2011 IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing

Supervised Rank Aggregation for Predicting Influencers in Twitter

Karthik Subbian and Prem Melville
IBM T.J. Watson Research Center, Yorktown Heights, NY 10598
Email: {ksubbian,pmelvil}@us.ibm.com

August 30, 2011

***Abstract*—Much work in Social Network Analysis has focused on the identification of the most important actors in a social network. This has resulted in several measures of influence and authority. While most of such sociometrics (e.g., PageRank) are driven by intuitions based on an actors location in a network, asking for the "most influential" actors in itself is an ill-posed question, unless it is put in context with a specific measurable task. Constructing a predictive task of interest in a given domain provides a mechanism to quantitatively compare different measures of influence. Furthermore, when we know what type of actionable insight to gather, we need not rely on a single network centrality measure. A combination of measures is more likely to capture various aspects of the social network that are predictive and beneficial for the task. Towards this end, we propose an approach to supervised rank aggregation, driven by techniques from Social Choice Theory. We illustrate the effectiveness of this method through experiments on a data set of 40 million Twitter users.**

***Keywords*-Social Network, Twitter, Influence Prediction, and Rank Aggregation.**

I. Introduction

The rise of Social Media, with its focus on user-generated content and social networks, has brought the study of authority and influence in networks to the forefront. For companies and other public entities, identifying and engaging with influential authors in social media is critical, since any opinions they express could rapidly spread far and wide.

Following this need, there has been a spate of recent work studying influence and the diffusion of information in social networks [1], [2], [3]. While these works are important in furthering our understanding of the dynamics of communication in networks, they do not directly give us measures of influence and authority in social media. On the other hand, there has been much work in the field of Social Network Analysis, from the 1930's [4] onwards, that has focused explicitly on sociometry, including quantitative measures of influence, authority, centrality or prestige. These measures are heuristics usually based on intuitive notions such as access and control over resources, or brokerage of information [5]; and has yielded measures such as Degree Centrality, Eigenvector Centrality and Betweeness Centrality [6].[1]

In this paper, we address the problem of identifying influence by posing it as a predictive task. In particular, we compare different measures of influence on their ability to accurately predict which users in Twitter will be virally rebroadcast (*retweeted*) in the near future. Formulating a concrete predictive task, such as this, allows us to quantitatively compare the efficacy of different measures of influence.

In addition to evaluating individual measures of influence, such as Degree Centrality and PageRank, we propose combining them to produce a more accurate measure of influence. Given that each measure produces an ordering of elements, we can leverage rank aggregation techniques from Social Choice Theory, such as Borda [8] and Kemeny optimal rank aggregation [9]. These classical techniques were designed to combine rankings to ensure *fairness* amongst voters and not to maximize performance on a predictive task; and as such are *unsupervised*. In this paper, we introduce Supervised Kemeny Ranking in order to aggregate individual rankings for the task of predicting influence in networks. We demonstrate the effectiveness of our approach in a case study of 40 million Twitter accounts. We also corroborate these results in a study of publication citation networks in [7].

In this paper, we make the following key contributions: (1) We propose a predictive, rather than a heuristic, perspective of influence, by formulating measurable predictive tasks. (2) We combine ideas from Sociometry and Social Choice Theory in novel ways. (3) We present a new approach to supervised rank aggregation. (4) We show the effectiveness of our approach on real-world network data. (5) We demonstrate that our approach is significantly better than current practice and other baselines that we devised.

II. Data Set and Task Definition

Our study was based on the Twitter discussion around Pepsi. What piqued our interest in Twitter and the role

[1] We have a detailed discussion of related work in the extended version of this paper [7].

978-0-7695-4578-3/11 $26.00 © 2011 IEEE
DOI

of influencers was the infamous sexist iPhone app called "AMP UP B4 U SCORE". An avalanche of Twitter users slammed the app ultimately leading to an apology from Pepsi. In this study, we found that the influence of twitter users heavily depends upon the number of rebroadcasts of his/her messages to millions of other users. In the context of Twitter, this suggests that a useful task would be to predict which twitterers will be significantly rebroadcast via retweets.

