Learning to Measure Influence in a Scientific Social Network

Shane Bergsma, Regan L. Mandryk, and Gordon McCalla

University of Saskatchewan, Saskatoon, Saskatchewan, Canada S7N 5C9
shane.bergsma@usask.ca, regan@cs.usask.ca, mccalla@cs.usask.ca

Abstract. In research, *influence* is often synonymous with *importance*; the researcher that is judged to be influential is often chosen for the grants, distinctions and promotions that serve as fuel for research programs. The influence of a researcher is often measured by how often he or she is cited, yet as a measure of influence, we show that citation frequency is only weakly correlated with influence ratings collected from peers. In this paper, we use machine learning to enable a new system that provides a better measure of researcher influence. This system predicts the influence of one researcher on another via a range of novel social, linguistic, psychological, and bibliometric features. To collect data for training and testing this approach, we conducted a survey of 74 researchers in the field of computational linguistics, and collected thousands of influence ratings. Our results on this data show that our approach significantly outperforms measures based on citations alone, improving prediction accuracy by 56%. We also perform a detailed analysis of the key features in our model, and make some important observations about the scientific and non-scientific factors that most predict researcher influence.

1 Introduction

This paper is concerned with understanding and quantifying researcher influence. We define influence as the capacity of a researcher to have an effect on another researcher's opinions, ideas, experimental approach, or choice of research topics.

Studying influence is important. Knowing about influential individuals can helps us understand how behaviours spread [1], and inform technology such as paper recommendation systems [2]. Also, since measures of influence are used to evaluate research, better measures of influence could help us make better strategic decisions, improving processes ranging from hiring, funding, promotion, and award-giving, to the assessment of organizations and funding programs.

Both quantitative and qualitative factors are commonly used to assess researcher influence. Quantitative measures based on citation counts, such as the h-index [3], provide a convenient and objective indicator of influence. But are these measures reliable? While citations "are correlated with other assessments of scientists' impact or influence, such as awards, honors, and Nobel laureateships," [4] citation-based measures are less useful in fields with "less heavy citation traffic" and are susceptible to manipulation [5]. Alternatively, qualitative assessments (e.g. recommendation letters) can be more reliable, but are time-consuming and subjective. Our aim is to develop an improved measure of researcher influence that can account for both quantitative and qualitative information, based on both scientific and non-scientific (social, psychological) factors.

Our main contribution is a learned model that predicts, for a pair of researchers, the influence that one researcher has on another. Formally, this model treats influence as function $\mathit{Infl}(x, y) \rightarrow [0, \infty)$: the influence of researcher x on researcher y, expressed as a non-negative real number. It is possible to aggregate these predictions in order to develop measures of total influence, but in this work we focus on pairwise, directed influence as a first step toward global models.

We train and test our approach within the scientific domain of computational linguistics. We first extracted a dataset of researchers from the information available through the ACL Anthology Network (Sec. 2) (derived from publications of the Association for Computational Linguistics). We then used the ACL data to build individual *scientific social networks* for the ACL authors; we elaborate on this concept in Sec. 3. Next, we solicited researchers to perform an online survey where they can select those members of their scientific networks that most influence them; respondents provided high-quality ratings (Sec. 4). We then used these ratings to train the influence model using techniques from supervised machine learning (Sec. 5). The heart of our model is a creative set of features (Sec. 6) that enables dramatic improvements in the prediction of influence, substantially improving accuracy over baselines based on citation counts alone (Sec. 8). In our results, we also provide an instructive analysis of what factors most affect influence. Overall, this paper provides important new tools, ideas, and directions for work at the intersection of machine learning and bibliometrics.

2 The ACL Anthology Network

To compute our function $\mathit{Infl}(x, y)$, we require information about researchers x and y and their relationship to one another. In this paper, we extract this information for researchers in the field of computational linguistics. We selected this field partly because of the availability of the ACL Anthology Network or AAN [6]. The AAN comprises a majority of worldwide papers in computational linguistics since the 1960s. The 2012 release of the AAN provides 20K full-text papers and 95K paper citations. Note, however, that these citations only include citations to other papers in the AAN. Each author is associated with a unique ID, from which we can trace citations and co-authorships within the AAN.

