

Aggregating Content and Network Information to Curate Twitter User Lists



Derek Greene

Barry Smyth

Gavin Sheridan

Pádraig Cunningham

RecSys-12: Recommender Systems & The Social Web



Clique: Graph & Network Analysis Cluster

School of Computer Science & Informatics

University College Dublin, Ireland



Task: User List Curation



- Twitter allows the grouping of users into topical **user lists**.
- **Storyful** maintains lists of users for news stories to monitor breaking news related to that story.

bahrain
Covering Bahrain

A public list
Curated by StoryfulPro

Following: 56 Followers: 17

People this list follows

- IAA_Bahrain** IAA Bahrain
The Official Twitter Channel for the Information Affairs Authority, Kingdom of Bahrain. If you have questions or comments please follow @IAA_Sara
- am_aswad** Ali M. Alaswad, X-MP
Name: Ali Mahdi Ali Alaswad Resigned MP-Bahrain(2010-2011) Relationship Status: Married Nationality: Bahrain Date of Birth: 1976
- ghassandhaif** Ghassan Dhaif
السلام عليك سيدي ومولاي يااباعبدالله الحسين
- abdadibh** Abdulla Alderazi
BHRS secretary General
- aibahrain** Amnesty Bahrain
Human Rights information and updates on Bahrain run by Amnesty International.
- bahrainiac14** المبرقة 14
Sarcastic Writer, blogger and human rights activist كاتبة ساخرة، مدونة، ناشطة في حقوق الانسان Bahrain
- abbasalomran** Abbas Al Omran
Human rights activist , Member Bahrain Center Human Rights
- ag47ent** Ali Mahdi
ITweet about Bahrain Revolution in AR\ENG ..I HATE PRO-DICTATORS (6abala) | A NORMAL GUY WHO IS SEEKING FOR FREEDOM http://ask.fm/ag47ent

Recently Following

This list's most recently followed accounts · view all

- IAA_Bahrain** IAA Bahrain · Follow
The Official Twitter Channel for the Information ...
- am_aswad** Ali M. Alaswad, X-MP · Follow
Name: Ali Mahdi Ali Alaswad Resigned MP-Bahrain...
- ghassandhaif** Ghassan Dhaif · Follow
السلام عليك سيدي ومولاي يااباعبدالله الحسين
- abdadibh** Abdulla Alderazi · Follow
BHRS secretary General

More lists by @storyfulpro · view all

- libya
- iraq
- southafrica
- jordan
- pakistan
- china
- bahrain
- greece
- nationalsecurity
- uspolitics
- saudi
- whitehouse
- afghanistan
- israel
- syria
- palestine
- japan
- usweather
- yemen
- egypt

About Help Blog Mobile Status Jobs Terms Privacy Shortcuts Advertisers Businesses Media Developers Resources © 2011 Twitter

bahrain
Covering Bahrain

A public list
Curated by StoryfulPro

Following: 56 Followers: 17

@storyfulpro/bahrain

- S_R_BH** S.R.BH
نبيل رجب: اهالي الدراز يخرجون الان احتجاجا على مقتل الشهيد والمواجهات مستمرة هناك | 21. | #lulu #humanrights Bahrain #feb14
- marhoon** Hussain Marhoon
الداخلية البحرينية تقول إنها ستحقق في مقتل الشاب أحمد جابر 16 عاماً اليوم pic.twitter.com/LWKhoQW5 #BAHRAIN #FEB14 #LULU
- LuaLuaTV** LuaLuaTV
مئات من الشباب الغاضب في منطقتي الشاخوره وأبوصبيع يغلقون شارع البديع قبل قليل #lualuatv #bahrain
- S_R_BH** S.R.BH
وزارة البلطجية "الداخلية" | تلقت غرفة العمليات بلاغاً من المستشفى الدولي يفيد بإحضار أحد المصابين، عقب ذلك تبين... fb.me/QgXMuvRz
- NABEELRAJAB** Nabeel Rajab
مواجهات في بني جمرة الان وكل المناطق تنتفض والمرتزة والبلطجية يعيثون فسادا في البلاد

Recently Following

This list's most recently followed accounts · view all

- IAA_Bahrain** IAA Bahrain · Follow
The Official Twitter Channel for the Information ...
- am_aswad** Ali M. Alaswad, X-MP · Follow
Name: Ali Mahdi Ali Alaswad Resigned MP-Bahrain...
- ghassandhaif** Ghassan Dhaif · Follow
السلام عليك سيدي ومولاي يااباعبدالله الحسين
- abdadibh** Abdulla Alderazi · Follow
BHRS secretary General

More lists by @storyfulpro · view all

- libya
- iraq
- southafrica
- jordan
- pakistan
- china
- bahrain
- greece
- nationalsecurity
- uspolitics
- saudi
- whitehouse
- afghanistan
- israel
- syria
- palestine
- japan
- usweather
- yemen

Supporting User List Curation



- **Idea:** Use data analysis to identify important users that form the “community” around a news story on Twitter.
- **Goal:** Recommend new users to expand an embryonic seed list, helping to find valuable content relating to the story.