One obvious indicator of influence could be the number of followers a user has (in-degree of the Follower Graph). However, many users follow 100K or more users and therefore this may not be sufficient indication of influence. For this reason, we consider two alternatives, the Retweet Graph and the Mention Graph, where edges correspond to retweets and mentions of users in the past. We generate two versions of both the Retweet and the Mention Graph, one collapsing all repeat connections from the same user *i* to the user *k* into just one edge. The second version uses the number of retweets/mentions as edge weights. For our influence measures (rankings) we use in-degree, out-degree and PageRanks (with a damping factor of 0.85).

We extracted the data to generate these graphs over a two week period from 11/11/09 to 11/26/09. This gives a Follower Graph with 40 million nodes (users) and 1.1 billion edges. We used the socio-metrics computed from these graphs to predict which users will have viral outbursts of retweets in the following week. We compare these predictions with the actual amount of retweets in the following week. For the purposes of testing, we monitored all retweets of a set of 9,625 users. This is the set we use for the train-test splits in our experiments.

We construct our prediction task from our data by dividing users in our test period into two classes – people who have been retweeted more than a threshold and below. In our data set, we selected 10% of the maximum number of retweets within a week as the threshold (100 retweets). We treat this as a binary classification problem, where the ranking produced by each measure is used to predict the potential for viral retweeting in the test time period. Since we are primarily concerned with how well these measures perform at ranking users, we compare the area under the ROC curve (AUC) based on using each measure [10].

We compared all measures of influence averaged over 20 trials of random stratified samples of 80% of the users (see Table I). We find that 9 of the 13 individual measures by themselves are quite effective at ranking the top potentially viral twitterers with an AUC $> 80\%$. Not surprisingly, the number of times that someone has been retweeted in the recent past produces very good rankings based on AUC. The number of followers and the number of people mentioned also produce reasonably good rankings

Measure	Definition	AUC
Followers	Follower Graph Indegree	88.18
Friends	Follower Graph Outdegree	76.03
Follower Pagerank	Follower Graph Pagerank	85.77
Distinct Past Retweets	Retweet Graph Indegree	90.17
People Retweeted	Retweet Graph Outdegree	87.04
Retweet Pagerank	Retweet Graph Pagerank	88.38
Past Retweets	Wtd. Retweet Indegree	90.18
Retweets Made	Wtd. Retweet Outdegree	86.80
Distinct Mentions Received	Mention Graph Indegree	60.71
People Mentioned	Mention Graph Outdegree	86.11
Mention Pagerank	Mention Graph Pagerank	70.43
Mentions Received	Wtd. Mention Indegree	60.53
Mentions Made	Wtd. Mention Outdegree	84.69

TABLE I
COMPARING RANKING MEASURES FOR IDENTIFYING VIRAL POTENTIAL, IN TERMS OF AUC(%)

in terms of AUC.[2] However the Spearman rank correlation between recent past retweets and followers is not very high (0.43), suggesting that there are multiple forces at work here.

III. Rank Aggregation

As each socio-metric captures only some aspect of the user's influence in the network, it is beneficial to combine them in order to more accurately identify influencers. One straightforward approach to combining individual measures is to use them as inputs to a classifier, such as logistic regression, which can be trained to predict the target variable (e.g., future retweets) on historical or held-out data. However, given that the individual influence measures produce an ordering of elements and not just a point-wise score, we can, instead leverage approaches to aggregating rankings for better results. The problem of rank aggregation or preference aggregation has been extensively studied in Social Choice Theory, where there is no *ground truth* ranking, and as such are unsupervised. In this section, we explain the necessary background for appreciating our proposed method Supervised Kemeny Ranking, which is a supervised order-based aggregation technique, that can be trained based on the ground-truth ordering of a subset of elements.

The Rank Aggregation Task: Let us begin by formally defining the task of rank aggregation. Given a set of entities S, let V be a subset of S; and assume that there is a total ordering among entities in V. We are given r individual rankers $\tau_1, ..., \tau_r$ who specify their order preferences of the m candidates, where m is size of V, i.e., $\tau_i = [d_1, ..., d_m], i = 1, ..., r, \text{if} d_1 > ... > d_m, d_j \in V, j = 1, ..., m$. If d_i is preferred over d_j we denote that by $d_i > d_j$. Rank aggregation function ψ takes input orderings from r rankers and gives τ, which is an aggregated ranking order. If V equals S, then τ is called a *full list* (total ordering), otherwise it is called a *partial list*

[2]Despite its popularity, PageRank does not perform as well as other measures.