3 Building Scientific Social Networks

While the domain of $\mathit{Infl}(x, y)$ could comprise all pairs of researchers, if we define $\mathit{Infl}(x, y)$ in terms of how often y cites x then, by definition, y is only influenced by the set of researchers $\{x\}$ whom y has cited. Rather than restrict our analysis to the set of citing researchers, we expand the set of researchers that might influence y to a group we call y's *scientific social network*.

A researcher x is part of y's scientific social network if and only if:

1. y cites a paper authored by x (y's citation network)
2. x is a co-author of y (y's degree-1 network)
3. x and y share a co-author (y's degree-2 network)
4. x is among y's most-similar authors by paper content (y's topic network)

We calculate the author similarity using a vector-space approach [7]. We first build *tf-idf* vectors for each author from the combined text of all their papers. We took steps to exclude names, affiliations, references, stopwords, and infrequent terms. We compute the similarity between two authors by computing the cosine similarity between their *tf-idf* vectors. y's topic network comprises y's 50 most similar authors, provided the cosine values exceed a minimum threshold (.06).

The average network size is 90.3 people, which includes overlapping contributions of, on average, 6 people via the degree-1 network, 52 people via degree-2, 32 via citations and 47 via topic similarity.

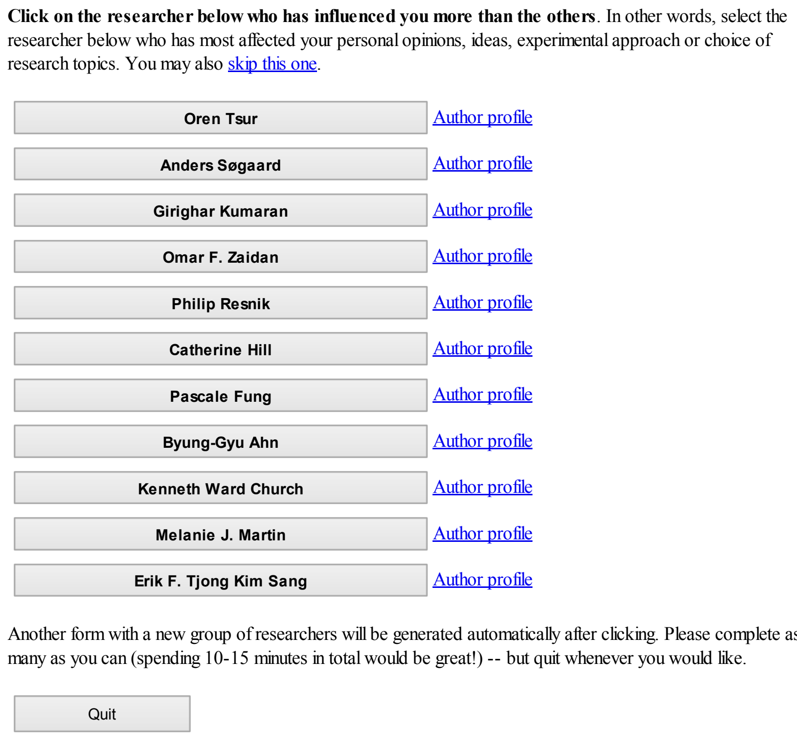
4 A Researcher-Specific Survey of Scientific Influence

To train and test our influence model, we require a gold standard set of influence ratings. The gold standard in bibliometrics has long been direct peer assessment [8, 5]. We used a survey to obtain peer assessments of the researchers within our data. We took steps to safeguard the confidentiality of the responses and our methodology received ethical approval from our institution.

