The Force Within: Recommendations Via Gravitational Attraction Between Items

Vikas Kumar
University of Minnesota
Minneapolis, Minnesota, USA
vikas@cs.umn.edu

Saeideh Bakhshi
Facebook Inc.
Menlo Park, California, USA
saeideh@gatech.edu

Lyndon Kennedy
Futurewei Technologies, Inc.
Santa Clara, California, USA
lyndonk@acm.org

David A. Shamma
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands
aymans@acm.org

ABSTRACT

Recommendation systems rely on various definitions of similarities. These definitions while having numerous design factors in different domains help identify and recommend relevant content. For example, similarity between users, or items, are measured based on, but not limited to, explicit feedback such as ratings, thumbs up; or/and implicit feedback such as clicks, views etc; or/and based on composition of item such as tags, metadata etc. In this paper, we explore a similarity model while very intuitive to find similar items using a very common natural law of attraction between bodies, that is gravitational law. We show how the two attributes, relative mass and distance between the bodies, of gravitation law can be interpreted for an effective personalized recommendations; in both spatial and non-spatial domains. Finally, we illustrate the use of distance and mass in a non-spatial domain and we exhibit the accuracy in recommendations against popular baselines.

CCS CONCEPTS

• **Information systems → Collaborative filtering**; **Similarity measures**; **Recommender systems**; *Personalization*;

KEYWORDS

Recommender Systems; Newtons's Gravitational Law; Similarity; MovieLens.

1 INTRODUCTION

With numerous independent online content providers, such as Facebook, Netflix, Spotify, and Amazon, people are often faced with a problem of needle in haystack. In such large information spaces they often rely on some help to find relevant content. Personalized recommendation systems among other alternatives play a crucial role in making such web services more usable and engaging for their users by identifying more meaningful and relevant contents, directly impacting the success and revenue of such businesses [2].

In most of the recommendation systems, the underlying model rely on similarities either in preferences of users or in items to find relevant and meaningful content [3, 13, 15]. For example, Amazon provides recommendation of items *similar* to recently browsed/bought, or based on alike users who have bought similar items. Furthermore, studies show descriptive similarities allow recommendations to be more transparent to users with potential of explanations [7]. For example, Pandora[1], a music recommendation site, provides a short explanation note on similarity in tonality, rhythm or use of chorus with user past preferences to explain why a song is chosen for the station.

There exist various models that further explore multiple dimensions of similarity, for example, location-similarity of users [11], content [17] or even latent factors [10]. However, recommendation systems literature rarely considered a model based on the natural law of attraction. In this paper, we take a step back on complexity of algorithms and define what we believe resonates with one of the most natural way of determining similarity between items—a

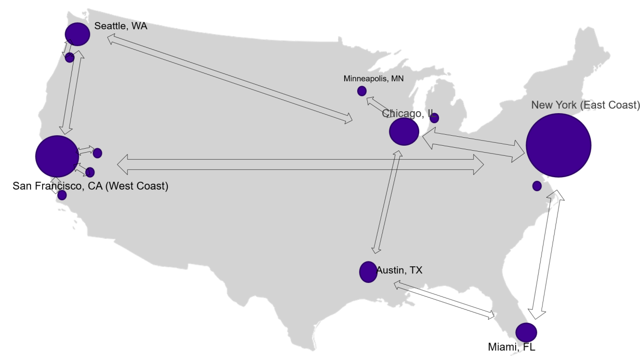

Figure 1: San Francisco is more similar to New York and Chicago (other bigger cities); and, places like Berkeley or New Jersey (read: moons) are more attracted to their nearby bigger cities San Francisco and New York respectively.

This research was conducted while the authors were at Yahoo Labs and Flickr.

UMAP'17, July 9–12, 2017, Bratislava, Slovakia

978-1-4503-4635-1/17/07…$15.00
DOI: http://dx.doi.org/10.1145/3079628.3079694

[1] http://www.pandora.com

method that provides recommendations using similarity defined by Newton's universal law of gravitation.