The image shows a Twitter list titled "bahrain" with the following details:

- Account:** @storyfulpro/bahrain
- Type:** A public list, Curated by StoryfulPro
- Stats:** Following: 56, Followers: 17
- Recent Tweets:**
 - S_R_BH** (S.R.BH) - 1 minute ago: نبيل رجب: اهالي الدراز يخرجون الان احتجاجا على مقتل الشهيد والمواجهات مستمرة هناك |.21|. #lulu #humanrights Bahrain #feb14
 - marhoon** (Hussain Marhoon) - 2 minutes ago: الداخلية البحرينية تقول إنها ستحقق في مقتل الشاب أحمد جابر 16 عاماً اليوم pic.twitter.com/LWKhoQW5 #BAHRAIN #FEB14 #LULU
 - LuaLuaTV** (LualuTV) - 2 minutes ago: مئات من الشباب الغاضب في منطقتي الشاخوره وأبوصيبع يلقون شوارع البديع قبل قليل #lualuatv bahrain
 - S_R_BH** (S.R.BH) - 3 minutes ago: وزارة البلطجية "الداخلية" | تلقت غرفة العمليات بلاغاً من المستشفى الدولي يفيد بإحضار أحد المصابين، عقب ذلك تبين... fb.me/QgXMuvRz
 - NABEELRAJAB** (Nabeel Rajab) - 3 minutes ago: مواجهات في بني جمرة الان وكل المناطق تنتفض والمرتزة والبلطجية يعيثون فسادا في البلاد

On the right side of the list, there are sections for "Recently Following" and "More lists by @storyfulpro".

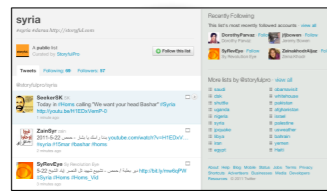
Four source boxes on the left point to the list:

- Bahrain By-Elections** (@voteBH) - Bahrain. The official twitter account for Bahrain By-Elections 2011. <http://www.vote.bh>
- محمد العرب** (@MohamedAlArab1) - twitter - facebook. الحقيقة الضائعة تحتاج دوما الى صحفي كقربان. <http://www.facebook.com/pages/Mohamed-alarab/192731494078104>
- IAA Bahrain** (@IAA_Bahrain) - Kingdom of Bahrain. The Official Twitter Channel for the Information Affairs Authority, Kingdom of Bahrain. If you have questions or comments please follow @IAA_Sara. <http://www.iaa.bh/>
- Bu Abdullah** (@Khaled_Bin_Ali) - Bahrain. Minister of Justice and Islamic Affairs and Waqf.

Curation System Overview



Seed Set



Bootstrap

Phase 1

Candidate List

Recommender

Ranked List

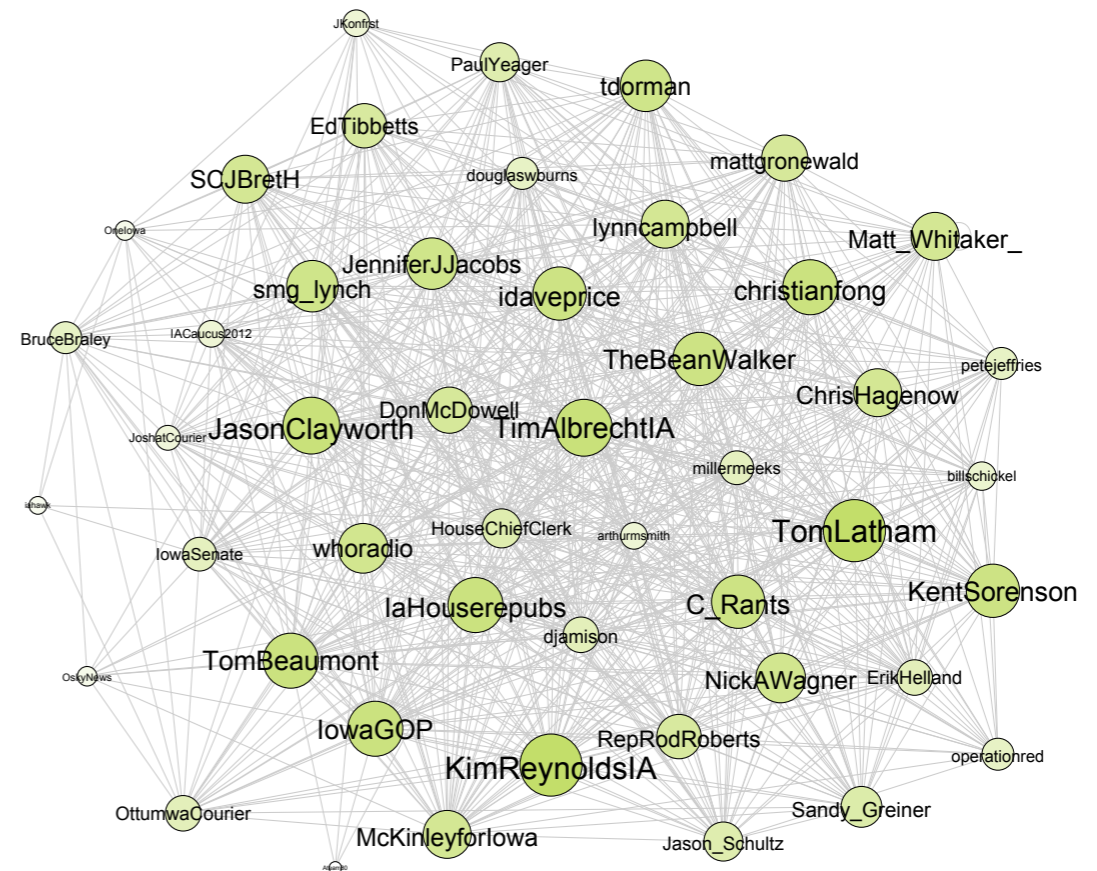
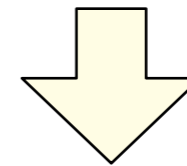
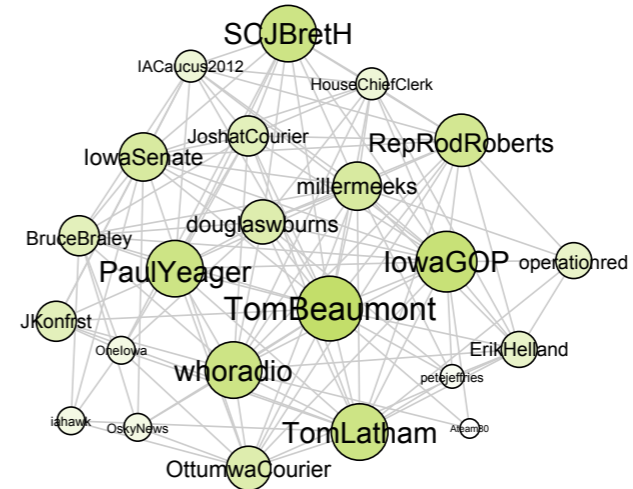
Phase 2