Respondents (users) clicked on a hyperlink that took them to an online form (Fig. 1). The screens were customized for each user, with 11 clickable names randomly drawn from that user's scientific social network. Eleven names were chosen as a reasonably large number that could still fit on one screen without the need to scroll. Users were instructed to "click on the researcher... who has most affected your personal opinions, ideas, experimental approach or choice of research topics." The task of selecting a single name, as opposed to ranking or ordering the names, was chosen both to make the task easy for the users, and because such choices can still imply a global influence ranking (Sec. 5). When a name is clicked, the selected name and the 10 unselected ones are recorded, and a new set of names is randomly chosen. While users have the option of skipping screens, one user reported that making one choice "sent me into a deep introspective philosophical debate that I will likely be weeks recovering from."

114 researchers were contacted and 74 responded, resulting in a response rate of 65.5%. Users completed 40 screens on average (2957 in total). 86% of users were male, while 76% were from North America. The users ranged from new researchers to those who have been publishing for several decades, and included 10 students, 9 postdocs, 34 professors/government scientists, and 21 in industry.

Survey reliability One measure of response quality is how often users contradict themselves, i.e., how often they rate researcher A higher than researcher B on one screen and then B higher than A on another. Because the average user completed 40 screens and had 90 people in their network, there were many chances to do



Click on the researcher below who has influenced you more than the others. In other words, select the researcher below who has most affected your personal opinions, ideas, experimental approach or choice of research topics. You may also skip this one.

Oren Tsur	Author profile
Anders Søgaard	Author profile
Girighar Kumaran	Author profile
Omar F. Zaidan	Author profile
Philip Resnik	Author profile
Catherine Hill	Author profile
Pascale Fung	Author profile
Byung-Gyu Ahn	Author profile
Kenneth Ward Church	Author profile
Melanie J. Martin	Author profile
Erik F. Tjong Kim Sang	Author profile

Another form with a new group of researchers will be generated automatically after clicking. Please complete as many as you can (spending 10-15 minutes in total would be great!) -- but quit whenever you would like.

Quit

Fig. 1. Screenshot of online survey form. Clicking on the 'Author profile' links will open the selected author's profile on the AAN website, providing information about the author's publications, affiliations, collaborators, etc.

this. However, we found that of all the thousands of ratings, only one user did this three times, while five other users did this only once.

Another indicator of response quality is the pattern of how frequently each name in the form is clicked, by position on the screen. Because the names are ordered randomly, the null hypothesis is that the clicks would be evenly distributed, with roughly the same number of clicks on names presented in the first position as those in the second, third, etc. The alternative hypothesis is that users click names in the first half significantly more than those in the bottom, since they read from top to bottom and may not bother to read all the names. While the clicks were fairly evenly distributed (Fig. 2), a one-tailed binomial test shows that users click on names in the top half significantly more (p=0.024). But since the names are randomly drawn, this behaviour results in random noise. Nevertheless, we are considering ways to account for and model this behaviour as part of our training algorithm. We also investigated if users were more likely to click on a name close to where they clicked on the previous screen. Again, any effect here would be noise rather than systematic bias. However, the observed next-click distribution closely tracked the expected distribution (Fig. 3).

Altogether, these analyses show the users did an excellent job on the surveys. However, there is some noise in the ratings, and while our training algorithm

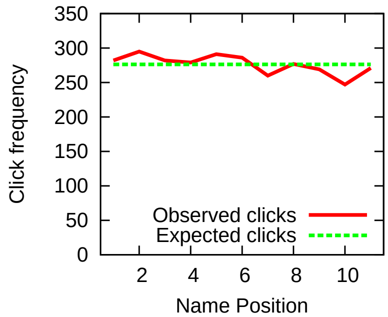


Fig. 2. Distribution of responses by position of researcher name on screen.

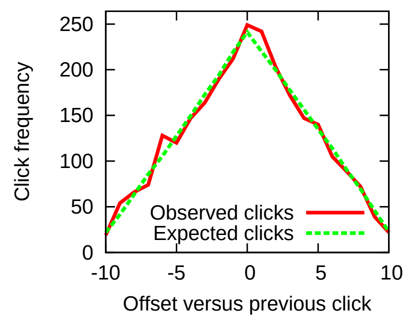


Fig. 3. Distribution of responses by distance from previous click on screen.

