasets, which may cause user distrust in the recommendations and dissatisfaction with the RS. Attacks on RSs are mounted by malicious users to "push" or promote an item, "nuke" or disparage an item, or simply to disrupt the recommendations; typically, attacks are motivated by financial gains, by a desire to "game" the system, or both. Although attack research indicates that item-based recommenders are resistant to a wide variety of push and nuke attacks, in previous work we have shown that push attacks on item-based recommenders can be effective using a multiple-target approach. In this paper, we explore nuke attacks on item-based recommenders using a multiple-target approach and variations on the Pearson Correlation calculation. We show that nuke attacks using a multiple-target approach can be configured to be effective against item-based recommenders. To evaluate the effectiveness of these attacks, we use new and existing robustness metrics and an experimental design that includes a variety of attack models, attack sizes, target item types, number of target items, and datasets.

1 Introduction

Online systems provide recommenders to help users determine which products and services to purchase. However, these systems unintentionally provide fertile ground for malicious users who, intent on gaming the system, take the opportunity to promote items ("push"), disparage items ("nuke"), or disrupt the recommender for financial gain or pleasure. Over the years, attacks on RSs have been documented in the media and, most recently, reports of "fake reviews" attacks on Amazon.com have been made public:[1] Amazon's complaint is that "While small in number, these reviews threaten to undermine the trust that customers, and the vast majority of sellers and manufacturers, place in Amazon, thereby tarnishing Amazon's brand". Fake reviews on TripAdvisor, also previously reported in the media, have been used to explicitly disparage hotel accommodations causing undue harm to hotel operators.[2]

Previous work has shown that item-based collaborative recommenders are more robust to push and nuke attacks (Lam and Riedl 2004; Mobasher et al. 2007) compared to user-based recommenders; that work used various attack models consisting of user profiles composed of items with average or randomly-selected ratings and also included single target items for attack purposes. The "power item and multiple-target" Power Item Attack (PIA-MT) model (Seminario and Wilson 2014), showed that the item-based algorithm is also vulnerable to **push attacks**. In this context, power items are those that explicitly exert a degree of influence (positive or negative) over many other items during item-based prediction calculations. Power items are selected using heuristic methods such as InDegree (based on Social Network Analysis concepts (Wasserman and Faust 1994)) and NumRatings (based on item popularity).

This study investigates the use of power items and multiple-targets to mount **nuke attacks** against item-based recommenders. Our main research question is, ***RQ1:*** *Can the Power Item Attack (PIA-MT) mount successful nuke attacks (from the attacker's viewpoint) against item-based recommenders?* And our hypothesis is, ***H1:*** *A small number of attackers ($< 5\%$ of all users) can mount successful nuke attacks (from the attacker's viewpoint) against item-based recommenders.* Success will be measured using robustness metrics outlined in Section 4.

2 Related Work

Attacks on RSs by providing false ratings have been called "shilling attacks" (Lam and Riedl 2004), or "profile injection attacks" (Mobasher et al. 2007; O'Mahony, Hurley, and Silvestre 2005). Since 2002, research in attacks on recommender systems has been performed (O'Mahony, Hurley, and Silvestre 2002) and a recent summary describes RS attack models, attack detection, and algorithm robustness (Burke, O'Mahony, and Hurley 2011). In (Wilson and Seminario 2013), a novel *Power User Attack* (PUA) model was defined to use a set of power user profiles with biased ratings that influence the results presented to other users. The

[1] http://www.reuters.com/article/2015/04/10/us-amazon-com-lawsuit-fake-reviews-idUSKBN0N02LP20150410#qY7xxdzAfg0Y5P6s.97

[2] http://www.dailymail.co.uk/travel/article-2059000/TripAdvisor-controversy-Reviews-website-launches-complaints-hotlines.html

Table 1: PIA-MT Attack User Profile Contents

$I_S, Selected\ Items$	$I_F, Filler$	$I_T, Target\ Items$
Power Items: ratings set with normal dist around *item* mean	Empty	Multiple Targets: ratings set to 1 for nuke attacks

PUA relies critically on the method of power user identification/selection, so a novel use of degree centrality concepts from Social Network Analysis (Wasserman and Faust 1994) was also developed and evaluated for identifying influential RS power users that can be used to generate synthetic users for attack purposes (Wilson and Seminario 2014).