(partial ordering).

All commonly-used rank aggregation methods, satisfy one or more of the following desirable $goodness$ properties: Unanimity, Non-dictatorial Criterion, Neutrality, Consistency, Condorcet Criterion and Extended Condorcet Criterion (ECC) [11]. We will focus on ECC (See Definition 3.1) as it offers an important property of Independence of Irrelevant alternatives.[3]

DEFINITION 3.1. *The Extended Condorcet Criterion [12] requires that if there is any partition $\{C, R\}$ of S, such that for any $d_i \in C$ and $d_j \in R$ a majority of rankers prefer d_i to d_j, then the aggregate ranking τ should prefer d_i to d_j.*

In addition to desirable theoretical properties, ECC proves to be very valuable in ranking in practice, as we will demonstrate in our experiments.

We will focus on two classical rank aggregation techniques in this paper, Borda and Kemeny, described below.

Borda Aggregation: In Borda aggregation [8] each candidate is assigned a score by each ranker; where the score for a candidate is the number of candidates below him in each ranker's preferences. The Borda aggregation is the descending order arrangement of the average Borda score for each candidate averaged across all ranker preferences. Though Borda aggregation satisfies neutrality, monotonicity, and consistency, it does not satisfy the Condorcet Criterion [13] and ECC. In fact, it has been shown that no method that assigns weights to each position and then sorts the results by applying a function to the weights associated with each candidate satisfies the Extended Condorcet Criterion [14]. This includes point-wise classifiers like logistic regression. This motivates us to consider order-based methods for rank aggregation that do satisfy ECC.

Kemeny Aggregation: A Kemeny optimal aggregation [9] is an aggregation that has the minimum number of pairwise disagreements with all rankers, i.e., a choice of τ that minimizes $K(\tau, \tau_1, ..., \tau_r) = \frac{1}{r}\sum_{i=1}^{r} k(\tau, \tau_i)$; where the function $k(\sigma, \tau)$ is the *Kendall tau* distance measured as $|\{(i,j)|i < j, \sigma(i) > \sigma(j), \text{but} \quad \tau(\text{i}) < \tau(\text{j})\}|$, where $\sigma(i)$ is used to denote the position of i in ranking σ.

Kemeny aggregation satisfies neutrality, consistency, and the Extended Condorcet Criterion. In addition, Kemeny optimal aggregation has a good maximum likelihood interpretation which we discuss in detail in [7].

IV. Supervised Kemeny Ranking

While Kemeny aggregation is optimal in the sense described above, it has two drawbacks when applied to our setting: (1) It is computationally very expensive, and (2) it does not distinguish between *good* and *bad* input rankings. Below we describe how we overcome these drawbacks.

Kemeny (and Borda) aggregation, being motivated from Social Choice Theory, strive for *fairness* and hence treat all rankers as equally important. However, fairness is not a desirable property in our setting, since we know that some individual rankers (measures) perform better than others in our target tasks. If we knew *a priori* which rankers are better, we could leverage this information to produce a better aggregate ranking. We propose Supervised Kemeny Ranking (SKR), which is based on such an approach.

The problem of computing optimal Kemeny aggregation is NP-Hard for $r \geq 4$ [14]. However, there have been some attempts to approximately solve Kemeny optimal aggregation [15]. Ailon et al. [16] presents a solution to the *feedback arc set problem* on tournaments, which can be applied to rank aggregation for a 2-approximation of Kemeny optimal aggregation. We use this approach, which we refer to as Approximate Kemeny; and we show here that it satisfies a relaxation of Kemeny optimality and the Extended Condorcet Criterion.

Approximate Kemeny can be described simply as a Quick Sort on elements based using the majority precedence relation $\succ$ as a comparator, where $d_i \succ d_j$ if the majority of input rankings has ranked d_i before d_j. Note that, the relation $\succ$ is not transitive, and hence different comparison sort algorithms can produce different rankings. In [14], Dwork et al. propose the use of Bubble Sort, which also leads to an aggregation that satisfies ECC, but comes with no approximation guarantees. This approach, which they refer to as Local Kemenization, is one of the baselines in our experiments.