The gravitational law measures affinity between two items based on their relative mass and distance between them. It suggests that a bigger or massive body have more influence than smaller bodies; and that, this influence reduces as bodies move farther apart from each other. For example, a smaller body like a moon to a planet even though being closer have lesser influence on the planet than other massive bigger bodies like its parent star or another giant planet. However, with respect to moon, it has the biggest influence from its massive mother planet. It is this relation between the bodies with respect to their mass and distance that allow to explore a natural form of similarity model between items in the space of recommendation systems. Intuitively, it is not hard to find nature of similar effects in recommendation systems. For instance, the rating similarity or influence between massively popular movies like *The Godfather* and *The Shawshank Redemption* is high, though very different in storyline, as they are rated by *most* of the users rather than just *alike* users. At the same time, *The Godfather*, a mafia movie, influence many other not-so-popular mafia movies like *Mean Streets*[2]. Likewise, in spatial-domain such as location-based restaurant recommendations one can easily see why San Francisco (SF), a large metropolitan and distinct urban-economy lifestyle very similar to New York (NYC) are more influential to each other than their neighboring cities. On the other hand, the relatively smaller cities like Berkeley (near SF) and New Jersey (near NYC) are primarily influenced by their neighboring bigger cities SF and NYC respectively than each other as illustrated in Figure 1 (analogous to influence we observe between moons and planets versus planets to planets).

In this paper, we model similarity between items in recommendation systems using gravitational model and compare against traditional definitions. We explore how two main features of this model, relative mass and distance, are open for interpretations and can be defined in various ways. For example, mass can be measured based on, but not limited to, ratings, cast, size, popularity, box-office collection, page-rank etc. Whereas, distance between items can be measured in one or more information spaces such as, but not limited to, geodesic distance for spatial items such as restaurants and POIs; and semantic distance for non-spatial items such as music, movies, and text. Finally, we conclude with results from one of the non-spatial domain of movie recommendations and discuss the ease of explain-ability of the model.

2 SIMILARITY MODELS

There exist multiple methods in recommendation literature to define similarity between two items, commonly referred to as neighbor based approach. For example, cosine-similarity calculates the similarity between items based on dot product of two vectors, each can be representing user ratings, or views/clicks etc. Likewise, there exists other common methods such as Jaccard, Pearsons as well as Log-likelihood for defining similarity [12] between items. Furthermore, similarity models can be a combination across multiple dimensions of similarity to calculate a more comprehensive value, such as, a linear combination of similarity between explicit feedback (user rating vectors), implicit feedback (views/clicks) or content (tags), as shown in Equation 1, where co-efficients α, β, γ can be learned using regression techniques [8].

$$\begin{aligned} \text{similarity}(i,j) =& \alpha \times \text{rating}_{\text{sim}}(i,j) \\ &+ \beta \times \text{views}_{\text{sim}}(i,j) \\ &+ \gamma \times \text{tags}_{\text{sim}}(i,j) \end{aligned} \tag{1}$$

More recent models calculate similarity among items based on latent factors, a model common with matrix factorization techniques [10]. There exists other sophisticated techniques such as tensor factorization [9] and restricted Boltzmann machines [14], capable to learn similarities from latent correlations between items to optimize for accurate recommendations [1]. However, though accurate they add more complexity to the understanding of similarity limiting the power of explainable recommendations. We believe that a method that resonates with the most natural way of defining affinity is like recognizing missing side of same coin.

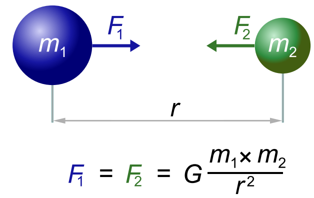

Figure 2: Newton's Law of Universal Gravitation. 🅭🅯🄎 Dennis Nilsson, Wikipedia

3 GRAVITATIONAL MODEL

Newton, around three centuries ago, defined one of the most common law of physics between two bodies known as gravitational law of attraction. The law defines a force of attraction (F) between two bodies proportional to the product of their masses (m_1 and m_2) and inversely proportional to the square of distance (r) between their centers as illustrated in Figure 2.