Previous work has shown that item-based collaborative recommenders are more robust to push and nuke attacks compared to user-based recommenders (Lam and Riedl 2004; Mobasher et al. 2007). E.g., Average and Reverse Bandwagon attacks against item-based recommenders were successful albeit to a much lesser degree compared to user-based recommenders (Mobasher et al. 2007). In (Wilson and Seminario 2014; Seminario and Wilson 2014), we showed that effective push attacks using synthetic attackers emulating power users can be mounted against user-based, item-based, and SVD-based recommenders. Another study showed that effective nuke attacks can be mounted against user-based and SVD-based recommenders (Seminario and Wilson 2015); that study also indicated that item-based recommenders remained robust to nuke attacks.

Therefore, the gap in this research is whether an attack model can be configured to generate effective nuke attacks against item-based algorithms. And this remains an open question in RS attack research that we explore in this study.

3 Power Item Attack Background

In order to study RS attacks based on *explicit* measures of influence, the *Power User Attack* (PUA) model uses a set of power user profiles with biased ratings that influence the results presented to other users (Wilson and Seminario 2013). To conduct attacks against item-based recommenders, the complementary notion of the Power Item Attack (PIA) containing "power items" (Seminario and Wilson 2014) was introduced, those items are most influential in nearest-neighbor prediction calculations. The Multiple Target variant, known as PIA-MT, was also introduced to generate effective push attacks on item-based recommenders. The PIA-MT is more effective than single-target PIA and attacks multiple items simultaneously, although it can be more susceptible to detection. Power items are part of the attack user profile (a row in the RS user-item matrix), as shown in Table 1, and are used to associate with the target items also present in the profile. The combination of power items and target items present in one or more attack user profiles are then used by the PIA-MT to correlate with other items in the dataset in order to influence the prediction calculation.

Power Items: The PIA-MT relies critically on the method of power item identification/selection. In this paper, power items are selected in a manner analogous to power user selection methods described in (Wilson and Seminario 2013) and consist of the following heuristic approaches:

In-Degree Centrality (InDegree or ID): This method is based on in-degree centrality where power items are those that participate in the largest number of neighborhoods (Wasserman and Faust 1994; Lathia, Hailes, and Capra 2008). In our implementation, for each item i compute similarity with every other item j applying significance weighting $n_{cij}/50$, where n_{cij} is the number of co-ratings and 50 was determined empirically by (Herlocker et al. 1999) to optimize RS accuracy. Next, discard all but the top-k neighbors for each item i. Count the number of similarity scores for each item j (# neighborhoods item j is in), and select the top-k item j's.

Number of Ratings (NumRatings or NR): This method is based on (Herlocker et al. 2004) where "power user" refers to users with the highest number of ratings. In an analogous fashion, the top-k items (based on the total number of ratings) were selected as the power items.

Random (Rand): This method selects power items ($I_S, Selected\ Items$ in Table 1) randomly to obtain a diverse cross-section of item characteristics such as popularity (highest number of ratings) and likability (highest ratings). Although not an "influential" power item selection method per se, Rand is used primarily to compare with results obtained from InDegree and NumRatings power item selection methods. The PIA-MT Random model differs from the Random attack model (Lam and Riedl 2004; Mobasher et al. 2007) in that power item ratings are set around the item mean (rather than system mean), and the presence of multiple (rather than single) targets.

