Supporting the Curation of Twitter User Lists

Derek Greene, Fergal Reid, Pádraig Cunningham
School of Computer Science & Informatics,
University College Dublin, Ireland

Gavin Sheridan
Storyful
Dublin, Ireland

1 Introduction

Media outlets can now break or cover stories as they evolve by leveraging the content produced by users of social media sites (*e.g.* videos, photographs, tweets). However, significant issues arise when trying to (a) identify content around a breaking news story in a timely manner, (b) monitor the proliferation of content on a certain news event over a period of time, and (c) ensure that this content is reliable and accurate. Storyful[1] is a social media news agency established in 2009 with the aim of filtering news, or newsworthy content, from the vast quantities of noisy data on social networks. To this end, Storyful invests considerable time into the manual curation of content on networks such as Twitter and YouTube. In some cases this involves identifying "gatekeepers" who are prolific in their ability to locate, filter, and monitor news from eyewitnesses.

Twitter users can organise the users they follow into Twitter *lists*. Storyful maintains lists of users relevant to a given news story, as a means of monitoring breaking news related to that story. Often these stories generate community-decided hashtags (*e.g. #occupywallstreet*) – but even with small news events, using such hashtags to track the evolution of a story becomes difficult. Spambots quickly intervene, while users with no proximity (in space, time or expertise) to the news story itself drown out other voices. Manual curation of lists is one way to overcome this problem, but is time consuming, and risks incomplete coverage. To support the list curation process, we propose methods for identifying the important users that form the "community" around a news story on Twitter. Specifically, given a small seed list of users supplied by a journalist, we use network analysis techniques to produce an expanded user list that provides comprehensive coverage of the story.

A number of authors have considered the related problem of producing personal recommendations for users to follow on Twitter, either by following user links or by analysing tweet content. Hannon *et al.* [5] proposed techniques for producing personal recommendations based on the similarity of the aggregated tweets or "profiles" of users that are connected to the ego in the Twitter social graph. Such techniques have primarily relied on a single view of the network to produce suggestions. However, we can view the same Twitter network from a range of different perspectives. For instance, Conover *et al.* [3] performed an analysis of Twitter data based on references to other Twitter screen names in a tweet, while researchers have also looked at the diffusion of content via *retweets* to uncover the spread of memes and opinions on Twitter [3, 7]. The idea is that both *mentions* and *retweets* provide us with some insight of the differing interactions between microblogging users.

In Section 2 we outline a set of recommendation criteria **and** network exploration methods used to support user list curation on Twitter. Rather than using a single view of the network to produce recommendations, we employ a *multi-view* approach that produces user rankings based on different graph representations of the Twitter network surrounding a given user list, and combines them using an SVD-based aggregation approach [9]. Notably, we consider the analysis of a *co-listed graph*, which has not been widely explored in the literature – we look at relations based on the weighted aggregation of co-assignments to Twitter user lists. At an aggregate level, this could be regarded as a form of crowd-sourced curation, where the assumption is that related pairs of users will be more frequently assigned to the same list than users who are dissimilar to one another. Information from multiple views is also used to control the exploration of the Twitter network – this is an important consideration due to the limitations surrounding Twitter data access.

[1] `http://www.storyful.com`

2 Methods

Here we provide a high-level overview of the list curation system, with a focus on the user recommendation problem. Full details of the methods employed in the system are provided in an extended version of this paper [4]. The overall process has three key phases as shown in Figure 1. The initial input to the system is a seed list of one or more users that have been manually labelled as relevant to a particular news story. Once the seed list has been supplied, the first operation of the system involves a *bootstrapping phase*, which retrieves follower ego networks around all seed list members, as well as user list membership, and a limited number of tweets.

After the bootstrap phase, the system will have two disjoint sets for the news story. The *core set* contains curated Twitter accounts, initially corresponding to the members of the seed list. The *candidate set* contains Twitter accounts that are not in the core set, but which exist in the wider network around the core – some of these users may potentially be relevant for curation, while others will be spurious. In the *recommendation phase*, a ranked list of the r top users from the candidate set is produced. This is done by analysing four different views of the network comprising the core set and the candidate users linked to the core set. The four views are the *friend graph*, the *mention graph*, the *retweet graph*, and the *weighted co-listed graph* [4].