(described in the following section) is robust to such effects, there is clearly an upper bound to the accuracy achievable on our test data.

5 An SVM Ranking Approach to Influence Prediction

While peer assessment ratings have typically been used to *evaluate* citation-based metrics [8, 5], here we use our peer assessments to *train* a model of influence.

Formally, this model expresses the influence of researcher x on researcher y as a function $\mathit{Infl}(x, y) \rightarrow [0, \infty)$. We model $\mathit{Infl}(x, y)$ using a linear function, $\mathit{Infl}(x, y) = \mathbf{w} \cdot \mathbf{f}(x, y)$, where $\mathbf{w} = (w_1, w_2 \dots w_N)$ are the parameters of the model: the N weights chosen by the machine learning algorithm during training. The function $\mathbf{f}(x, y) = (f_1(x, y), f_2(x, y) \dots f_N(x, y))$ is an N-dimensional feature function that describes the relationship between researcher x and y. The particular binary and real-valued features that we use are described in Sec. 6.

Our survey was designed to provide us with *relative* influence ratings for learning the weights of the $\mathit{Infl}(x, y)$ model. When a user y selects the kth researcher, x_k, from the set of 11 researchers, $\{x_i\}$, the user is indicating that this kth researcher was more influential than the others presented in that set. This implies relative preference rankings, $\mathit{Infl}(x_k, y) > \mathit{Infl}(x_i, y), \forall i \neq k$. Ten such preferences are implied for each screen, and with roughly 3000 screens completed by our survey respondents, we obtain close to 30,000 implied preferences.

Our training algorithm has the basic goal of finding the set of weights, w, that can satisfy as many of these preferences as possible. Joachims [9] describes the similar problem of optimizing a ranking function for search engines, via the preferences implied by click-through data. We follow Joachims in using a support vector machine (SVM) solution. The SVM objective is to find the set of weights that results in the maximum separation between $\mathit{Infl}(x_k, y)$ and $\mathit{Infl}(x_i, y)$ (with slack variables and regularization). With this objective, Joachims shows how the constraints can be reduced to the form $\mathbf{w} \cdot (\mathbf{f}(x_k, y) - \mathbf{f}(x_i, y)) > 1$, which is essentially a classification SVM on pairwise difference vectors. We can thus solve the optimization problem efficiently using standard SVM software (Sec. 7).

6 Features for Predicting Scientific Influence

The job of the features is to provide information to explain why a user, y, deems one researcher, x, to be more influential than another. One of our key hypotheses is that a variety of non-scientific factors can help explain scientific influence. For motivating these factors, we draw on the ideas of Cialdini [10], a popular work on the principles of human influence. Of course, these principles may operate differently here, and the effectiveness of our features may be the result of other not-yet-understood processes. However, our aim is to both improve our system, and gain insights about what sorts of information are most useful.

The particular features were selected based on development experiments.

Basic Network These features encode the information that we originally used to create the scientific social networks (Sec. 3). Features indicate the (a) no. of papers where y cites x, (b) no. of times y and x are co-authors, (c) no. of times x and y share a common co-author, and (d) cosine similarity of x and y's papers.

ReverseCite Cialdini [10]'s first principle of influence is *reciprocity*: we are influenced by people who do something for *us.* As a potential instance of this in science, we include a feature for the number of times x cites y, i.e., the opposite of what we typically measure. Note we are not assessing here whether pure citation reciprocity is at play; we are assessing whether the fact someone cites you makes you more likely to rate them as genuinely *influencing* you.

Authority We are also influenced by people in positions of authority. This is best exemplified by Milgram's famous study where participants performed acts against their personal conscience when instructed to do so by an authority figure [11]. To encode the authority of x, we include features for x's total number in the data of (a) citations and (b) co-authors; these measures are known to correlate with authoritative positions such as program committee membership [12]).

