Supervised Rank Aggregation for Predicting Influence in Networks

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

August 18, 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 Twitter and citation networks.**

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. For users, when presented with a vast amount of content relevant to a topic of interest, ordering content by the source's authority or influence assists in information triage, thus overcoming the ever-increasing information overload.

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].

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 [7] and Kemeny optimal rank aggregation [8]. 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; and we further corroborate these results in a study of publication citation networks.

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 primary study was based on the Twitter discussion around Pepsi. What piqued our interest in Twitter and the role 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). In addition to degree and eigenvector centralities, there are other important socio-metrics based on the paths between vertices like, Closeness and Betweeness Centrality. We exclude them, as they come at the prohibitive computational cost of calculating all-pairs shortest paths in a graph ($O(V^3)$).[1]

We extracted the data[2] 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 [9]. For some applications it is more important to correctly rank relevant elements at the top of list, which we also measure by Average Precision (AP) for the top k users [10].

Measure	**Definition**	**AUC**	**AP**
Followers	Follower Graph Indegree	88.18	0.4366
Friends	Follower Graph Outdegree	76.03	0.2821
Follower Pagerank	Follower Graph Pagerank	85.77	0.4397
Distinct Past Retweets	Retweet Graph Indegree	90.17	0.7246
People Retweeted	Retweet Graph Outdegree	87.04	0.3976
Retweet Pagerank	Retweet Graph Pagerank	88.38	0.5135
Past Retweets	Wtd. Retweet Indegree	90.18	0.7406
Retweets Made	Wtd. Retweet Outdegree	86.80	0.4707
Distinct Mentions Received	Mention Graph Indegree	60.71	0.5690
People Mentioned	Mention Graph Outdegree	86.11	0.5923
Mention Pagerank	Mention Graph Pagerank	70.43	0.3631
Mentions Received	Wtd. Mention Indegree	60.53	0.2737
Mentions Made	Wtd. Mention Outdegree	84.69	0.2895

TABLE I
COMPARING RANKING MEASURES FOR IDENTIFYING VIRAL POTENTIAL, IN TERMS OF AUC(%) AND AVERAGE PRECISION@100.

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 and Average Precision. The number of followers and the number of people mentioned also produce reasonably good rankings in terms of AUC and Average Precision respectively.[3] 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. This underscores the fact that each aspect (network of followers, diffusion of past retweets, and interactions through replies and mentions) contributes to ones potential to reach a large audience. By focusing on selecting a single centrality measure to capture influence we would miss out on the opportunity to more precisely detect potential viral users.

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

[1] In related work, we have been working on a scalable algorithm for computing Betweeness Centrality, exploiting hierarchical parallelization.

[2] We will make all our data publicly available.

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

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* (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 primarily focus on ECC, defined below:

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.*

The ECC property is highly preferred in our domains, as it eliminates the possibility of inferior candidates being introduced strategically in order to manipulate the choice between superior candidates. In other words, it offers the property of Independence of Irrelevant Alternatives. Additionally, ECC is a relaxed form of Kemeny optimal aggregation (defined below), where the partition C and R are arranged in the "true" order, but not necessarily the elements within partitions C and R. In addition to the 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, describe below.

Borda Aggregation: In Borda aggregation [7] 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 [8] 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 $|\{(i, j)|i < j, \sigma(i) > \sigma(j), \text{but } \ \tau(\text{i}) < \tau(\text{j})\}|$, where $\sigma(i)$ is used to denote the position of i in ranking σ.

Kemeny aggregation satisfies neutrality, consistency, and the Extended Condorcet Criterion. Kemeny optimal aggregation also has a good maximum likelihood interpretation. Suppose there is an underlying "correct" ordering σ of S, and each order $\tau_1, ..., \tau_r$ is obtained from σ by swapping pairs of elements with some probability less than $1/2$. That is, the τ's are "noisy" versions of σ. A Kemeny optimal aggregation of $\tau_1, ..., \tau_r$ is one (not necessarily unique) that is maximally likely to have produced the τ's.

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. In fact, given the ordering of a (small) set of candidates, we can estimate the performance of individual rankers and use this to produce a better ranking on a new set of candidates. 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.