Target Items: The PIA-MT also relies critically on the target items that are selected for inclusion in the attack user profiles. In prior research (Lam and Riedl 2004; Mobasher et al. 2007; Wilson and Seminario 2013), target items were selected either randomly, because of their association with $I_S, Selected\ Items$, or to represent a diverse set of items based on their popularity, likability, and entropy (ratings dispersion). Items with low popularity or "New" items have few ratings and are usually easier to attack because their average rating can be easily manipulated by a group of attackers. From previous research (Seminario and Wilson 2014), it was shown that New target items are more vulnerable to attack than New and Established targets. For this study, we use target items that are challenging to attack in addition to those that are selected randomly: ***Most Liked (ML)*** items with the highest ratings; ***Most Popular (MP)*** items with the most number of ratings; and ***Random (RND)*** a diverse set of items selected randomly from the dataset.

The attack intent, in this case *nuke*, is applied to the multiple target items simultaneously at run time. To conduct the attacks, synthetic attack user profiles (as in Table 1) were generated that contained power items (InDegree, NumRatings, or Random) and target items (ML, MP, or RND), as described in (Seminario and Wilson 2014).

4 Evaluation Metrics

The main objective of a nuke attack in ratings-based systems is to remove target items from, or to prevent them from showing up in, users' top-N lists. To achieve this objective,

the attacker must manipulate the computed predictions for the target items so they are sufficiently reduced and, consequently, those target items are moved off the top-N lists and replaced by non-target items with higher prediction values.

To evaluate attacks on collaborative recommenders, robustness metrics were developed (Mobasher et al. 2007; Burke, O'Mahony, and Hurley 2011) such as Hit Ratio (HR) which measures the percentage of users with the target item in their top-N lists, Prediction Shift (PS) which measures the change in prediction calculation from before to after the attack, and Rank (R) which indicates the ordinal location of the target item in the top-N list. Average values for each of these metrics are calculated over all users and target items. After a successful nuke attack, we would expect to see a low $\overline{HR}$, a negative $\overline{PS}$, and a high $\overline{R}$ assuming that the target item had a higher $\overline{HR}$ and lower $\overline{R}$ before the attack. For attacks using multiple targets in the attack user profiles, the Number of Target Items Per User ($NTPU$) metric was developed (Seminario and Wilson 2014) that measures the average number of target items per user (over all target items and users) and Normalized $NTPU$ ($NNTPU$) which is the product of $\overline{HR}$ and $NTPU$ and is used to compare attack results. The $NTPU$ and $NNTPU$ metrics indicate the extent to which a multiple-target attack has successfully caused target items to appear in top-N lists; higher values indicate presence of more target items in top-N lists.

In our experimentation we found that with nuke attacks, and multiple-targets in particular, some of the above metrics may not always provide a *practical* interpretation of the attack results, i.e., when $\overline{R}$ is well beyond the top-N threshold prior to the attack (see Table 3) and remains so afterwards, does it really matter that there were small differences in $\overline{PS}$ and $\overline{R}$? Therefore, to better evaluate the attacks in this study, we have developed two new metrics that focus on the key objectives of the nuke attack:

Average HR Shift ($\overline{HRS}$): Measures the change in $\overline{HR}$ from before the attack to after the attack and is expressed as a percentage. A positive $\overline{HRS}$ would indicate an increase in $\overline{HR}$ occurred as a result of the attack. For successful nuke attacks, we expect to see negative $\overline{HRS}$.

Normalized NTPU Shift (NNS): Measures the change in $NNTPU$ from before the attack to after the attack and is expressed as a number. A positive NNS would indicate an increase in $NNTPU$ occurred as a result of the attack. For successful nuke attacks, we expect to see negative NNS.

5 Experimental Design

To address our research question and hypothesis, we conducted two experiments using the PIA-MT attack model:

Experiment 1 (E1): Power Item Multiple-Target Nuke Attack, Positive Similarities only. Positive Pearson Correlation similarity has been used widely in attack research, e.g., (Lam and Riedl 2004; Mobasher et al. 2007) to avoid potential prediction inconsistencies.[3] Our experimentation found that with this assumption, the PIA-MT was not successful against item-based recommenders; in fact, for some target item types the results resembled effective push attacks.