Similarity It has also been established that people are most influenced by *similar* people [10]. We measure similarity in three ways:

(1) Affiliation: Since our data includes affiliation information for each researcher, we include features for whether x and y have the same affiliation and whether the final token in their affiliations match (indicating they are in the same country, state, or province). We also labeled each affiliation with whether it represents an academic, government, or industrial organization type; we then include a feature for whether x and y have the same organization type.

(2) Name: Two researchers may also share a similar ethnicity. We capture this in our system by including a feature for the semantic similarity of x and y's first names. We compute this similarity using recent data from Bergsma et al. [13], who built clusters of names based on communication patterns on Twitter. The name clusters were shown to be useful for predicting a user's ethnicity, race, and native language. Our feature is the cosine similarity between x and y's cluster-membership-vector (the vector of similarities to each cluster centroid).

(3) Gender: People of one gender may be more influenced by people of the same gender. For each x and y, we compute the most likely gender of their names via the gender data of Bergsma and Lin [14], and include a feature for which of the four x/y gender configurations (*M/M*, *M/F*, *F/M*, *F/F*) is under consideration.

SocialProof A final, widely-exploited form of influence is commonly called *social proof*: we often adopt the behaviour of our peers as a default habit [10]. In our case, if y's academic peers are influenced by researcher x, then y might be too. An example of this kind of default thinking in research is the hundreds of citations to a paper by Salton, where the paper does not actually exist [15]; many researchers were simply copying the citations of their peers without looking up the original paper. To capture this kind of influence, we add a feature to our classifier for the number of times that y's co-authors cite x.

TimeInactive In development experiments, we discovered one other feature that was a good *negative* predictor of influence: time since the researcher's last publication, in years. After researchers become inactive, their influence decreases.

All When we use all the above features together, we refer to it as the *All* system. When using *All*, we also incorporate a new technique for improving the feature types that rely on citations. Rather than just counting the number of papers where y cites x, we count how often the *surname* of x is mentioned in the papers of y. While more noisy than citations, there are two primary reasons for using surname counts: (1) Zhu et al. [16] recently showed that citations mentioned more than once in a paper are more likely to be rated by the authors as being "influential", and surname counts indirectly capture citation counts, and (2) since the AAN data is a closed set (Sec. 2), we can indirectly capture citations to papers by the researchers that occur in non-AAN venues (e.g., papers by researchers who regularly publish in both machine learning and NLP).

7 Experiments

Our experiments address two main questions: (1) How much can our approach improve over the standard way of measuring influence via citations? and (2) How important are each of the scientific and non-scientific factors in the prediction of influence? For the latter question, we test the value of each feature type by seeing how much accuracy drops when features are removed from the system.

For evaluation, we divide the gold ratings by *user*: we take the ratings from 54 users for training and from 20 other users for final test data. We train our models using SVM-Rank [17]. Since we do not have a surfeit of data, we performed development experiments and tuned the SVM regularization parameter by performing 54 rounds of leave-one-out evaluation on the training data. For the final results, we train on all 54 users and test on the held-out test data.

Our unit of evaluation is a screen completed by a user, y: one selected researcher x and 10 alternatives. Our evaluation metrics are: (1) Top-1 Acc.: the proportion of screens for which we perfectly predict the user-selected researcher

Type	System	Top-1 Acc. (%)	MRR (%)	Signif.
Baseline	Random	9.1	27.5	*p<0.001*
	Most-cited	29.6	51.4	*p<0.001*
RankSVM	Features: *ResearcherID*	27.6	47.2	*p<0.001*
RankSVM	Features: *Basic Network*	37.4	56.0	*p<0.001*
	Features: *All*	**46.1**	**63.4**	-
	Features: *All - ReverseCite*	45.2	62.9	*Not signif.*
	Features: *All - Authority*	45.0	62.9	*p<0.2*
	Features: *All - Similarity*	44.4	62.2	*p<0.1*
	Features: *All - SocialProof*	45.2	62.8	*Not signif.*
	Features: *All - TimeInactive*	44.2	62.4	*p<0.05*

Table 1. Main results: Baselines, *All* system, and with features removed. *Signif.* gives p-value (McNemar's) for whether system's Top-1 Acc. is significantly worse than *All*.