Likewise, we postulate that a similar model could help us describe the affinity between items that exist in the information space of online systems. An advantage of such model lies in open interpretation of relative effect size and distance. Leveraging multiple definitions implicitly favors the model to be adaptable in diverse information spaces. For example, the *mass* of an item can be represented in both simple scalar or vector forms. The basic requirement can be that it reflects the relative size (or importance), such as popularity of item, with respect to other items. For instance, in music recommendation, mass of an artist can be based on number of ratings, number of views or number of plays etc. A vector mass of the artists can also be modeled as a representation of item size over multiple dimensions. For example, artist represented with popularity in specific genres of music, or preference with specific cluster of users[4]. It can also be combination of heterogeneous parameters such as brand value, number of followers, number of recent popular releases etc.

Similar to mass, relative *distance* can also be represented in a scalar or vector form. For spatial items, such as restaurants, POIs,

[2] The similarities are as observed on MovieLens platform.

etc, the geodesic distance provide an easy alternative. However, for non-spatial items, it can be challenging. One alternative is based on semantic dissimilarity measured using the content or composition of items. The available metadata for items, or user provided tags can play an important role in defining the semantic distance. We discuss the tag-based semantic distance in our experiment. As an alternative, we believe Wikipedia with its global reach on various topics can determine the semantic relatedness [6] for various types of items in online information space. Atlasify[3], an open source project using Wikibrain API[4], is an example on how semantic relatedness can help determine the distance between items with no spatial footprint.

4 EXPERIMENT

To study if the gravitational model is effective, we setup an offline experiment using MovieLens 10M ratings dataset. The dataset is publicly available from MovieLens platform[5]. It consists of 10 million ratings and 100,000 tags for 10,000 movies by 72,000 users.

4.1 Model

We use the popularity of movies, measured by log of number of ratings, as value for relative size or mass of the movie. That is, more the number of ratings, more popular the movie is and thus higher is the mass of the movie. The log is used to minimize the effect of movies like *Toy Story* and *The Matrix* which are like black holes due to their massive popularity compared to other items in the system.

For distance, movies being non-spatial items we use semantic dis-similarities. Using tag applications we calculate the similarity in user perceived composition of movies using cosine. Each movie is represented as vector, of equal length, with series of relevance value as determined by tag genome [16] for each tag-movie pair. Higher the similarity between movies in this semantic space, lesser is the distance between them. Finally, using the Newton's model, we combine the mass and distance to determine the gravitational similarity[6] between two movies, m_i and m_j, as shown in Equation 2 as $g_{\text{sim}}(m_i, m_j)$:

$$g_{\text{sim}}(m_i, m_j) = \frac{\text{mass}(m_i) \times \text{mass}(m_j)}{[\text{distance}(m_i, m_j)]^2} \tag{2}$$

where:

$\text{mass}(m_i) = \log(\text{num\_ratings}_{m_i})$
$\text{distance}(m_i, m_j) = (1 - \text{tag}_{\text{sim}}(m_i, m_j))$

4.2 Personalized Recommendations

To generate personalized recommendation for each user, traditional item-item collaborative filtering technique is used where instead of the usual cosine similarity between items, we use the gravitational similarity in the equation to predict user (u) ratings for an unknown movie (m) where $\text{Nb}(m)$ is the Neighbors of m.

$$\text{pred}(u, m) = \frac{\sum_{i \epsilon \text{Nb}(m)} g_{\text{sim}}(i, m) \times rating(i, u)}{\sum_{i \epsilon \text{Nb}(m)} g_{\text{sim}}(i, m)} \tag{3}$$

For top-N recommendations we select the N most highly rated predictions from the list of each user. Finally, we compare the prediction and recommendation accuracy of our model against traditional techniques, discussed in next section.