Experiment 2 (E2): Power Item Multiple-Target Nuke Attack, Positive and Negative Similarities. To obtain more accurate correlation between target items and other items, we adjusted the item-based algorithm to include all similarities (positive and negative) during the prediction calculation. The results of this attack are much improved over E1 and indicate that the use of full similarity correlation contributes to the effectiveness of the attack. We also note that RS accuracy (MAE) changed by small, albeit statistically significant amounts ($p < 0.001$), when moving to full correlation similarities: MAE change was +0.022 for MovieLens[4] ML100K[5] and -0.008 for ML1M[6] datasets.

Evaluation Metrics: Evaluations were performed using the metrics[7] described in § 4. The top-N list of recommendations for Hit Ratio calculations use N=40, based on analysis in (Lam and Riedl 2004) that the median recommendation search ends within the first 40 items displayed.

Datasets and Algorithms: We used MovieLens[4] ML100K[5] and ML1M[6] datasets. The CF item-based weighted (IBW) algorithm (Sarwar et al. 2001) uses Pearson Correlation similarity with a threshold of 0.0 (positive correlation) and -1.0 (positive and negative correlation), and significance weighting of n/50 where n is the number of co-rated items (Herlocker et al. 1999). We used IBW from Apache Mahout[8] and added functionality to implement similarity thresholding (0.0) and significance weighting (n/50). Also, Mahout "centers" the data for Pearson, making it mathematically equivalent to cosine similarity.

Power Item Selection: The InDegree (ID), NumRatings (NR), and Random (Rand) methods described in § 3 were used. The number of power items included in the attack user profile varied for each dataset. For ML100K, we used 166 (10% of items in the dataset), 83 (5%), and 17 (1%) power items. For ML1M, we used 184 (5% of items in the dataset), 37 (1%), and 4 (0.1%) power items.

Target Item Selection: Target items were selected as described in § 3. We varied the number of target items used for attacking each dataset. For ML100K, we used 50 (3% of items in the dataset) and 10 (0.6%) target items. For ML1M, we used 184 (5% of items in the dataset), 37 (1%), and 18 (0.5%) target items. Statistical characteristics of each target set, i.e., average number of ratings, average rating (μ), standard deviation of rating (σ), and average rating entropy (S), are given in Table 2.

Attack Parameter Selection: The Attack Intent is Nuke, i.e., target item rating is set to min (= 1). The Attack Size or number of synthetic attack user profiles in each attack varied by dataset: 50 (5% of users in the dataset) and 10 (1%) attackers for ML100K, 60 (1% of users in the dataset) and 6 (0.1%) for ML1M. Attack sizes, also expressed as $(\frac{\#attackers}{\#users} * 100)\%$, were selected based on previous re-

[3] Other researchers (Herlocker et al. 1999; Sarwar et al. 2001) used similarities $>= 0$ to improve performance and accuracy.

[4] http://www.grouplens.org

[5] nominal 100,000 ratings, 1,682 movies, and 943 users.

[6] nominal 1,000,209 ratings, 3,883 movies, 6,040 users.

[7] Note: Mean Reciprocal Rank may be explored in future work.

[8] http://mahout.apache.org

Table 2: Target Item Ratings Statistics

Target type & #	$Avg\#$	μ	σ	S
ML100K-ML-10T	1.6	5.000	0.000	0.000
ML100K-ML-50T	130.6	4.471	0.610	1.109
ML100K-MP-10T	486.3	3.753	1.019	1.935
ML100K-MP-50T	356.8	3.864	0.972	1.860
ML100K-RND-10T	104.5	3.257	0.900	1.721
ML100K-RND-50T	73.8	3.133	1.003	1.770
ML1M-ML-18T	37.1	4.885	0.177	0.330
ML1M-ML-37T	501.6	4.687	0.429	0.796
ML1M-ML-184T	614.0	4.364	0.755	1.413
ML1M-MP-18T	2508.8	4.224	0.877	1.632
ML1M-MP-37T	2228.4	4.150	0.889	1.678
ML1M-MP-184T	1420.0	3.907	0.943	1.816
ML1M-RND-18T	204.5	3.182	0.903	1.717
ML1M-RND-37T	248.1	3.236	0.953	1.800
ML1M-RND-184T	226.0	3.171	0.969	1.846

search (Mobasher et al. 2007; Burke, O'Mahony, and Hurley 2011), where a 5%-10% attack size was shown to be effective. Attack profiles were generated as described in § 3.