Curator

Core List'

Updater

Candidate List

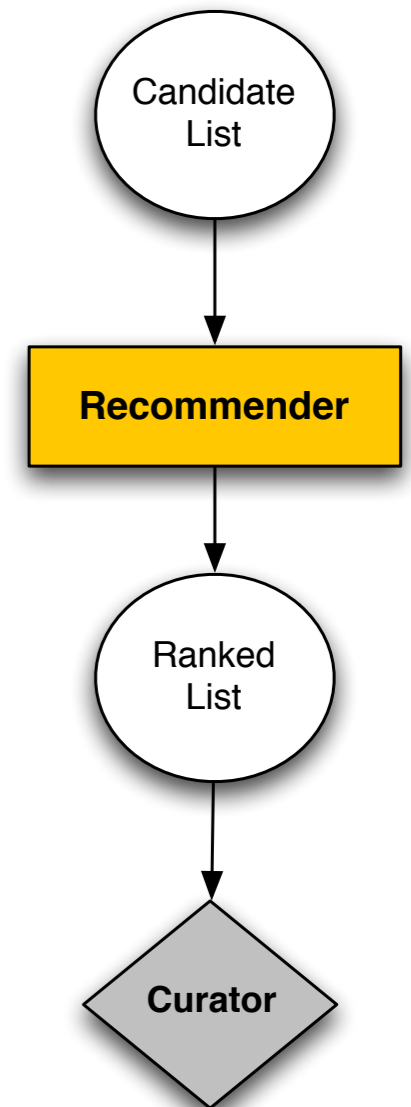


General Recommender

- Use embryonic seed list as initial training data.
- Compute training data centroid.
- Measure cosine similarity between centroid and vectors representing candidate non-seed users.
- ➔ Produce ranked list of top K non-seed users to present to a human curator.

Which criterion should we use for recommendation?

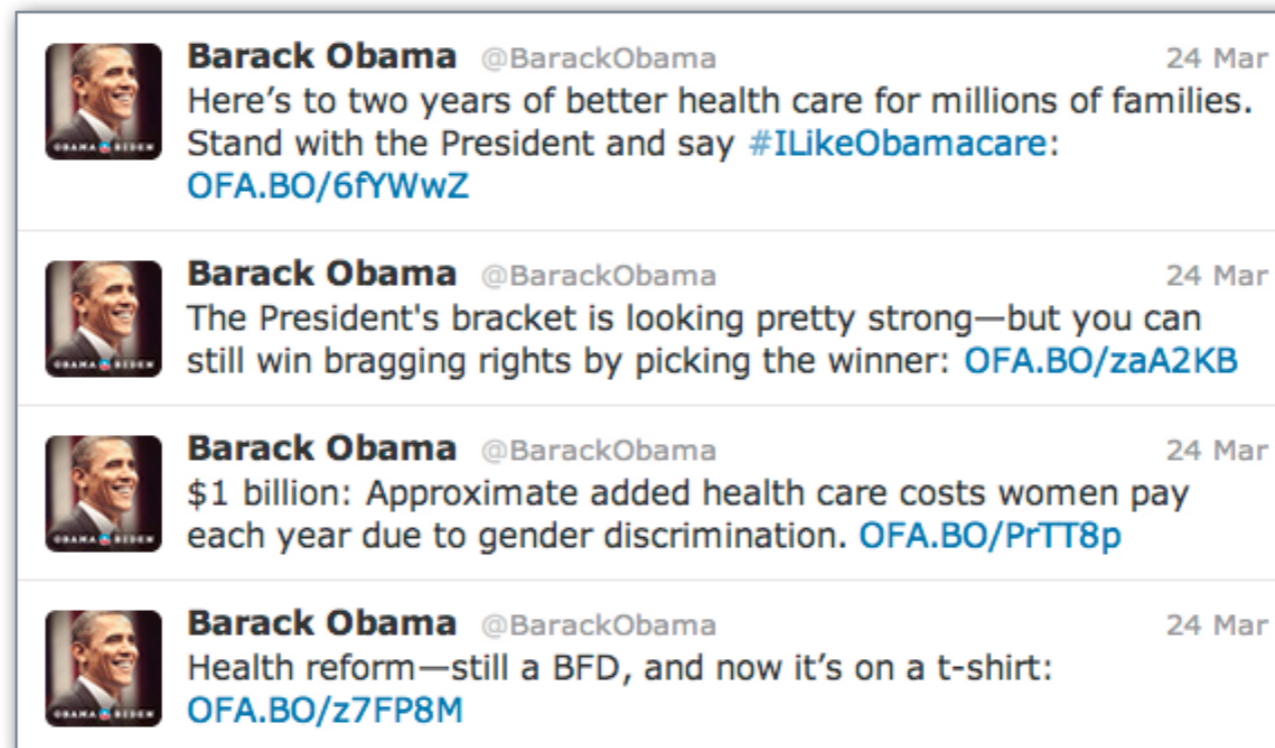
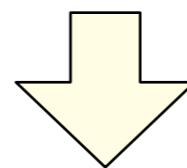
- ➔ Propose variety of vectorised approaches, using criteria based on Twitter network and content analysis.