4.3 Metrics

For prediction accuracy, we use 90% of the 10 million ratings for learning similarities and use rest 10% for test. As a metric, the traditional RMSE is used to determine accuracy of the predictions. We compare our results against other well known collaborative filtering techniques i.e., ItemItem, UserUser, and Matrix Factorization. For recommendation accuracy, we determine Precision, Recall and Mean Average Precision for each test user in the dataset. The test data is created by randomly sampling 10% of users. We hide 20% of their ratings and use rest 80% for training purpose. A recommendation is considered relevant if we find user has rated the movie 4 stars or higher in 20% of hidden ratings. Final metrics are average of observed values over all test users. Similar to prediction accuracy, we again compare the results with other baselines.

5 RESULTS

The results for both prediction and recommendation metrics are shown in Table 1. Lower RMSE for Gravity model at 0.92 compared to traditional collaborative filtering techniques i.e. ItemItem and UserUser, is a significant result[7]. Performing better than traditional CF techniques highlights the effectiveness of this similarity model in predicting ratings. Nevertheless, we find matrix factorization still a hard algorithm to beat in predictions and stands out with best RMSE of 0.91.

But, as known, RMSE numbers do not really paint the right picture of accurate recommendations [2]. Real users rarely notice difference in predicted ratings of a movie from 4.5 to 5.0 while, such minuscule differences play a huge role on overall RMSE. TopN recommendations in such cases provide a better alternative. Related recommendation metrics Precision, Recall and MAP thus help measure how well a set of recommendation is relevant for users, and Gravitational model stands out against all the other techniques. We observe statistically significant ($p < 0.001$) and much higher level of accuracy on each of the metric of Precision (0.08), Recall (0.28) and MAP (0.275).

We believe the significant improvement is in part due to the bias of gravitational model to popular items. A popular movie in general is highly likely to be rated by users. This popularity bias, though questionable, and probably a concern with ideal recommendations, we believe capture a very important implicit bias of users. A recent work by Harper et al. [5] confirms this implicit bias in users choice for recommendations on same MovieLens platform. They show users to choose recommendations that provide more popular items over recommendations with less popular ones.

We also study two more metrics, shown in Table 1, namely diversity and spread. With implicit bias towards popular items, we would expect the recommendations to be less diverse. We find this to be true for Gravitational model compared to other techniques but not worse than UserUser. However, with better spread, i.e. how

[3] http://www.atlasify.com/about.html
[4] https://shilad.github.io/wikibrain/
[5] http://grouplens.org/datasets/movielens
[6] The similarity values are normalized between 0.0 and 1.0 post the calculation.

[7] The difference in predicted ratings are found to be statistically significant with p-value<0.001 in Kruskal-Wallis test.

Table 1: RMSE and TopN results in MovieLens 10M dataset, best performance marked as (*). Gravitational based model does significantly well on recommendation metrics (than prediction metrics) for top 20 recommendations with comparable diversity and spread.

Metrics	ItemItem	UserUser	MatrixF	Gravity
RMSE	1.030	1.060	0.910*	0.920
Precision@20	0.001	0.019	0.010	0.080*
Recall@20	0.001	0.060	0.040	0.280*
MAP@20	0.024	0.088	0.046	0.275*
Intra-list Diversity@20	0.960	0.022	0.315	0.024
Spread@20	0.310	0.100	0.120	0.140

many distinct items does recommendation able to cover while recommending for users reflect that our preliminary model of gravitational similarity to be an effective algorithm for recommendations.

6 CONCLUSION

To our knowledge, this is the first work to consider and discuss the gravity and effective definition of Newton's law of attraction in recommendation systems. We propose and show the how using gravitational model in design of similarity result in efficient recommendations. We also discuss how With possible alternate interpretations of effective size and distance, the expression of gravitational model can be adapted for various recommendation domains.