By extension from Quick Sort, it can be easily shown that Approximate Kemeny runs in $O(rm \log m)$. We show below that Approximate Kemeny also produces an aggregation that satisfies the following optimality criterion. (The proofs for Lemma 4.1 and Theorem 4.1 are provided in the extended version of this work [7].)

DEFINITION 4.1. *A permutation τ is* locally Kemeny optimal *[14], if there is no full list τ^+ that can be obtained from τ by a single transposition of an adjacent pair of elements, such that, $K(\tau^+, \tau_1, ..., \tau_r) < K(\tau, \tau_1, ..., \tau_r)$.*

LEMMA 4.1. *The final aggregation τ of the Approximate Kemeny procedure produces a locally optimal Kemeny order.*

THEOREM 4.1. *Let τ be the final aggregation of the Approximate Kemeny procedure. Then τ satisfies the Extended Condorcet Criterion with respect to the input rankings $\tau_1, \tau_2, ..., \tau_r$.*

The pseudo-code for Supervised Kemeny Ranking is presented in Algo. 1. In order to accommodate supervision,

[3]We discuss in detail various other interesting interpretations of ECC in [7]

Algorithm 1 Supervised Kemeny Ranking (SKR)

Input: $\tau_i = [\tau_{i1}, ..., \tau_{im}], \forall i = 1, ..., r$, ordered arrangement of m candidates for r rankers.
$w = [w_1, ..., w_r]$ – where w_i is the weight of ranker i
$\mu = [\mu_1, ..., \mu_m]$ – initial ordered arrangement of m candidates
k – the number candidates to consider in each ranker's preference list ($k \leq m$)
Output: τ – rank aggregated arrangement of candidates in decreasing order of importance

1) Initialize majority table $M_{i,j} \Leftarrow 0, \forall i, j = 1, ..., m$
2) For each ranker $p = 1$ to r
3) For each candidate $i = 1$ to k-1
4) For each candidate $j = i + 1$ to k
5) $M_{\tau_{pi},\tau_{pj}} \Leftarrow M_{\tau_{pi},\tau_{pj}} + w_p$
6) Quick sort μ, using M_{μ_i,μ_j}. If $M_{\mu_i,\mu_j} - M_{\mu_j,\mu_i} > 0$ then μ_i is greater than μ_j. If $M_{\mu_i,\mu_j} - M_{\mu_j,\mu_i} = 0$ then μ_i is equal to μ_j. If $M_{\mu_i,\mu_j} - M_{\mu_j,\mu_i} < 0$ then μ_i is less than μ_j.
7) Return τ

we extend Approximate Kemeny aggregation to incorporate weights associated with each input ranking. The weights correspond to the relative utility of each ranker, which may depend on the task at hand. For the task of influence prediction in Twitter, we weigh each ranker based on its (normalized) AUC[4] computed on a training set of candidates, for which we know the target variable i.e., the true retweet rates. For Supervised Kemeny Ranking we incorporate weights directly in sorting the elements through Quick Sort. Instead of comparing candidates based on the preference of the simple majority of individual rankers, we use a weighted majority.

Instead of using total orderings provided by each ranker, we can also use partial orderings (for a subset of candidates). Since identifying relevant candidates at the top of the list is usually more important, we use the partial orderings corresponding to the top k candidates for each ranker. In our experiments, unless otherwise specified, we use the top-ranked 15% of candidates for each ranker.

V. Empirical Evaluation

We compared Supervised Kemeny Ranking to using individual rankings, logistic regression using all input rank scores as features, Local Kemenization [14], Borda aggregation, and a supervised version of Borda aggregation. We also compared to SVMRank [17], which is a supervised approach that tries to optimize performance on AUC.