Fernando Pereira, Dekang Lin, Chris Manning, Aravind Joshi, Ken Church, Mark Steedman, Hal Daume, Graeme Hirst, Jason Eisner, Chris Quirk, Ray Mooney, Philip Resnik

Table 2. Most-highly-ranked researchers: Ranked in descending order of the learned weights of the *ResearcherID* model (trained on ratings of 54 survey respondents).

from the 11 choices, and (2) Mean reciprocal rank: (MRR) of the correct researcher, a standard IR metric for evaluating rankings (closer to 1 is better).

We compare to three baseline systems:

1. Random: Select a researcher, x, randomly from the 11 choices
2. Most-cited: Select the researcher, x, that y cites the most
3. *ResearcherID*: A learned classifier with one feature: the ID of researcher x

ResearcherID ignores y and thus the relationship between x and y, and thus lets us test whether a global ranking of researchers can generate accurate predictions.

8 Results

Table 1 presents the main results of our study. *All* is substantially and significantly better than the baselines. In fact, the *Basic Network* system is also significantly better than *Most-cited* ($p<0.01$, McNemar's test), showing the power of combining even simple features via machine learning. The feature ablation shows that *Similarity* and *TimeInactive* are the most important new features, but Top-1 Acc. drops by a percentage or two when any of the feature classes are removed. The overall strong performance of *All* is thus due to the collective contributions of all feature types, even though these contributions may not individually be statistically significant. So, even when we consider how many papers x and y have co-authored, and how often y has cited x, etc., there is still valuable information in more subtle clues, such as whether x and y are in the same country, whether x has cited y, whether y's colleagues have cited x, etc.

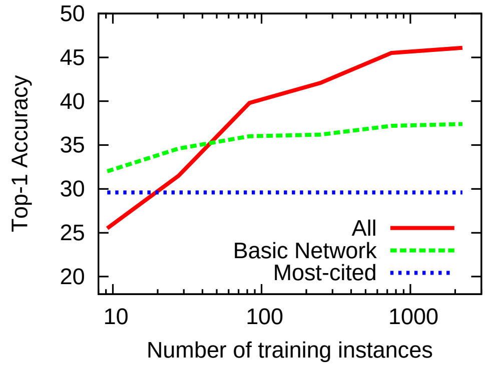


Fig. 4. Learning curve of different systems (x-axis is log-scale): *All* is better than the alternatives after 50 training screens, and continues to improve across all training sizes.

Degree-1: Number of co-authored papers	0.82
Similarity of first names	0.75
Affiliation perfect match	0.56
Topic-Sim: Cosine similarity of publication term vectors	0.49
Cites: Number of times *you* mention the *researcher's* last name	0.42
Degree-2: Number of shared co-authors	0.26
ReverseCite: Number of times the *researcher* mentions *your* last name	0.19
Affiliation match of country/state	0.18
Affiliation match of type: academic, government, or industrial	0.18
Authority: Number of overall citations researcher has	0.18
SocialProof: Number of times your *co-authors* cite the researcher	0.12
Cites: Number of times *you* cite the researcher	0.11
TimeInactive: Number of years since researcher's last published paper	-0.27

Table 3. Features and their learned weights in the *All* system.

The *ResearcherID* system performs worse than *Most-cited*, illustrating the importance of modeling the relationship between an influencer and a user, rather than just relying on a global ranking of influencers (further experiments also showed adding these features impairs the *All* system). It is nevertheless interesting to consider the resulting ranking of researchers using *ResearcherID* (Table 2). These researchers are certainly some of the leaders of the ACL community.

Fig. 4 shows how system performance depends on the amount of training data. Even with very little training data (50 or so completed screens), the *All* system exceeds both the *Most-cited* baseline and the *Basic Network* system performance. At the same time, *All* continues to improve up to the full 3K training instances, and so collecting further data may well increase accuracy further.