Test Variations: We used 2 datasets, 2 item similarity threshold values, 3 power item selection methods, 6 power item sizes, 3 target item types, and 5 attack sizes.

6 Results and Discussion

E1: Power Item Multiple-Target Nuke Attack, Positive Similarities only. To conduct this experiment, we first select target items according to one of 3 criteria (Most Liked, Most Popular, Random), select power items using one of 3 methods (InDegree, NumRatings, Random), generate synthetic attack user profiles using power items and target items, append the synthetic profiles to each of two datasets (ML100K, ML1M) and begin the attack process. To mount each attack, we iterate through each user and request recommendations until all target items have been recommended (if possible); when a target item is presented as a recommendation, we make note of the "hit" if it is presented within the top-N recommendations, and store away prediction and rank order information. When all users have been processed, we compute $NTPU$, $NNTPU$, NNS, $\overline{HR}$, $\overline{HRS}$, $\overline{PS}$, and $\overline{R}$. This process is repeated for each variation described in § 5. Baseline target item values (before the attack) for key metrics in each dataset configuration are shown in Table 3. Successful nuke attacks will have negative $\overline{HRS}$ and NNS.

Results shown in Figure 1 indicate a Hit Ratio Shift that is mostly non-negative, i.e, the attacks have not resulted in significantly reducing the number of target items in users' top-N lists. For Most Liked (ML) target items, the number of target items in the top-N lists increased or remained the same; for Most Popular (MP) targets (not shown), there is no change to an already low number of target items (see Table 3). And for Random (RND) targets we see mixed results: a slightly negative $\overline{HRS}$ for attacks with the most number of targets (184) and positive $\overline{HRS}$ for the other cases. The results in Figure 2 confirm that the attacks were not successful (from the attacker's viewpoint) with across-the-board increases in the number of target items per user for ML and RND targets; NNS for MP items was very close to zero for all variations. For ML100K (not shown), we also observed across-the-board increases in $\overline{HRS}$ and NNS, indicating ineffective nuke attacks.

Table 3: Baseline Target Item Metrics Before Attack

Target type & #	$\overline{HR}$	$\overline{R}$	Avg $Rating$	$NNTPU$
Experiment 1				
ML1M-ML-18T	49.2%	902	3.943	0.648
ML1M-ML-37T	76.4%	1125	3.867	1.412
ML1M-ML-184T	92.5%	1181	3.834	2.598
ML1M-MP-18T	0.0%	1675	3.741	0.000
ML1M-MP-37T	0.3%	1661	3.746	0.003
ML1M-MP-184T	1.8%	1756	3.733	0.024
ML1M-RND-18T	8.7%	1736	3.718	0.088
ML1M-RND-37T	20.1%	1704	3.728	0.215
ML1M-RND-184T	73.7%	1720	3.725	1.183
Experiment 2				
ML1M-ML-18T	17.4%	1178	3.873	0.182
ML1M-ML-37T	46.0%	1216	3.863	0.575
ML1M-ML-184T	87.9%	1223	3.867	2.365
ML1M-MP-18T	0.5%	1617	3.756	0.006
ML1M-MP-37T	2.0%	1599	3.767	0.022
ML1M-MP-184T	12.3%	1671	3.749	0.207
ML1M-RND-18T	13.4%	1721	3.685	0.143
ML1M-RND-37T	15.9%	1700	3.695	0.171
ML1M-RND-184T	86.5%	1717	3.685	1.829