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.*

Proof: Every element in the final order is compared at least once with its neighboring elements in the quick sort procedure. As such, d_i is placed immediately to the left of d_j only if d_i is preferred to d_j by a majority of input rankings. So, swapping any such adjacent elements can only increase the number of input rankings that disagree with this ordering, thus increasing the total Kendall tau distance. Hence Approximate Kemeny is locally Kemeny optimal. ■

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 $\tau_1, \tau_2, ..., \tau_r$.*

Proof: The proof follows directly from Lemma 6 of [14]. If the claim is false then there exist rankers $\tau_1, \tau_2, ..., \tau_r$, an Approximate Kemeny aggregation τ, and a partition (T, U) of the elements where for all $a \in T$ and $b \in U$ the majority among $\tau_1, \tau_2, ..., \tau_r$ prefers a over b, but there is a $c \in T$ and a $d \in U$ such that $d > c$ in τ. Let (d, c) be a closest such pair in τ. Consider the immediate successor of $d \in \tau$, and call it e. If $e = c$ then c is adjacent to $d \in \tau$ and transposing this adjacent pair of elements produces a τ^+ such that $K(\tau^+, \tau_1, ..., \tau_r) < K(\tau, \tau_1, ..., \tau_r)$, contradicting Lemma 4.1 that τ is a locally Kemeny optimal aggregation of the $\tau_1, \tau_2, ..., \tau_r$. If e does not equal c, then either e is in T, in which case the pair (d, e) is a closer pair in τ than (d, c) and also violates the Extended Condorcet Criterion, or e is in U, in which case (e, c) is a closer pair than (d, c) that violates the Extended Condorcet Criterion. Both cases contradict the choice of (d, c). ■

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

The pseudo-code for Supervised Kemeny Ranking is presented in Algo. 1. In order to accommodate supervision, 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 computed on a training set of candidates, for which we know the target variable i.e., the true retweet rates. When evaluating on Average Precision, we use weights based on Average Precision instead. 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. This can be achieved simply by using weighted votes during the creation of the majority table M – which represents the sum of weights of the rankers who prefer the row candidate to the column candidate for each pairwise comparison.

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/AP) weights in Borda aggregation. This is relatively straightforward, where instead of simple averages, we take weighted averages of Borda scores. A similar approach to supervised Borda was used in [18], where weights were based on average precision of each ranker for a meta-search task. While, supervised versions of Borda appear in prior work, to our knowledge, we present the first supervised version of Kemeny aggregation.[4]

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].

A. 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 and Average Precision@100, over 10 runs of random stratified train-test splits for different amounts of data used for training. These results are summarized in Tables II and III.

We note that, in terms of AUC, in general, aggregation techniques perform better than using Past Retweets, which is the best individual ranker. However, apart from Supervised Kemeny Ranking, this is not always the case for Average Precision. So one must use rank aggregation with caution, depending on the desired performance metric. 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 all points in terms of Average Precision, and 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, in both AUC and AP.[5]

Our ablations studies show that every component of Supervised Kemeny Ranking does contribute to its superior 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. As mentioned earlier, the Bubble Sort variation, as used by Dwork et al. [14] comes with no approximation guarantees, which makes a perceptible difference in practice. In addition to using AUC-based weights for Supervised Kemeny Ranking, we also experimented with alternative weighting schemes in Algo. 1, such as, $(AUC - 0.5)$ and $(log(AUC/(1 - AUC)))$. However, in experiments (not presented) the simple AUC based weights outperformed other weighting schemes by a margin of $2 - 5\%$.

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.

While Fig. 1 shows the performance in terms of area under the ROC curve for different sample sizes, in Fig. 2 we present the ROC curves for a single point (1,920 training samples). We contrast Supervised Kemeny Ranking, with the methods most commonly used in practice, namely, number of followers and follower PageRank (e.g., as done by Twitaholic.com and Tunrank.com). Note that, all other

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

[5]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]).