Content-Based Criteria



- **Tweet profiles:** Create a term vector for each user, containing the aggregation of their 50/100/200 most recent tweets.

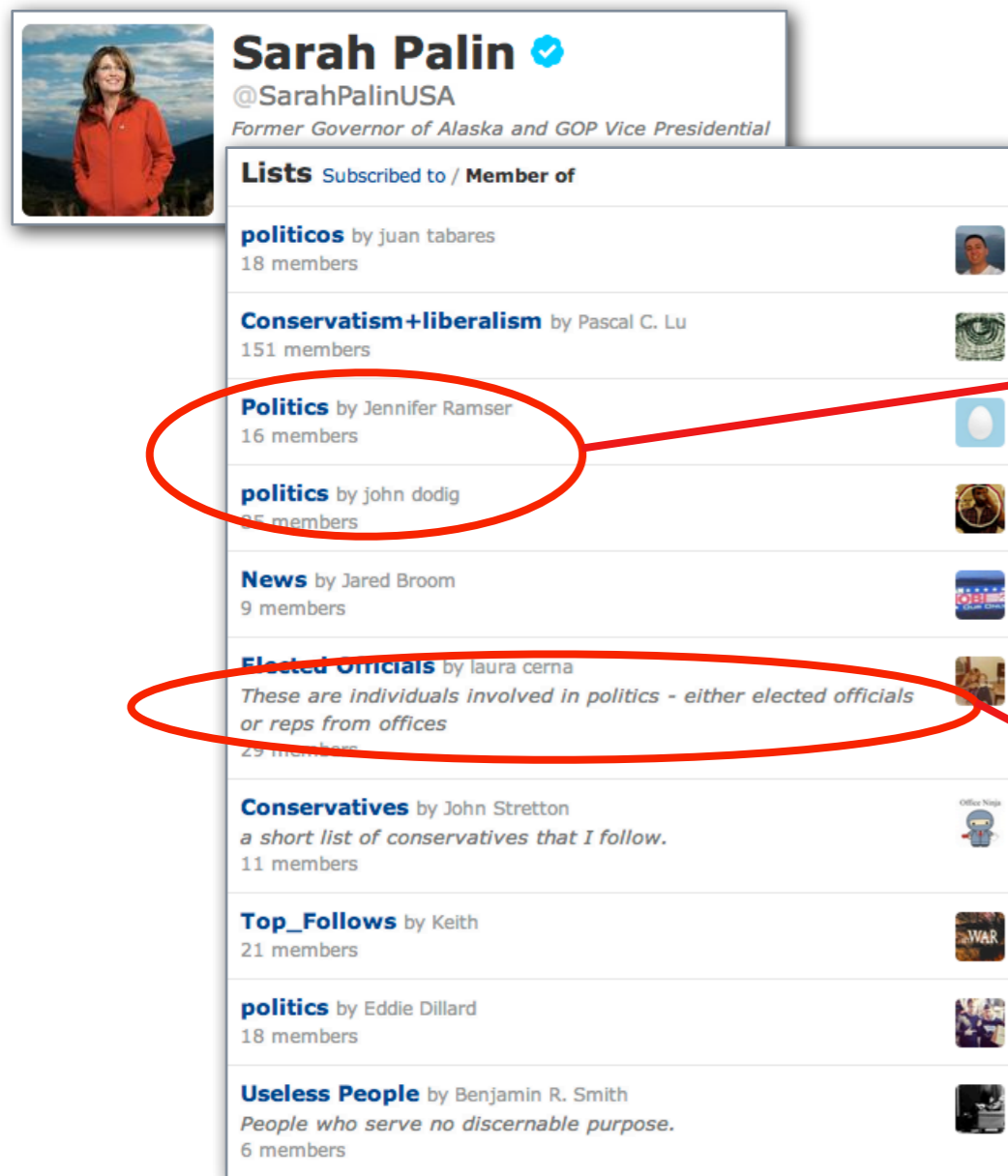
A screenshot of four tweets from Barack Obama (@BarackObama) dated 24 Mar. Each tweet includes a profile picture of Barack Obama, his name, handle, date, and text. The tweets are: 1. 'Here's to two years of better health care for millions of families. Stand with the President and say #ILikeObamacare: OFA.BO/6fYWwZ'; 2. 'The President's bracket is looking pretty strong—but you can still win bragging rights by picking the winner: OFA.BO/zaA2KB'; 3. '\$1 billion: Approximate added health care costs women pay each year due to gender discrimination. OFA.BO/PrTT8p'; 4. 'Health reform—still a BFD, and now it's on a t-shirt: OFA.BO/z7FP8M'.

User	president	health	care	reform	gender	costs	...
@BarackObama	2	2	2	1	1	1	...

Content-Based Criteria



- **List names/descriptions:** Create a term vector for each user, containing the aggregation of names and/or descriptions of user lists to which they have been assigned.



User	politics	news	useless	...
@SarahPalinUSA	2	1	1	...

User	elected	politics	purpose	...
@SarahPalinUSA	1	1	1	...