Overall, the E1 results did not indicate effective nuke attacks, i.e., there was little or no decline in the number of target items in users' top-N lists. In many cases, results are more similar to push rather than nuke attacks[9], i.e., positive $\overline{HRS}$ and NNS. Several factors contribute to this, such as strength of the attack (i.e., attack size, number and influence of attack profile items), number and characteristics of targets, and Pearson Correlation similarity (i.e, positive-only vs. positive and negative). Results in Figure 1 indicate that relatively weak attacks (fewer number of targets) are unable to impact the high Hit Ratio of target items before the attack (Table 3). In addition, ratings distribution of low ratings in the dataset (about 6% 1's and 11% 2's in both ML100K and ML1M) can have a significant impacts on predictions (Burke, O'Mahony, and Hurley 2011). Since IBW predictions are based on items in the user profile that are similar to the target item(s), there are fewer opportunities for user profile items to correlate highly with 1-rated target items and is further exacerbated by using positive-only Pearson correlations. Hypothesis H1 cannot be accepted for E1 (ML100K, ML1M) since no attack configuration achieved significant negative $\overline{HRS}$ and NNS results.

E2: Power Item Multiple-Target Nuke Attack, Positive and Negative Similarities: The E2 experiment was conducted using the same process as E1. The only change

[9]Positive Prediction Shift for nuke attacks has also been observed before (Mobasher et al. 2007; Seminario and Wilson 2015).

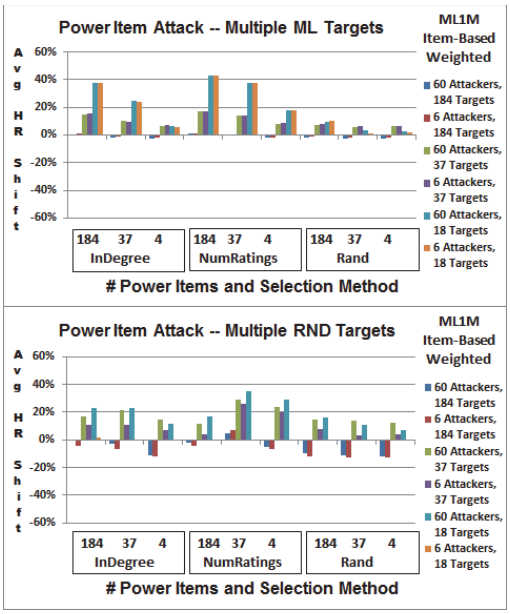

Figure 1: E1 – Average Hit Ratio Shift using ML1M with Positive Pearson Correlation Similarity

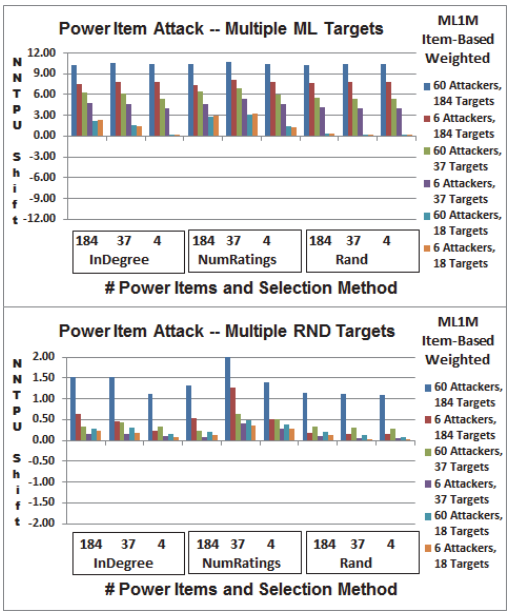

Figure 2: E1 – Normalized NTPU Shift using ML1M with Positive Pearson Correlation Similarity