Nevertheless, we also recognize certain limitations to this work that we can address in future. In our current approach, we do not consider or measure the nature of new items that model can recommend, which can potentially impact the design of recommendation. We believe that an online experiment with actual users would clearly benefit our model to further understand the role of gravity in recommendations. We also aim to study the model in various other domains like news, music, restaurants as we believe that these domains will exhibit the implicit popularity bias, an important factor observe in the gravitational model.

We further believe that there is a valuable byproduct of the gravitational model in online information spaces; that is, Visualization. With the items modeled by their intuitive size and distance between them, cluster of items could be visualized for navigational purposes in the same way as we navigate through solar systems in a galaxy (ex: Figure 1). One can choose or navigate through stars (most influential item of the cluster) and find similar planets (other influential items in cluster) and their moons (least influential but highly similar) to further explore categories, or other nearby influential star.

REFERENCES

[1] Amir Albadvi and Mohammad Shahbazi. 2009. A hybrid recommendation technique based on product category attributes. *Expert Systems with Applications* 36, 9 (2009), 11480–11488.

[2] Xavier Amatriain. 2016. Past, Present, and Future of Recommender Systems: An Industry Perspective. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*. ACM, 1–1.

[3] Robin Burke. 2002. Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction* 12, 4 (2002), 331–370.

[4] Evangelia Christakopoulou and George Karypis. 2016. Local Item-Item Models For Top-N Recommendation. In *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16)*. ACM, New York, NY, USA, 67–74. DOI: http://dx.doi.org/10.1145/2959100.2959185

[5] F. Maxwell Harper, Funing Xu, Harmanpreet Kaur, Kyle Condiff, Shuo Chang, and Loren Terveen. 2015. Putting Users in Control of Their Recommendations. In *Proceedings of the 9th ACM Conference on Recommender Systems (RecSys '15)*. ACM, New York, NY, USA, 3–10. DOI: http://dx.doi.org/10.1145/2792838.2800179

[6] Brent Hecht, Samuel H Carton, Mahmood Quaderi, Johannes Schöning, Martin Raubal, Darren Gergle, and Doug Downey. 2012. Explanatory semantic relatedness and explicit spatialization for exploratory search. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*. ACM, 415–424.

[7] Jonathan L Herlocker, Joseph A Konstan, and John Riedl. 2000. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*. ACM, 241–250.

[8] LEE Jung-Hyun. 2004. User preference mining through hybrid collaborative filtering and content-based filtering in recommendation system. *IEICE TRANSACTIONS on Information and Systems* 87, 12 (2004), 2781–2790.

[9] Alexandros Karatzoglou, Xavier Amatriain, Linas Baltrunas, and Nuria Oliver. 2010. Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering. In *Proceedings of the fourth ACM conference on Recommender systems*. ACM, 79–86.

[10] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *Computer* 8 (2009), 30–37.

[11] Justin J Levandoski, Mohamed Sarwat, Ahmed Eldawy, and Mohamed F Mokbel. 2012. Lars: A location-aware recommender system. In *Data Engineering (ICDE), 2012 IEEE 28th International Conference on*. IEEE, 450–461.

[12] Sean Owen, Robin Anil, Ted Dunning, and Ellen Friedman. 2011. *Mahout in action*. Manning Shelter Island.

[13] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. 1994. GroupLens: an open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*. ACM, 175–186.

[14] Ruslan Salakhutdinov, Andriy Mnih, and Geoffrey Hinton. 2007. Restricted Boltzmann machines for collaborative filtering. In *Proceedings of the 24th international conference on Machine learning*. ACM, 791–798.

[15] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. 2001. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*. ACM, 285–295.

[16] Jesse Vig, Shilad Sen, and John Riedl. 2012. The tag genome: Encoding community knowledge to support novel interaction. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 2, 3 (2012), 13.

[17] Valentina Zanardi and Licia Capra. 2008. Social ranking: uncovering relevant content using tag-based recommender systems. In *Proceedings of the 2008 ACM conference on Recommender systems*. ACM, 51–58.