Network-Based Criteria



- **Followed-by Profiles:** Users are similar if they are "co-followed" by the same set of users. Represent each user as a binary follower profile vector.

User X is followed by...

<i>User X</i>	@TeamGB	@bradwiggins	@Mo_Farah	@britishswimming	
@MichaelPhelps	1	0	0	1	
@chrishoy	1	1	1	0	...
@TomDaley1994	1	1	0	1	

- **Mentioned-by Profiles:** Users are similar if they are mentioned in tweets by the same users (i.e. "co-mentioned").
- **Retweeted-by Profiles:** Users are similar if their posts are retweeted by the same users (i.e. "co-retweeted").

Network-Based Criteria



- **Co-Listed Information:**

- Other media outlets and individuals will also be simultaneously curating user lists on topics in the wider Twitosphere.
- ➔ Would like to “crowdsource” these efforts to support list curation.

A screenshot of a Twitter list titled 'Irishrugby' by Gerry Thornley. The list members are: David Wallace (@wa22y), Brian O'Driscoll (@BrianODriscoll), Ian madigan (@Ian_madigan), Trevor Hogan (@TrevHogan), Rob Kearney (@KearneyRob), Paul O'Connell (@Paul_OConnell), and Ronan O Gara (@RonanOGara10). Red circles highlight Brian O'Driscoll and Rob Kearney. A vertical red line connects the two circles, indicating they are co-listed.

A screenshot of a Twitter list titled 'Ireland rugby squad' by RTÉ Rugby. The list members are: Chris Henry (@chrishenry678), Sean Cronin (@SeanCronin2), Rob Kearney (@KearneyRob), Rhys Ruddock (@RhysRuddock), andrew trimble (@andrew_trimble), Tommy Bowe (@TommyBowe14), and Brian O'Driscoll (@BrianODriscoll). Red circles highlight Rob Kearney and Brian O'Driscoll. A vertical red line connects the two circles, indicating they are co-listed.

Users @BrianODriscoll and @KearneyRob are co-listed on both lists.

Experimental Evaluation



- Collected 10 datasets around "Super Tuesday" GOP nominations in March 2012, with annotated seed users from Storyful.

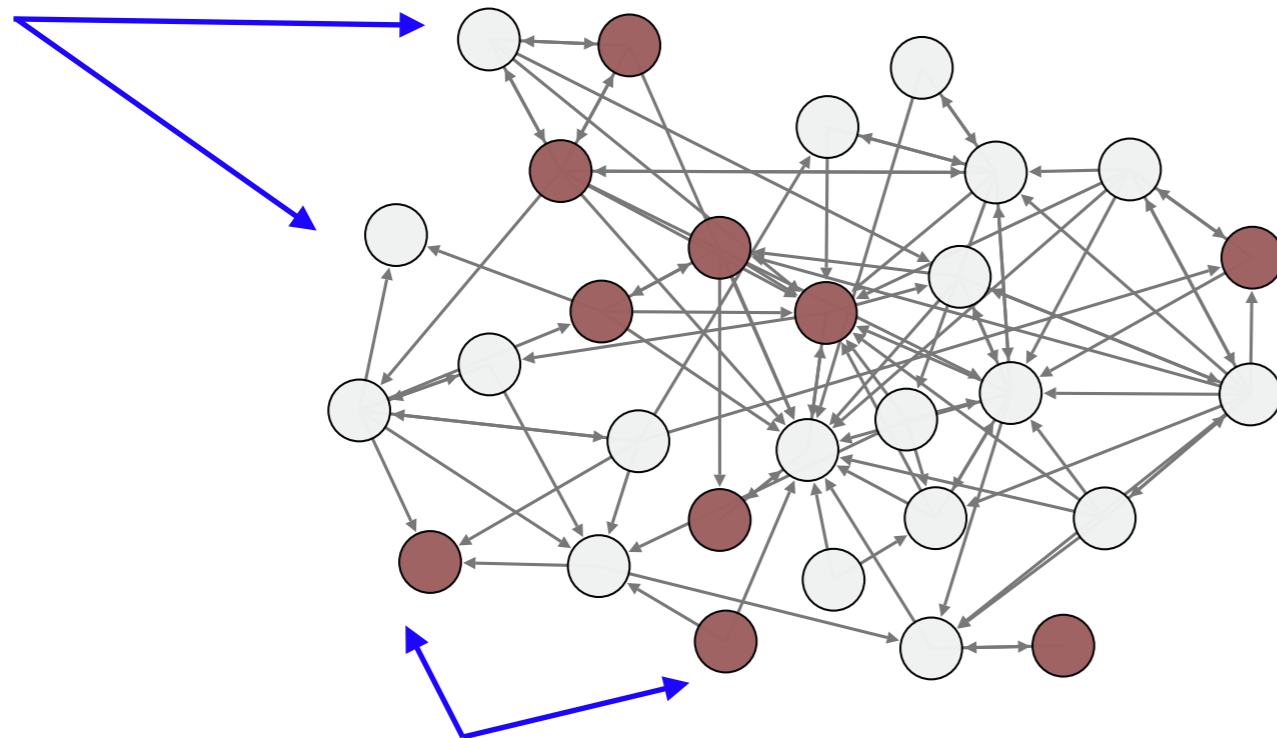


Dataset	Users	Core Users	Tweets	Friends	Followers	Listed
<i>Alaska</i>	948	41	185	208	269	89
<i>Georgia</i>	966	34	211	235	295	126
<i>Idaho</i>	743	20	186	264	273	47
<i>Massachusetts</i>	821	24	209	244	293	122
<i>North Dakota</i>	363	26	203	147	192	93
<i>Ohio</i>	1051	97	178	171	207	115
<i>Oklahoma</i>	693	32	205	178	211	109
<i>Tennessee</i>	979	48	199	170	204	112
<i>Vermont</i>	864	36	182	169	190	66
<i>Virginia</i>	877	46	200	160	199	115