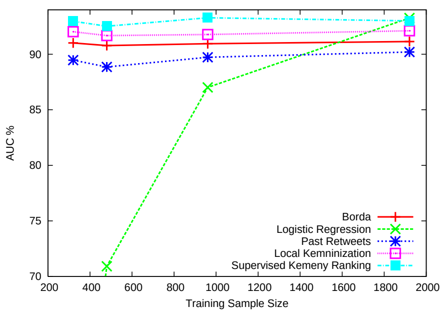

Fig. 1. AUC performance of rank aggregation techniques with increasing training data.

Ranking Method	Training Samples 320	480	960	1920
Supervised Kemeny Ranking	**0.7242**	**0.6837**	**0.6991**	**0.6783**
Past Retweets	0.7210	0.6610	0.6766	0.6668
logistic regression	0.3255	0.4862	0.6662	0.6219
Borda	0.2600	0.2600	0.2333	0.2133
Supervised Borda	0.3000	0.2733	0.2366	0.2334
Local Keminization	0.5240	0.4938	0.4768	0.4891
SVM Rank	0.1732	0.3180	0.3990	0.3996
Unsupervised, Total Orderings	0.6982	0.5998	0.6706	0.6357
Supervised, Total Orderings	0.6994	0.6024	0.6826	0.6521
Unsupervised, Partial Orderings	0.7018	0.6622	0.6745	0.6619
Supervised, Bubble Sort	0.5273	0.4963	0.4772	0.4930

TABLE III

RANK AGGREGATION PERFORMANCE MEASURED IN AVERAGE PRECISION@100 FOR VARIOUS TRAINING SET SIZES.

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.

baselines in this paper are devised by us, and are much better than these approaches. We observe that Supervised Kemeny Ranking performs 5 to 8% better in terms of AUC and 54 to 55% better in terms of AP compared to current practice.

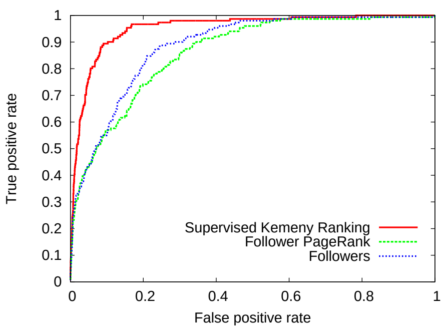

Fig. 2. ROC curves comparing Supervised Kemeny Ranking to popular measures in practice.

B. Citation Network Study: In addition to Twitter data, we also performed a case study on publication citation networks. For this we used a collection of papers with their citations that was used in the KDD Cup contest held in 2003.[6] This data consists of 1,716 papers in the field of High Energy Physics Theory (hep-th), published on arXiv.org during a 6 month period. The data set also contains the number of times each paper was downloaded during the 60 day period after it was published on arXiv.org. This download information gives us an extrinsic proxy for the influence of a paper. As such, we define the task of predicting highly influential papers, as measured by downloads, based on the citation data of the papers. If a paper received 600 or more downloads, we consider it as a high-influence paper (77 papers); else we consider it to have little or no influence.

First, we constructed a citation graph based on all publications in hep-th, which was also provided as part of KDD Cup 2003. In this citation graph, each node represents a paper and each edge represents a citation. As of May 1, 2003, there were 29,014 papers and 342,427 citations in total in the hep-th data. Next, for each of the 1,716 papers with download information, we used this citation graph to compute 5 influence measures - Indegree, Outdegree, Pagerank, Hub and Authority score [21].

We ran experiments as before, using 20% of the data (343 papers) for training the supervised methods, and setting k to 1,200 in Algo. 1. The results in terms of AUC and Average Precision for each method are presented in Table IV. As expected, the number of papers citing a given paper (in-degree) is a good indicator of how often the paper will be downloaded. Furthermore, having more citations from highly-cited papers, as captured by PageRank is a better indicator of influence in this data. Note that, this was not the case in predicting viral potential in Twitter. The number of papers a paper is citing (out-degree) and Hub-score have some, though weaker, ability to predict influence. This is probably because some survey papers do become influential if they refer to many good papers in that area.