For Supervised Borda, we incorporate performance-based (AUC) weights in Borda aggregation. This is relatively straightforward, where instead of simple averages, we take weighted averages of Borda scores. While supervised versions of Borda appear in prior work [18], to the best of our knowledge, we present the first supervised version of Kemeny aggregation.[5]

In order to verify the effectiveness of each component of Supervised Kemeny Ranking, we performed several ablation studies. In particular, we compared Supervised Kemeny Ranking to the following variations of Algo. 1:

- *Unsupervised, Total Orderings*: Using uniform weights ($w_i = 1, \forall i$), and $k = |S|$, which reduces to the unsupervised approximation to Kemeny aggregation on total orderings.
- *Supervised, Total Orderings*: $k = |S|$, i.e., Supervised Kemeny Ranking on total orderings.
- *Unsupervised, Partial Orderings*: Using uniform weights ($w_i = 1, \forall i$).
- *Supervised, Bubble Sort*: Using Bubble Sort instead of Quick Sort in Step 6. This can be viewed as a supervised version of Local Kemenization [14].

Twitter Network Study: We compared our approach, Supervised Kemeny Ranking, to the different supervised and unsupervised techniques described above on the task of predicting viral potential, as in Sec. II. As inputs to each aggregation method we use the 13 different measures listed in Table I. Each measure is used to produce a total ordering of preferences over the 9,625 candidates (twitter users), where ties are broken randomly. We compared the 10 aggregation methods (see Table II) to individual rankers, but in the interest of space we only list the best individual measure (Past Retweets) in the table. We averaged performance, measured by AUC , over 10 runs of random stratified train-test splits for different amounts of data used for training. These results are summarized in Table II .

We note that, in terms of AUC, in general, aggregation techniques perform better than using Past Retweets, which is the best individual ranker. The results also show that our version of Supervised Borda performs better than traditional Borda aggregation. However, Local Kemenization, outperforms Supervised Borda, showing the benefit of Kemeny-based aggregation versus Borda's score-based aggregation. Our approach, of Supervised Kemeny Ranking, further improves on this result, with the best performance at 3 of 4 points in terms of AUC. Logistic Regression is a little better than Supervised Kemeny Ranking at one point in terms of AUC. However, overall logistic regression is less effective than the other aggregation methods, occasionally performing worse than the best individual ranker. Supervised Kemeny Ranking, also outperforms SVMRank, consistently on all training sample sizes.[6]

Our ablations studies show that every component of Supervised Kemeny Ranking does contribute to its superior

[4]We also evaluate average precision and other weighting schemes in [7]

[5]A very preliminary version of our work appears in [19]

[6]Note that, while some absolute differences may appear small, a relative improvement of 1% is considered to be substantial in ranking domains such as web search (see Fig. 1 of [20]).

Ranking Method	Training Samples 320	480	960	1920
Supervised Kemeny Ranking	**92.97**	**92.52**	**93.28**	93.00
Past Retweets	89.47	88.86	89.73	90.20
logistic regression	46.87	70.92	87.02	**93.26**
Borda	91.02	90.78	90.95	91.14
Supervised Borda	91.50	91.09	91.22	91.62
Local Keminization	92.03	91.68	91.78	92.11
SVM Rank	87.98	89.33	92.15	92.79
Unsupervised, Total Orderings	88.49	88.29	89.91	89.35
Supervised, Total Orderings	88.89	88.36	89.92	89.51
Unsupervised, Partial Orderings	92.73	92.42	92.72	92.58
Supervised, Bubble Sort	92.23	91.88	92.03	92.27

TABLE II
RANK AGGREGATION PERFORMANCE MEASURED IN AUC(%) FOR VARIOUS TRAINING SET SIZES.

performance. In particular, we see that supervised variants of Algo. 1 perform better than unsupervised variants. Also, focusing on the top k elements from each individual ranker (*partial orderings*) is more effective that using total orderings. Finally, using the Quick Sort approximation to Kemeny aggregation makes a notable difference over using Bubble Sort.

Learning curves comparing our approach to existing baselines are presented in Fig. 1. We observe that, while logistic regression performs well with ground truth on a large number of candidates, its performance drops significantly with lower levels of supervision. In contrast, the rank aggregation methods are fairly stable, consistently beating the best individual ranking and performing better than logistic regression in the more realistic setting of moderately-sized training sets. The consistently good performance of Supervised Kemeny Ranking confirms the advantages of supervised locally optimal order-based ranking compared to score-based aggregation, such as Borda, and unsupervised methods.