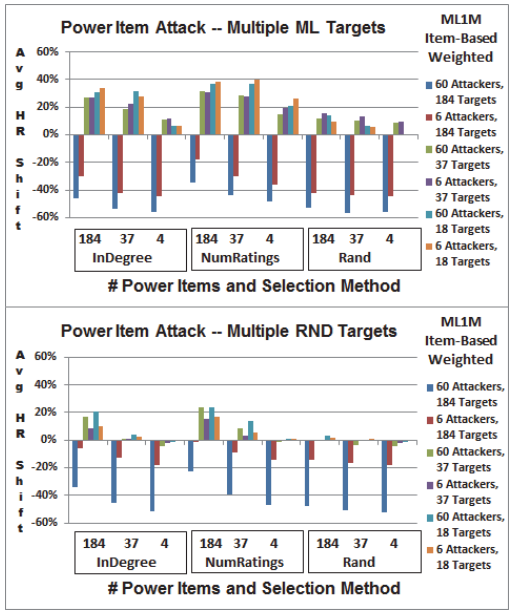

Figure 3: E2 – Average Hit Ratio Shift using ML1M with Positive and Negative Pearson Correlation Similarity

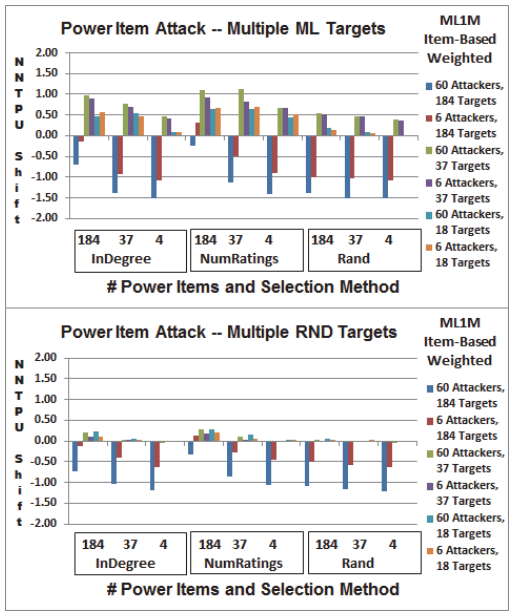

Figure 4: E2 – Normalized NTPU Shift using ML1M with Positive and Negative Pearson Correlation Similarity

was to use the full range of Pearson similarities, both positive and negative. The objective was to exploit the negative correlations between power items (typically rated above average) and target items (ratings set to 1 for nuking). Results in Figures 3 and 4 indicate that the attacks with the highest number of target items (184) were quite effective in removing target items from top-N lists and reducing the number of target items per user for all target item and power item types. We also observed that, for ML1M, the boundary between effective and non-effective multiple-target nuke attacks lies between 37 and 184 targets (1% to 5% of all items), and 6 and 60 users (0.1% to 1% of all users). Although not tested here, it appears that effective attacks could be mounted with $< 5\%$ of all items (as targets) thus avoiding detection. An interesting result to note is that $\overline{HRS}$ and NNS display a phenomenon simi-

lar to that observed in previous work (Mobasher et al. 2007; Seminario and Wilson 2014), i.e., as the number of power items *decreases* (from 184 to 37 to 4 in Figures 3 and 4), the attack effectiveness *increases*. This occurs for all three power item selection methods at the level of 184 targets and may be caused by including too many item ratings that could make the profile dissimilar to a given user (Mobasher et al. 2007).

For ML targets, NNS was reduced significantly for all target item levels compared with E1 results. The $\overline{HRS}$ for the highest target item size (184) was also significantly improved from E1. For MP targets (not shown), there was a significant reduction in the $\overline{HRS}$ and NNS metrics albeit from a small number of hits. For RND targets, the improvement in the $\overline{HRS}$ and NNS metrics was also significant. For ML100K (not shown), we observed significant reductions (14%) in $\overline{HRS}$ and NNS metrics for MP and RND target items for 50 target item attacks; for ML targets, there were significant increases in $\overline{HRS}$ and NNS indicating ineffective attacks. The use of positive and negative correlations for prediction calculations contributed significantly to attack effectiveness in E2. However, as found in E1, results for middle and low end target item sizes (37 and 18) still indicate characteristics of push rather than nuke attacks and will be analyzed in future work.