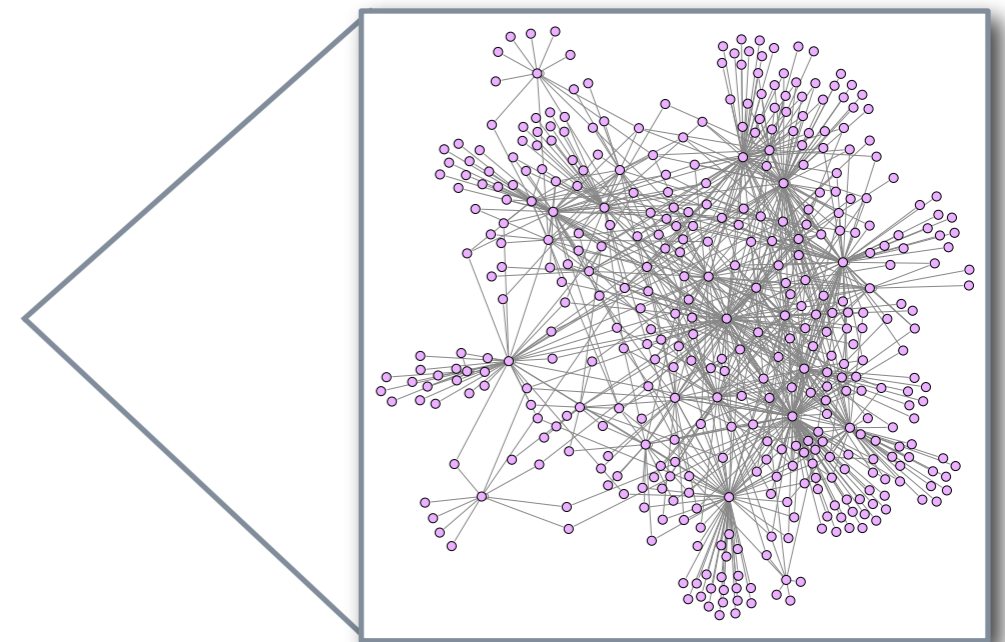
Experimental Setup

- Ran 250 X k -fold cross validation experiments per dataset:
 - Held out a proportion of the seed set as test data.
 - Used the remaining seed users as training data.
 - Ranked users in test data using each criterion.
 - Calculated *precision* and *recall* relative to the test data.