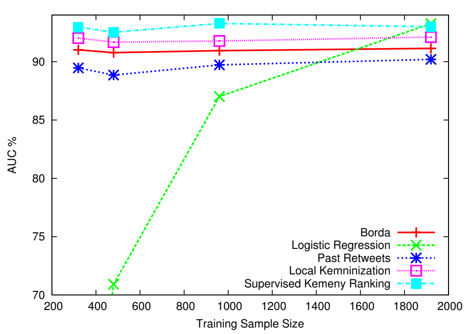

Fig. 1. AUC performance of rank aggregation techniques with increasing training data.

VI. Conclusion

Understanding influence within blog and micro-blog networks has become a crucial technical problem with increasing relevance to marketing and information retrieval. We address the problem of assessing influence by casting it in the form of a predictive task; which allows us to objectively compare different measures of influence in light of standard classification and ranking metrics. Furthermore, we propose a novel supervised rank aggregation method, which combines aspects of different influence measures to produce a composite ranking mechanism that is most effective for the desired task. We have successfully applied this approach in a study involving 40 million Twitter accounts, for the task of predicting the potential for viral out-breaks.

References

[1] E. Bakshy, B. Karrer, and L. Adamic, "Social influence and the diffusion of user-created content," in *ACM EC*, 2009.

[2] M. Goetz, J. Leskovec, M. Mcglohon, and C. Faloutsos, "Modeling Blog Dynamics," in *ICWSM*, 2009.

[3] G. Kossinets, J. Kleinberg, and D. Watts, "The structure of information pathways in a social communication network," in *KDD*, 2008.

[4] J. Moreno, *Who Shall Survive? Foundations of Sociometry, Group Psychotherapy and Sociodrama*. Nervous and Mental Disease Publishing Co., 1934.

[5] D. Knoke and R. Burt, *Applied Network Analysis*. Newbury Park, CA: Sage, 1983, ch. Prominence.

[6] S. Wasserman and K. Faust, *Social Network Analysis: Methods & Applications*. Cambridge, UK: Cambridge University Press, 1994.

[7] K. Subbian and P. Melville, "Supervised rank aggregation for predicting influence in networks," in *CoRR abs/1108.4801*, 2011.

[8] J. Borda, "Memoire sur les elections au scrutin," in *Histoire de l'Academie Royale des Sciences*, 1781.

[9] J. Kemeny, "Mathematics without numbers," in *Daedalus*, vol. 88, 1959, pp. 571–591.

[10] T. Fawcett, "An introduction to roc analysis," in *Pattern Recognition Letters*, vol. 27, 2006, pp. 861–874.

[11] K. Arrow, "Social choice and individual values." New Haven: Cowles Foundation, 2nd Edition 1963.

[12] M. Truchon, "An extension of the condorcet criterion and kemeny orders," in *J. Eco. Lit.*, 1998.

[13] H. Young and A. Levenglick, "A consistent extension of condorcet's election principle," in *J. on App. Math.*

[14] C. Dwork, R. Kumar, R. Naor, and D. Sivakumar, "Rank aggregation methods for the web," in *WWW*, 2001.

[15] F. Schalekamp and A. van Zuylen, "Rank aggregation: Together we're strong," in *ALENEX*, 2009, pp. 38–51.

[16] N. Ailon, M. Charikar, and A. Newman, "Aggregating inconsistent information: Ranking and clustering," *J. ACM*, vol. 55, no. 5, 2008.

[17] T. Joachims, "Training linear svms in linear time," in *KDD*, 2006.

[18] J. A. Aslam and M. Montague, "Models for metasearch," in *SIGIR*, 2001.

[19] P. Melville, K. Subbian, C. Perlich, R. Lawrence, and E. Meliksetian, "A predictive perspective on measures of influence in networks," in *Proceedings of the Workshop on Information in Networks*, 2010.

[20] Z. Zheng, H. Zha, T. Zhang, O. Chapelle, K. Chen, and G. Sun, "A general boosting method and its application to learning ranking functions for web search," in *NIPS*, 2007.