The criteria used to rank users in these graphs are in-degree, HITS with priors [6], and a novel normalised in-degree measure [4]. Since some criteria are only meaningful when applied to certain graphs, for the evaluations described in this paper, we use the following five combinations:

- Normalised in-degree applied to the core friend graph.
- HITS with priors applied to the core friend graph.
- Weighted in-degree applied to the co-listed, mention, and retweet graphs.

The various graph/criterion combinations can potentially produce rankings of users that differ significantly. To combine rankings, we use SVD-based aggregation, which has previously been shown to be effective for this task [9]. This aggregation process produces a single ranking that is presented to a human curator who can then select a subset to migrate to the core set (*i.e.* to augment the existing Twitter user list). For the purpose of the evaluation described here, we automatically select the top five users at each iteration.

The system can iterate at this point by updating the core set and repeating the recommendation process (see Fig. 1). This update process is not completely straightforward as the underlying network is evolving due to changes in follower links, list memberships, and newly-posted tweets containing mentions and retweets. The iterations are repeated for as long as the news story is being monitored.

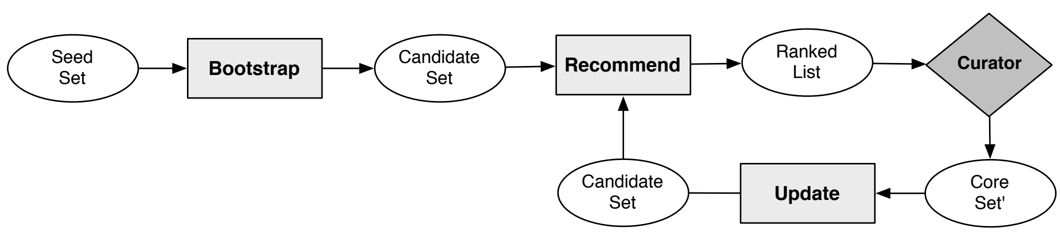
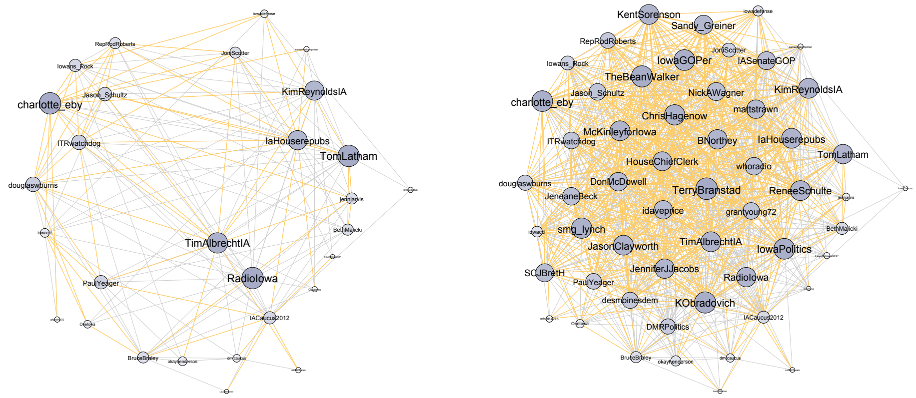


Figure 1: Curation support system, with the *bootstrap*, *recommendation*, and *update* phases.

3 Case Study 1: Iowa

We evaluated the recommendation system on a list previously curated by Storyful, covering Iowa politics during the 2012 US Presidential Primaries[2]. At the time of initial data collection – 16 September 2011 – this list contained 128 users. To evaluate the user recommendation process, we use cross validation, randomly dividing the complete Iowa user list into four disjoint subsets, which we then expand independently of each other, before comparing the results against the original complete list. Fig. 2(a) shows the 32 user subgraph of the induced by Iowa set #1, using a force directed layout [2]. We then applied the Fig. 1 workflow to each of the sets, for six iterations.

[2] `http://twitter.com/#!/trailmix12/iowa`



(a) Initial core set

(b) Final core set

Figure 2: Follower graph for core set members in the Iowa set #1 after (a) the initial bootstrap phase, (b) six complete iterations. Larger nodes with a more saturated colour are indicative of nodes with a higher in-degree (*i.e.* users with more followers within the core set). Highlighted edges indicate reciprocated follower links between users. Layout positions are preserved for both figures.