We recognize the fact that attackers are unable to set similarity parameters in publicly-available recommenders, however, we note for system operators that *this particular attack indicates a vulnerability in the item-based algorithm that can result in either push or nuke impacts*. From an attacker's perspective, a low-cost[10] and effective attack is the goal. Although attackers may find it difficult, albeit not impossible, to specify ID power items, finding NR power items as well as ML and MP target items should be simple using publicly-available data. We also note for system operators that *a low cost/knowledge PIA-MT attack with Rand power items and RND targets can result in attacks as effective as those using InDegree and NumRatings power items.* Our hypothesis H1 is accepted for E2 (partially for ML100K, fully for ML1M) since all attack configurations (InDegree, NumRatings, Random) were able to achieve significant reductions in $\overline{HRS}$ and NNS results using a small number of attackers.

7 Conclusion

This paper evaluated power item nuke attacks against item-based collaborative recommenders using new and existing robustness metrics. System operators should note that the use of positive and negative Pearson Correlation similarity can enable an attacker's ability to mount effective PIA-MT nuke attacks against an item-based recommender. Results also indicate that weaker nuke attacks have similar robustness characteristics as effective push attacks, i.e., the number of target items appearing in users' top-N lists increase after attack. Our future work in this area will be to model how the weak multiple-target nuke attacks on item-based recommenders morph into push attacks, and to further investigate the effectiveness of the Rand/RND PIA-MT attack compared to the InDegree and NumRatings variants.

[10]Cost of generating attack user profiles and mounting the attack.

References

Burke, R.; O'Mahony, M. P.; and Hurley, N. J. 2011. Robust collaborative recommendation. In Ricci, F., et al., eds., *Recommender Systems Handbook*. Springer.

Herlocker, J. L.; Konstan, J. A.; Borchers, A.; and Riedl, J. 1999. An algorithmic framework for performing collaborative filtering. In *Proc of the ACM SIGIR Conf*.

Herlocker, J. L.; Konstan, J. A.; Terveen, L. G.; and Riedl, J. 2004. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*.

Lam, S. K., and Riedl, J. 2004. Shilling recommender systems for fun and profit. In *Proceedings of the 13th international conference on World Wide Web*. ACM.

Lathia, N.; Hailes, S.; and Capra, L. 2008. knn cf: A temporal social network. In *Proceedings of the 2nd ACM Recommender Systems Conference (RecSys '08)*.

Mobasher, B.; Burke, R.; Bhaumik, R.; and Williams, C. 2007. Toward trustworthy recommender systems: An analysis of attack models and algorithm robustness. *ACM Trans. Internet Technol.*

O'Mahony, M. P.; Hurley, N.; and Silvestre, G. C. M. 2002. Promoting recommendations: An attack on collaborative filtering. In *Proceedings of DEXA'02*.

O'Mahony, M. P.; Hurley, N.; and Silvestre, G. C. M. 2005. Recommender systems: Attack types and strategies. In *Proceedings of the 20th National Conference on Artificial Intelligence (AAAI-05)*.

Sarwar, B.; Karypis, G.; Konstan, J.; and Riedl, J. 2001. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the World Wide Web Conference*.

Seminario, C. E., and Wilson, D. C. 2014. Attacking item-based recommender systems with power items. In *Proceedings of the 8th ACM Conference on Recommender Systems*, RecSys '14. ACM.

Seminario, C. E., and Wilson, D. C. 2015. Nuke 'em till they go: Investigating power user attacks to disparage items in collaborative recommenders. In *Proceedings of the 9th ACM Conference on Recommender Systems*, RecSys '15. ACM.

Wasserman, S., and Faust, K. 1994. *Social Network Analysis: Methods and Applications*. New York, NY: Cambridge University Press.

Wilson, D. C., and Seminario, C. E. 2013. When power users attack: assessing impacts in collaborative recommender systems. In *Proceedings of the 7th ACM conference on Recommender Systems*, RecSys '13. ACM.

Wilson, D. C., and Seminario, C. E. 2014. Evil twins: Modeling power users in attacks on recommender systems. In *Proceedings of the 22nd Conference on User Modelling, Adaptation and Personalization*.