Table 3 provides the feature weights of *All* and thus a picture of how *All* computes its scores. We regard these numbers cautiously as the feature *values* themselves have different dynamic ranges; also when there are two similar features, the SVM divides credit between them (e.g. the two *Cites* features).

Since not all the above information is always available, we also ran experiments with reduced feature sets. First, because some bibliographic databases may not provide full paper texts, we ran our system using only paper *meta*-data (e.g. citations, co-authorships). This results in a Top-1 Acc. of 42.8% (significantly worse than *All*, $p<0.01$). Next, there are some collections, such as arxiv.org, which provide full-text, author, and affiliation data, but not citations. When excluding all citation-derived data (but notably not our last-name-count features), we obtain 42.6% (also worse than *All*, $p<0.01$). While worse than *All*, these systems are still much better than the typical *Most-cited* approach, and in the latter case, can be achieved without the use of citations at all!

9 Discussion and Related Work

Why are citations alone not the best predictor of influence? First of all, studies have documented that there are many reasons to cite a paper aside from acknowledging its influence [4]. You may simply be paying homage to an early pioneer, or perhaps criticizing another person's work. Secondly, our work here hints at influence beyond the medium of publication. For example, two people at the same institution are more likely to have a general influence on one another, even when such sharing is not manifested in citations or even co-authorship.

Recognizing that not all citations imply influence, Zhu et al. [16] used machine learning to predict the most important citations in a paper. Like us, they used peer assessment to gather training data. Unlike our approach, their method does not capture the valuable and effective social factors that we consider above.

A few recent papers have made use of the AAN in order to investigate the spread of scientific ideas [18, 19]. Radev et al. [20] considered citation networks and collaboration networks separately, but did not integrate these into a joint model of influence, as we do. Johri et al. [21] used topic models to identify different types of collaboration in ACL articles (e.g. apprenticeship, synergistic, etc.); we could potentially exploit this data as an additional information source in order to refine the co-authorship features in our system.

There are ties between research in bibliometrics and research in broader social networks. For example, studies of influence have been performed within online social networks [1]. Academic social networks have also been used as case-study social-networks in many publications, including the foundational paper of the important *link prediction problem* in social networks [22].

Our models of influence have immediate application in the real world. For example, Fig. 5 shows the predicted influence on various researchers of the first author of this paper versus whether they responded to his request to complete the survey. When the predicted influence was high, people were more likely to respond to the request. Since it takes effort to contact people and configure their surveys, we could save time by using the predicted influence as a guide for whom to contact. As another example, note that our results suggest that the

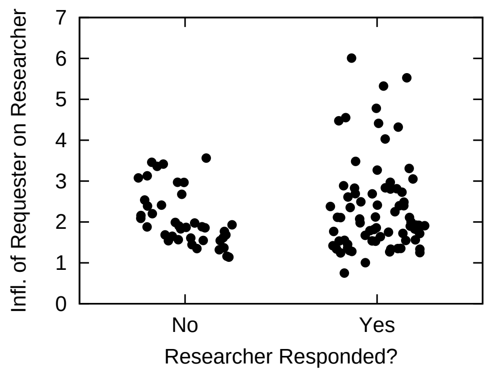


Fig. 5. The predicted influence on researchers (via *All*) of the person requesting participation in the survey, and whether the researchers complied with the request.

more people you cite, the more influence you will have. Unlike other conferences, AI'2014 does not allow authors an extra page for citations. This might be limiting the potential influence of work published at this venue!

10 Conclusion and Future Work

We have proposed a novel approach to the prediction of scientific influence. We extracted a dataset of researchers from the field of computational linguistics, and used readily-available, publication-derived data to create *scientific social networks* for each researcher. To generate training data for our approach, we asked a subset of these researchers to perform an online survey to identify the other researchers who most influenced them. Our survey enjoyed a 66% response rate, and users provided very high-quality ratings, rarely contradicting themselves and clicking across the whole range of options as expected. Our system, trained on these ratings, strongly outperformed the baselines and standard citation-based approach. The overall large gains in accuracy were attributed to the combination of small contributions from each of a variety of novel and interesting features.