Iteration	**Precision**					**Recall**				
	Set #1	*Set #2*	*Set #3*	*Set #4*	*Mean*	*Set #1*	*Set #2*	*Set #3*	*Set #4*	*Mean*
1	0.95	0.97	0.97	1.00	**0.97**	0.27	0.28	0.28	0.29	**0.28**
2	0.93	0.98	0.95	1.00	**0.96**	0.30	0.32	0.31	0.33	**0.32**
3	0.94	0.96	0.96	0.98	**0.96**	0.34	0.35	0.35	0.36	**0.35**
4	0.90	0.94	0.96	0.96	**0.94**	0.37	0.38	0.39	0.39	**0.38**
5	0.88	0.93	0.93	0.95	**0.92**	0.39	0.41	0.41	0.42	**0.41**
6	0.82	0.92	0.89	0.89	**0.88**	0.40	0.45	0.43	0.43	**0.43**

Table 1: Precision and recall scores for four subsets of the Iowa user list.

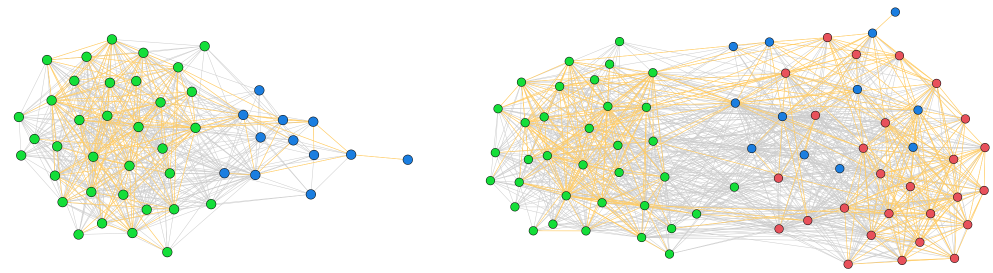
During each iteration, the system generated $r = 50$ recommendations. After each complete iteration the system selected the top five highest ranked users (based on SVD aggregation) to add to the core set. The six iterations thus yielded 30 additional core users for each of the four sets. An example of the final expanded core set for Iowa set #1 is shown in Fig. 2(b). Note that several high-degree nodes were added to the core set, such as the user @*TerryBranstad*, Governor of Iowa. By the final iteration, users were selected from a complete candidate set with average size of $\approx 62k$ users. Note that after each iteration, we filtered out users with $> 50,000$ followers; furthermore, we only retrieved the first 1,000 links, lists, and tweets, for each user queried. We also removed users who had not tweeted in the previous two weeks, or had fewer than 25 tweets in total.

We evaluate the recommendations in terms of *precision* and *recall*. The results for the four subsets across all six iterations are listed in Table 1. We observe that, in terms of recall, increasing the user list size by 30 accounts does not lead to a significant fall in precision – average precision relative to the complete original list remains at 0.88 by iteration six. Meanwhile, recall increases steadily in all cases – the average is 0.43. Note the maximum achievable recall by iteration six is 0.48, and is lower in previous iterations. We observe that the Iowa user list studied here is a relatively homogeneous group of users pertaining to a focused story – the users are predominantly Republicans involved in the Iowa caucuses. Therefore, unlike the study in [3] which analysed Twitter relations across the entire U.S. during 2010 midterm elections, here a pronounced partisan divide is not evident.

4 Case Study 2: Bahrain

For our second study, we analysed a significantly different dataset. As a seed list we use a Twitter list covering the current political situation in Bahrain which was also manually curated by Storyful[3]. As of 27 September 2011, this list contained 51 users. A small number of these have a "loyalist" or "pro-government" stance, while the remaining users could be regarded as being either "non-

[3] `http://twitter.com/#!/storyfulpro/bahrain`



(a) Initial core set (b) Final core set

Figure 3: Follower graph for core set members in the Bahrain dataset (a) at the initial bootstrap phase, (b) after four complete iterations. Blue nodes denote users in the original user list labelled as "loyalist", while the remaining members of the user list are coloured green. The additional nodes that have been selected, based on recommendations using *Bahrain-L* as a seed list, are coloured red.

loyalist", or "neutral" observers with an interest in Bahrain. This natural division in the seed list raises an interesting question – does starting with a seed list that takes a particular stance on a given news story lead to the construction of localised network "silos", which may lead an automated system to give biased user recommendations? To investigate this, we generate recommendations based on a seed list *Bahrain-L* containing a subset of 14 users labelled as "loyalist". We ran four complete iterations using the same mechanism as the previous evaluation. This resulted in a core set containing 34 users, from a candidate set of 51,114 users.