Training Users



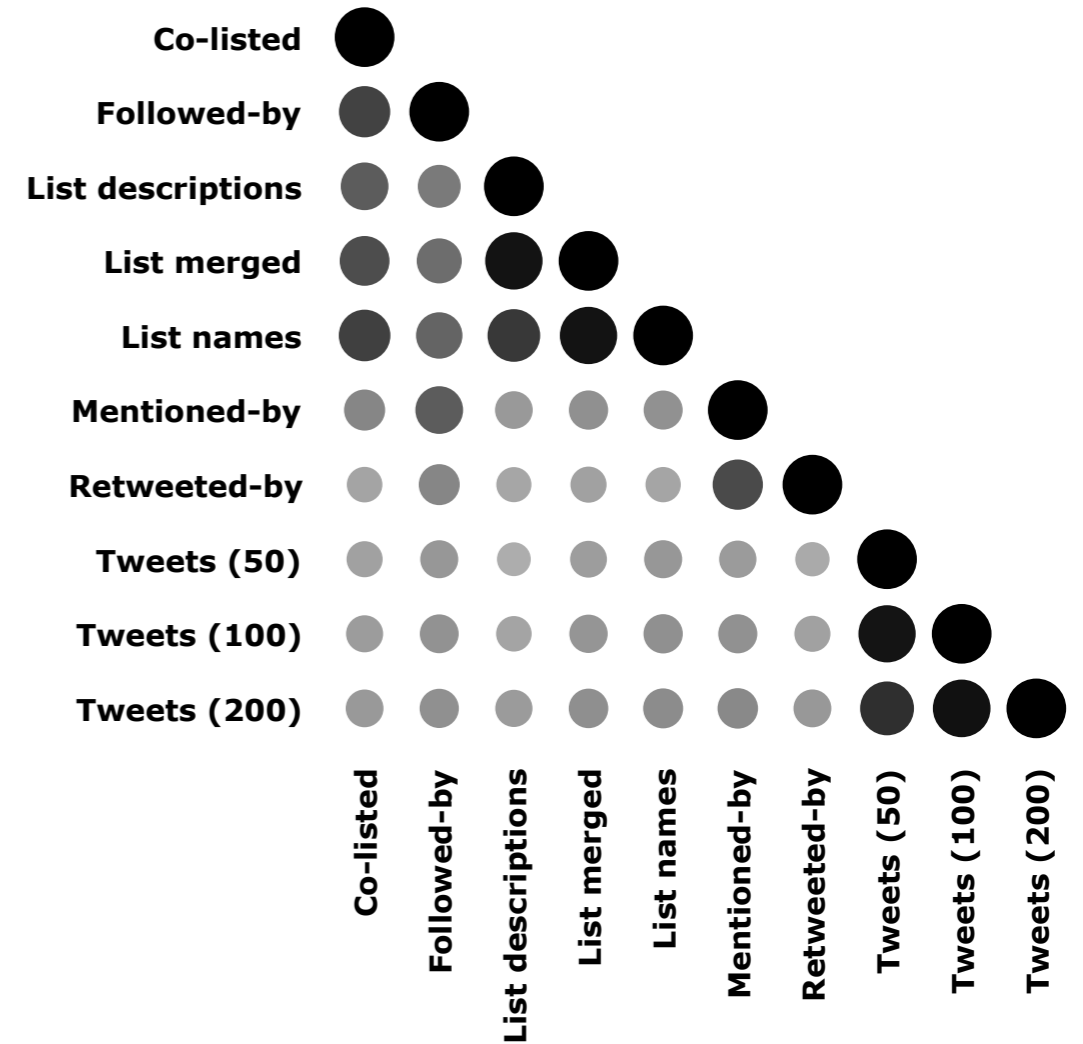
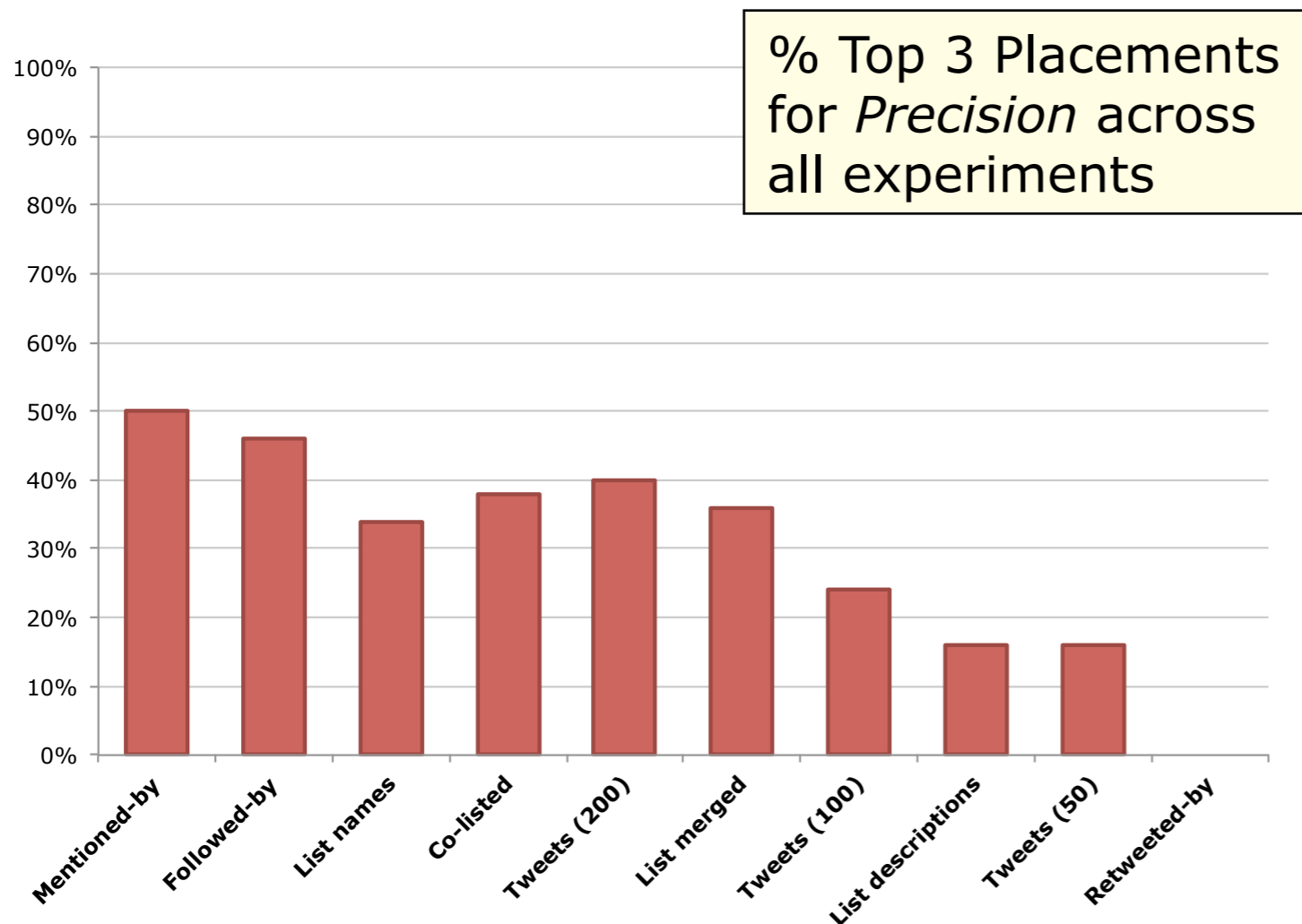
Test Users



Comparison of Criteria



- Looked at diversity among the recommendations from the different criteria...
- We see several distinct signals present across the different views.