The next step for this line of research is to apply our approach to other domains. Do we need peer assessments to train the system in each domain? Could we calibrate our system with only a few ratings? Or are the correlations we observed in our domain actually universal across science? We are also interested in aggregating our scores, both to compare the overall influence of researchers, as well as to answer questions such as, "how much influence does researcher x have in a particular geographic region or sub-community?" We are also interested in studying the dynamics of influence as people move into and out of the field.

References

1. Aral, S., Walker, D.: Identifying influential and susceptible members of social networks. Science **337**(6092) (2012) 337–341

2. Tang, T.Y., Winoto, P., McCalla, G.I.: Further thoughts on context aware paper recommendations for education. In Manouselis, N., et al., eds.: Recommender Systems for Technology Enhanced Learning. Springer (2013) In press, 16 p.
3. Hirsch, J.E.: An index to quantify an individual's scientific research output. Proceedings of the National academy of Sciences of the United States of America **102**(46) (2005) 16569–16572
4. Bornmann, L., Daniel, H.D.: What do citation counts measure? A review of studies on citing behavior. Journal of Documentation **64**(1) (2008) 45–80
5. Van Raan, A.F.: Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. Scientometrics **67**(3) (2006) 491–502
6. Radev, D.R., Muthukrishnan, P., Qazvinian, V.: The ACL anthology network corpus. In: Proc. ACL Workshop on Natural Language Processing and Information Retrieval for Digital Libraries. (2009) 54–61
7. Turney, P.D., Pantel, P.: From frequency to meaning: Vector space models of semantics. Journal of Artificial Intelligence Research **37**(1) (2010) 141–188
8. Lawani, S.M., Bayer, A.E.: Validity of citation criteria for assessing the influence of scientific publications: New evidence with peer assessment. Journal of the American Society for Information Science **34**(1) (1983) 59–66
9. Joachims, T.: Optimizing search engines using clickthrough data. In: Proc. KDD. (2002) 133–142
10. Cialdini, R.B.: Influence: The Psychology of Persuasion. HarperCollins (2007)
11. Milgram, S.: Behavioral study of obedience. The Journal of Abnormal and Social Psychology **67**(4) (1963) 371–378
12. Liu, X., Bollen, J., Nelson, M.L., Van de Sompel, H.: Co-authorship networks in the digital library research community. Information processing & management **41**(6) (2005) 1462–1480
13. Bergsma, S., Dredze, M., Van Durme, B., Wilson, T., Yarowsky, D.: Broadly improving user classification via communication-based name and location clustering on twitter. In: Proc. NAACL-HLT. (2013) 1010–1019
14. Bergsma, S., Lin, D.: Bootstrapping path-based pronoun resolution. In: Proc. Coling-ACL. (2006) 33–40
15. Dubin, D.: The most influential paper Gerard Salton never wrote. Library trends **52**(4) (2004) 748–764
16. Zhu, X., Turney, P., Lemire, D., Vellino, A.: Measuring academic influence: Not all citations are equal. Journal of the American Society for Information Science and Technology (JASIST) (2013) to appear
17. Joachims, T.: Training linear SVMs in linear time. In: Proc. KDD. (2006)
18. Hall, D., Jurafsky, D., Manning, C.D.: Studying the history of ideas using topic models. In: Proc. EMNLP. (2008) 363–371
19. Gerrish, S., Blei, D.M.: A language-based approach to measuring scholarly impact. In: Proc. ICML. (2010) 375–382
20. Radev, D.R., Joseph, M.T., Gibson, B., Muthukrishnan, P.: A bibliometric and network analysis of the field of computational linguistics. Journal of the American Society for Information Science and Technology (JASIST) (2009)
21. Johri, N., Ramage, D., McFarland, D., Jurafsky, D.: A study of academic collaborations in computational linguistics using a latent mixture of authors model. In: Proc. 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities. (2011) 124–132
22. Liben-Nowell, D., Kleinberg, J.: The link prediction problem for social networks. In: Proc. CIKM. (2003) 556–559