Fig. 3(a) shows the subgraph induced by the original complete curated list of 51 users on the follower graph – the split between loyalist users and other users is evident from the positions calculated by force directed layout. In particular, the latter group of users form a densely connected core, while most of the "loyalist" nodes are not well-connected with the rest of the subgraph. Fig. 3(b) shows a subgraph induced by the union of the curated list, with the set of nodes selected based on the recommendation process using *Bahrain-L* alone as the seed set. We observe that none of the 37 non-loyalist nodes from the curated list were selected during the four iterations. In contrast, we see that the new users are closely connected with the other loyalist users, forming a second dense core. While we might expect this if recommendations were only generated based on follower links, recall that rankings based on mentions and retweets are also being aggregated to select new users. In fact, the addition of these rankings appears to further compound the "silo" effect which is evident from Fig. 4. This contrasts with the findings in [3], where user mentions crossed the partisan divide. More generally, these results highlight that political polarisation can be strongly pervasive in microblogging networks, as is the case in other forms of social media [1].

Our analysis suggests there is little interaction on Twitter between users with opposing stances on Bahraini politics. On the one hand, this highlights a weakness of the proposed recommendation techniques in the case of highly-polarised stories. Alternative criteria, which emphasise diversity over homogeneity (*e.g.* representative sampling via clustering or community finding), may provide a solution. This would be analogous to techniques in active learning that identify diverse examples [8]. On the other hand, these results also highlight the continued importance of the role of the curator in (a) selecting a suitably diverse seed list, (b) actioning recommendations produced by the system.

5 Conclusions

In this paper, we have proposed a comprehensive approach for automating aspects of the Twitter list curation process, based on novel network exploration and multi-view recommendation techniques. We showed that by using different starting subsets of a manually-curated list, we can recall the original human annotations while maintaining high precision. The next phase of our work will study diffusion patterns of newsworthy multimedia resources (*e.g.* links to images and videos) in the network surrounding a user list. We also intend to study the proposed recommendation and network exploration techniques beyond Twitter, looking at multiple views across different social networks.

This work is supported by Science Foundation Ireland under grant 08/SRC/I140: Clique: Graph and Network Analysis Cluster.

References

[1] L.A. Adamic and N. Glance. The political blogosphere and the 2004 US election: Divided they blog. In *Proc. of 3rd International Workshop on Link Discovery*, pages 36–43. ACM, 2005.

[2] M. Bastian, S. Heymann, and M. Jacomy. Gephi: An open source software for exploring and manipulating networks. In *Proc. 3rd International Conference on Weblogs and Social Media (ICWSM'09)*, pages 361–362. AAAI, 2009.

[3] M. D. Conover, J. Ratkiewicz, M. Francisco, B. Gonçalves, A. Flammini, and F. Menczer. Political polarization on twitter. In *Proc. 5th International Conference on Weblogs and Social Media (ICWSM'11)*. AAAI, 2011.

[4] D. Greene, F. Reid, G. Sheridan, and P. Cunningham. Supporting the Curation of Twitter User Lists. *ArXiv e-prints*, (1110.1349), October 2011.

[5] J. Hannon, M. Bennett, and B. Smyth. Recommending twitter users to follow using content and collaborative filtering approaches. In *Proc. 4th ACM Conference on Recommender Systems (RecSys'10)*, pages 199–206. ACM, 2010.

[6] J.M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632, 1999.

[7] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonçalves, A. Flammini, and F. Menczer. Detecting and tracking political abuse in social media. In *Proc. 5th International Conference on Weblogs and Social Media (ICWSM'11)*. AAAI, 2011.

[8] B. Settles. Active Learning Literature Survey. Technical Report 1648, University of Wisconsin–Madison, 2009.

[9] G. Wu, D. Greene, and P. Cunningham. Merging multiple criteria to identify suspicious reviews. In *Proc. 4th ACM Conference on Recommender Systems (RecSys'10)*, pages 241–244. ACM, 2010.