In this study we find that not all aggregation techniques are better than using individual rankers. In particular, high in-degree is very correlated with high download rates, as reflected by Average Precision. So depending on the data and the evaluation metrics, one should always consider

[6] http://www.cs.cornell.edu/projects/kddcup/

Measure	AUC %	AP
PageRank	81.09	0.4470
Indegree	80.42	**0.5376**
Authority	80.39	0.5324
Outdegree	64.33	0.2820
Hub	61.07	0.2867
Supervised Kemeny Ranking	**81.70**	0.4950
logistic regression	76.02	0.5330
Borda	77.47	0.2363
Supervised Borda	78.27	0.2787
Local Keminization	76.62	0.3668
SVMRank	77.59	0.4625
Unsupervised, Total Orderings	80.12	0.3518
Supervised, Total Orderings	80.30	0.4902
Unsupervised, Partial Orderings	80.23	0.4928
Supervised, Bubble Sort	79.17	0.4798

TABLE IV
COMPARING RANKING METHODS FOR IDENTIFYING INFLUENTIAL PAPERS, BASED ON AUC AND AVERAGE PRECISION@60.

using the best individual ranker along with alternative aggregation methods. Nevertheless, in terms of AUC, Supervised Kemeny Ranking still produces the best ranking, outperforming individual rankers and other aggregation techniques. The results on the ablation studies are similar to before, further corroborating the contribution of each component of the Supervised Kemeny Ranking algorithm.

VI. Related work

An associated growing area of research attempts to explain content and link structures in social media, together with their temporal evolution, based on tensor factorizations and higher order extensions of techniques such as Singular Value Decomposition (SVD) [22], [23]. Recently, Weng et al. [24] propose TwitterRank, a variant of PageRank that also takes topical similarity between users into account.

Another interesting approach to quantitatively evaluating the ranking of blogs is through the task of cascade detection - selecting a set of blogs to read which link to most of the stories that propagate over the blogosphere. Current solutions [25], [26] to this task do not attempt to address the task of assigning an influence score to individual bloggers, since they are focused on optimal set selection. However, there is a lot of potential for using such approaches to identify influencers.

In related work on rank aggregation, Liu et al. [27] present an alternative supervised approach for the task of web-search – where they build on a Markov Chain (MC) based approach to rank aggregation. However, it has been shown that Local Kemenization improves on MC-based approaches [14], which in turn, we show is outperformed by Supervised Kemeny Ranking.

In concurrent work on the analysis of Twitter, Cha et al. [28] also conclude that number of followers alone reveals little about a user's influence. We go further in our work, by comparing many more socio-metrics on different tasks, and providing approaches to improve influence prediction through rank aggregation. In recent work, Suh et al. [29] analyze factors that correlate with retweeting. While they consider in- and out-degrees of the follower graph, they do not look at other graphs, such as the retweet graph, or other socio-metrics, such as PageRank. Furthermore, since their study only uses randomly sampled tweets, they are limited to a very small subset of retweets. In contrast, we collect all retweets for all users in our study.

In addition to SVMRank, there have been several recent advances in learning to rank [30], [31], driven largely by the application to web search. All of these approaches produce a ranked list as an output. In their seminal work, Dwork et al. [14], showed how rank aggregation can be used to improve on meta-search, by combining individual search rankings. Since, we demonstrate that Supervised Kemeny Ranking performs better than their Local Kemenization approach, we are hopeful that it can be used to aggregate the rankings from different learning to rank methods, to improve results on web search and other applications.

In recent work, Ghosh and Lerman [32] evaluate various influence models based on geodesic-path based distance measures and topological ranking measures. They propose a Normalized α-centrality algorithm and evaluate its effectiveness on measuring influential users in Digg.com. Their work aims to find the best individual socio-metric and does not intend to improve the predictive accuracy by combining various influence models. However, as we have shown in this paper, often individual socio-metrics fail to capture all critical factors that are relevant for predicting influence in networks. Presumably, one could use the Normalized α-centrality algorithm as another input ranker to Supervised Kemeny Ranking, to further improve predictive performance.