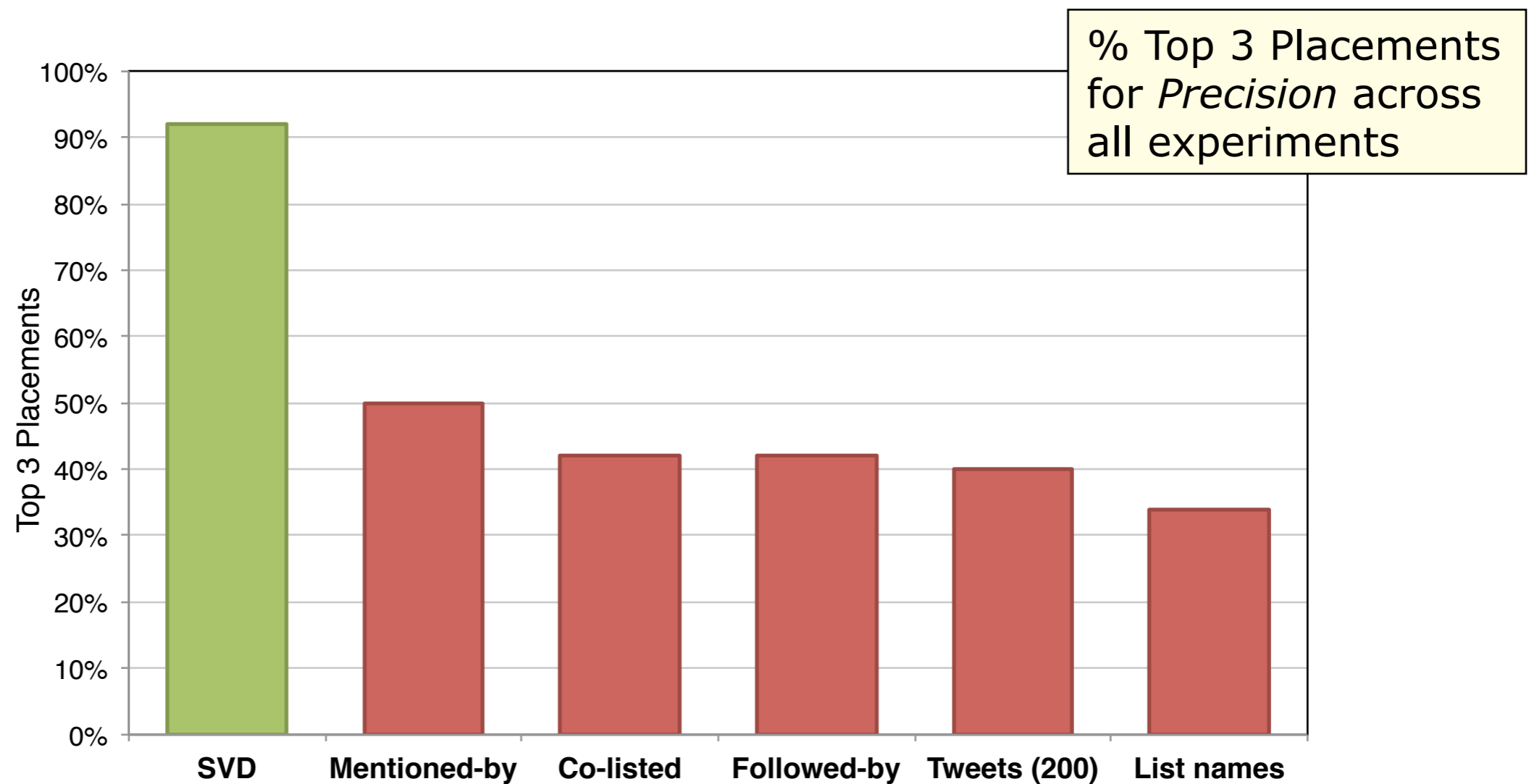


→ No single criterion performs consistently well on all datasets.

Aggregating Multiple Rankings



- **SVD Aggregation:** Combine rankings generated on top 5 individual criteria into a single matrix, apply Singular Value Decomposition, then rank values in first singular vector.



Comparison of SVD Aggregation V Top 5 Individual Criteria

User List Curation System



storyful: Curation System

User Lists

[Create New List...](#)

[Import List...](#)

[Manage Blacklists](#)

>> [afghanistan](#)

>> [africa](#)

>> [alabama](#)

>> [albania](#)

>> [algeria](#)

>> [andalusia](#)

>> [angola](#)

>> [argentina](#)

>> [arizona](#)

>> [armenia](#)

>> [art](#)

>> [asia-pacific](#)

>> [austria](#)

>> [azerbaijan](#)

>> [bahrain](#)

Curation System Home

Recent Activity:

Last bootstrapped: [\[greece\]](#) [\[portugal\]](#) [\[uk\]](#)

Last recommended: [\[greece\]](#) [\[france\]](#) [\[it-security\]](#)

Last updated: [\[greece\]](#) [\[france\]](#) [\[it-security\]](#)

Last modified: [\[master-list\]](#)

Pending Instructions:

Project	Process Type	Requested At	Scheduled For
it-security	CONFIGURE	2012-06-22 12:08	2012-06-22 12:08
angola	UPDATE	2012-06-22 01:04	2012-06-22 12:16
afghanistan	RECOMMEND	2012-06-22 11:51	2012-06-22 12:24
portugal	UPDATE	2012-06-22 01:22	2012-06-22 12:35
france	UPDATE	2012-06-22 02:52	2012-06-22 13:22

Active Processes:

Project	Process Type	Started At	Status
armenia	BOOTSTRAP	2012-06-22 11:59	RUNNING
kosovo	BOOTSTRAP	2012-06-22 11:58	RUNNING
uk	UPDATE	2012-06-22 11:48	RUNNING

- **Summary:**

- Proposed variety of criteria for user list building.
- Performed a comprehensive comparison of individual criteria, demonstrating weaknesses of both content and network-based strategies.
- Demonstrated that more accurate results can be achieved using SVD aggregation.

- **Future Work:**

- Stratification of networks to target recommendations for niche communities - avoiding the "filter bubble".
- Support for curation across multiple social networks.

Any Questions ?



derek.greene@ucd.ie



@derekgreene



<http://cliquecluster.org>