The work by Agarwal et al. [33] does a empirical study on identifying influential people in blog networks. They propose 4 main features that produce influence in the bloggers network, based on recognition, activity, novelty, and eloquence. They weigh these four features to produce a combined score for each blogger. In [34], Sayyadi and Getoor predict the popularity of a paper using its expected future citations. They propose $FutureRank$, which combines the PageRank score of a paper in the citation network, the authority score in the authorship network, and the recency of the publication. Both [33] and [34] propose a score-based model, where they combine the scores from a set of features defined on the underlying network data. Note that, neither of the methods are supervised and they require further enhancements to accommodate such supervision. In addition, their methods are score-based aggregations, and not order-based. Both Dwork et al. [14] and this paper shows clearly the inefficiency of weighted combination of score-based algorithms compared to order-based.

VII. Conclusion and Future Work

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 applied this approach to a case study involving 40 million twitter accounts, and have examined the task of predicting the potential for viral out-breaks. We further corroborated these results on the task of identifying influential papers based on citation networks. Empirical results show that our proposed approach, Supervised Kemeny Ranking, performs better than several existing rank aggregation techniques, as well as other supervised learning benchmarks.

The problem of choosing the optimal Kemeny order can be formulated as a mixed-integer programming problem as discussed in [35]. However, the problem of finding the optimal weights for Supervised Kemeny Ranking is much more difficult, as it involves a quadratic objective function, with two sets of variables; one for selecting the optimal weights and one for the optimal order. An efficient algorithm to solve this optimization could significantly improve results, and is a promising direction for future work.

Acknowledgements

The authors would like to thank Estepan Meliksetian at IBM Research for his help in the necessary data extraction for the experiments. The authors also are grateful to Claudia Perlich at Media6Degrees and Richard Lawrence at IBM Research for their suggestions and comments on this work.

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

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

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

[10] R. Baeza-Yates and B. Ribeiro-Neto, "Modern information retrieval." Addison Wesley Co, 1999.

[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 *SIAM J. on App. Math*, vol. 35(2), 1978.

[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.

[21] J. Kleinberg, "Authoritative sources in a hyperlinked environment," in *J. ACM*, 1999.

[22] T. Kolda and B. Bader, "The TOPHITS model for higher-order web link analysis," in *SDM Workshop on Link Analysis, Counterterrorism and Security*, 2006.

[23] Y. Chi, S. Zhu, X. Song, J. Tatemura, and B. Tseng, "Structural and temporal analysis of the blogosphere through community factorization," in *KDD*, 2007.

[24] J. Weng, E.-P. Lim, J. Jiang, and Q. He, "TwitterRank: Finding topic-sensitive influential Twitterers," in *WSDM*, 2010.

[25] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. Vanbriesen, and N. Glance, "A Cost-effective outbreak detection in networks," *KDD*, 2007.

[26] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *KDD*, 2003.

[27] Y.-T. Liu, T.-Y. Liu, T. Qin, and H. Li, "Supervised rank aggregation," in *WWW*, 2007.

[28] M. Cha, H. Haddadi, F. Benevenuto, and K. Gummadi, "Measuring user influence in twitter: The million follower fallacy," in *ICWSM*, 2010.

[29] B. Suh, L. Hong, P. Pirolli, and E. H. Chi, "Want to be retweeted? large scale analytics on factors impacting retweet in twitter network," in *SocialCom*, 2010.

[30] Y. Freund, R. Iyer, R. Schapire, and Y. Singer, "An efficient boosting algorithm for combining preferences," in *JMLR*, 2003.

[31] C. J. C. Burges, K. M. Svore, P. N. Bennett, A. Pastusiak, and Q. Wu, "Learning to rank using an ensemble of lambda-gradient models," in *JMLR*, 2011, pp. 253–35.

[32] R. Ghosh and K. Lerman, "Predicting influential users in online social networks," in *SNA-KDD*, 2010.

[33] N. Agarwal, H. Liu, L. Tang, and P. S. Yu, “Identifying the influential bloggers in a community,” in *WSDM*, 2008.
[34] H. Sayyadi and L. Getoor, “Futurerank: Ranking scientific articles by predicting their future pagerank,” in *SDM*, 2009.
[35] V. Conitzer, A. J. Davenport, and J. Kalagnanam, “Improved bounds for computing kemeny rankings,” in *AAAI*, 2006.